The following publication Hao, M., Lin, Y., Shen, G., & Su, W. (2023). Nonparametric inference on smoothed quantile regression process. Computational Statistics & Data Analysis, 179, 107645 is available at https://doi.org/10.1016/j.csda.2022.107645.

# Penalized integrated check loss based nonparametric inference for composite quantile regression model

Meiling Hao[1], Yuanyuan Lin[2], Guohao Shen[3] and Wen Su[4]

[1]School of Statistics, University of International Business and Economics, Beijing, China

[2]Department of Statistics, The Chinese University of Hong Kong, Hong Kong

[3]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong

[4]Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong

**Abstract:** We investigate global estimation in semiparametric quantile regression models. For estimating unknown functional parameters, we propose an integrated quantile regression loss function with penalization. We first obtain a vector-valued functional Bahadur representation of the resulting estimators, and then derive the asymptotic distribution of the proposed infinite-dimensional estimators. Furthermore, a resampling approach that generalizes the minimand perturbing technique is adopted to construct confidence intervals and to conduct hypothesis testing. Extensive simulation studies demonstrate the effectiveness of the proposed method, and applications to the real estate dataset and world happiness report data are provided.

**Key Words**: Asymptotic normality; Bahadur representation; Quantile regression process.

*Correspondence:* Wen Su, Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong. E-mail: jenna.wen.su@connect.hku.hk

# 1 Introduction

Quantile regression has been gaining much popularity as an alternative to linear regression for its far-reaching applications in numerous disciplines such as econometrics, genetics, and social sciences (Koenker and Hallock, 2001; Cade and Noon, 2003; Briollais and Durrieu, 2014). Let $Y$ denote the response variable and $\mathbf{X}$ represent a $p$-dimensional vector of covariates, then the $\tau$-th conditional quantile of $Y$ given $\mathbf{X}$ is $Q_Y(\tau|\mathbf{X}) = \inf\{t : P(Y \leq t|\mathbf{X}) \geq \tau\}$. For a given quantile level $\tau \in (0, 1)$, Koenker and Bassett (1978) defined the quantile regression model as follows:

$$Q_Y(\tau|\mathbf{X}) = \mathbf{X}^\top \boldsymbol{\beta}_\mathbf{0}(\tau), \tag{1}$$

where $\boldsymbol{\beta}_0(\tau) = (\beta_{01}(\tau), \beta_{02}(\tau), \ldots, \beta_{0p}(\tau))^\top$ is the $p$-dimensional vector of quantile regression coefficients. Quantile regression models the conditional distribution of the outcome at specific quantiles given covariates, thus such model explicitly accounts for the effects of heterogeneity within data and shows its robustness against extreme outliers. Note that the classical median regression can be considered as a special case of quantile regression when $\tau = 0.5$ under the linearity assumption. Refer to Koenker (2005) and Koenker et al. (2017) for a comprehensive review on quantile regression.

Following the pioneering work by Koenker and Bassett (1978), numerous extensions and enhancements have been developed to address specificities in various types of datasets and tackle different issues. Under the framework of linear models, Portnoy and Koenker (1989, 1997) incorporated the adaptive L-estimators and subsequently proposed algorithms to significantly shorten computational time; Zhou and Portnoy (1998) focused on statistical inference for heteroscedastic linear models; Yu and Jones (1998) explored the kernel weighted local linear fitting. Moreover, He (1997) and Koenker and Xiao (2002) proposed estimation approaches to tackle issues arising from potential quantiles crossing, while He and Zhu (2003) and Bondell, Reich and Wang (2010) formulated various tests for quantile regression models. Koenker and Geling (2001) extended quantile regression to survival analysis, followed by many literatures on censored quantile regression including Portnoy (2003), Peng and Fine (2009), Wang and Wang (2009), Wu and Yin (2013), Gorfine, Goldberg and Ritov (2017), Yang, Narisetty and He (2018), and Jiang et al. (2020). Composite quantile regression has been explored by Zou and Yuan (2008), Kai, Li and Zou (2011), Wang, Li and He (2012), Wang and Li (2013). Furthermore, quantile regression has been extended to Bayesian approaches (e.g., Müller and Quintana, 2004; Dunson and Taylor, 2005; Chung and Dunson, 2009; Reich, Fuentes and Dunson, 2011; Feng, Chen and He, 2015; Qu and Yoon, 2015) as well as high dimensional settings (e.g., Kato, 2011; Wang, Wu and Li, 2012; He, Wang and Hong, 2013; Jiang, Wang and Bondell, 2013; Zheng, Peng and He, 2015; He et al., 2020).

Qu and Yoon (2015) presented estimation methods and asymptotic theory for the analysis of a nonparametric conditional quantile process. Chao, Volgushev and Cheng (2017) established weak convergence of quantile regression process (QRP) for nonparametric and semiparametric estimators. Volgushev et al. (2019) proposed a two-step estimation procedure to specifically address massive data by first using the divide-and-conquer algorithm to estimate the conditional quantile functions, and then aggregating results from each subset through quantile projection. However, these methods only considered quantile process pointwise/locally. Composite quantile regression simultaneously considers multiple quantiles and accounts for the effect of heterogeneity by allowing the intercept term to vary at different quantiles, providing a more holistic view than the locally concerned quantile regression that considered only a singleton. Such regression has been explored by Zou and Yuan (2008), Wang and Wang (2009), Kai, Li and Zou (2011), Wang, Li and He (2012), Wang and Li (2013). Zheng, Peng and He (2015) introduced a globally concerned quantile regression model that considered a union of multiple disjoint intervals and adaptive weight functions.

When model (1) is assumed for all $\tau \in (0, 1)$, such a globally concerned quantile regression with quantile regression function $\boldsymbol{\beta}_0(\cdot) = (\beta_{01}(\cdot), \beta_{02}(\cdot), \ldots, \beta_{0p}(\cdot))^\top$ defined on (0,1) suffices to characterize the conditional distribution of $Y$ given $\mathbf{X}$. To the best of our knowledge, there is limited globally concerned quantile method that provides a quantile regression process to address the heterogeneity in covariates for ordinary datasets. To fill this research gap, we propose a new globally concerned quantile regression framework with functional parameters, allowing regression coefficients to vary across different quantiles. We derive the asymptotic properties of the proposed estimator by first constructing a vector-valued Sobolev space, followed by a Bahadur representation of the estimator. In particular, different from the novel approaches developed by Shang and Cheng (2015) and Cheng and Shang (2015), we establish the vector-valued Bahadur representation of functional estimators based on a nonsmooth objective function, which brings both theoretical and computational challenges. The resampling methodology is implemented to obtain confidence intervals and conduct hypothesis testing. The proposed tests have three potential applications: (i) testing alternative model specifications; (ii) testing stochastic dominance; and (iii) testing the significance of treatment effects and heterogeneity.

The rest of the paper is organized as follows. Section 2 provides an estimation approach for unknown functional quantile regression coefficient. Section 3 presents some preliminaries about vector-valued reproducing kernel Hilbert space (VRKHS), the vector-valued functional Bahadur representation (VFBR), and the asymptotic properties of the proposed estimator. Section 4 discusses the inference procedure about functional coefficients with the resampling approach. Section 5 reports some results from simulation studies conducted for evaluating the proposed method. An application to a world happiness report dataset is provided in Section 6, and some concluding remarks are given in Section 7. All technical proofs are given in the online Supplementary Material.

# 2 Estimation Method

Suppose that the observed data consist of $n$ independent and identically distributed (i.i.d) replicates of $\left(Y, \mathbf{X}^\top\right)^\top$ from the globally-concerned quantile regression model, denoted by $\left\{\left(Y_i, \mathbf{X}_i^\top\right)^\top, i = 1, \ldots, n\right\}$. To estimate the unknown quantile regression function $\boldsymbol{\beta}_0(\cdot)$, the objective function can be written as

$$l_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\int_{\triangle}\rho_\tau\{Y_i - \mathbf{X}_i^\top\boldsymbol{\beta}(\tau)\}\, d\tau,$$

where $\triangle$ is a closed subset of $(0,1)$ being the quantile region of interest, $\rho_\tau(u) = u\{\tau - \mathbf{1}(u \leq 0)\}$ is the quantile check loss function, and $I(\cdot)$ is the indicator function. In the following, we assume that $\boldsymbol{\beta}_0(\cdot) \in \mathcal{H}^m(\triangle)$, with

$$\mathcal{H}^m(\triangle) = \{\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \beta_2(\cdot), \ldots, \beta_p(\cdot))^\top : \triangle \to \mathbb{R}^p | \beta_k^{(j)}(\cdot) \text{ is absolutely continuous for}$$
$$j = 0, 1, \ldots, m-1, \beta_k^{(m)}(\cdot) \in L_2(\triangle), k = 1, 2, \ldots, p\},$$

where $m \geq 2$ and is assumed to be known, $\beta_k^{(j)}(\cdot)$ is the $j$th derivative of $\beta_k(\cdot)$, and $L_2(\triangle)$ is the $L_2$ space defined in $\triangle$. Define $\boldsymbol{\beta}^{(j)}(\cdot) = (\beta_1^{(j)}(\cdot), \beta_2^{(j)}(\cdot), \ldots, \beta_p^{(j)}(\cdot))^\top, j = 1, 2, \ldots, m$ and $J(\boldsymbol{\beta}, \tilde{\boldsymbol{\beta}}) = \int_\triangle {\boldsymbol{\beta}^{(m)}}^\top(\tau)\tilde{\boldsymbol{\beta}}^{(m)}(\tau)\, d\tau$. The penalized objective function of $\boldsymbol{\beta}(\cdot)$ is defined as

$$\iota_{n,\lambda}(\boldsymbol{\beta}) = l_n(\boldsymbol{\beta}) + \frac{\lambda}{2} J(\boldsymbol{\beta}, \boldsymbol{\beta}),$$

where $J(\boldsymbol{\beta}, \boldsymbol{\beta})$ is the roughness penalty and $\lambda$ is the smoothing parameter converging to zero as $n \to \infty$. The estimator of $\boldsymbol{\beta}_0$ is defined as

$$\hat{\boldsymbol{\beta}}_{n,\lambda} = \arg \min_{\boldsymbol{\beta} \in \mathcal{H}^m(\triangle)} \iota_{n,\lambda}(\boldsymbol{\beta}).$$

The penalized estimator $\hat{\boldsymbol{\beta}}_{n,\lambda}$ is continuous, since the penalty of the derivatives plays the same role as that defined in Koenker, Ng and Portnoy (1994) to avoid the sharp oscillation. Then, we can obtain a smooth estimator of $\boldsymbol{\beta}_0(\cdot)$.

# 3 Asymptotic Properties

In this section, we introduce the vector-valued reproducing kernel Hilbert space (VRKHS) and present obtained main results including the convergence rate and asymptotic normality of the proposed estimator.

## 3.1 Vector-Valued Reproducing Kernel Hilbert Space

Before stating the main results, we first introduce some notation. Define $\iota_\lambda(\boldsymbol{\beta}) = E\{\iota_{n,\lambda}(\boldsymbol{\beta})\}$. Let $F_{Y|\mathbf{X}}$ and $f_{Y|\mathbf{X}}$ be the conditional distribution and density functions of $Y$ given $\mathbf{X}$, respectively. And let $f_{Y|\mathbf{X}}^{(1)}$ be the derivative of the density functions of $Y$ given $\mathbf{X}$. For ease of presentation, we introduce additional notation related to Fréchet derivatives. Let $\mathcal{S}_n(\boldsymbol{\beta})$ and $\mathcal{S}_{n,\lambda}(\boldsymbol{\beta})$ be the Fréchet derivatives of $l_n(\boldsymbol{\beta})$ and $\iota_{n,\lambda}(\boldsymbol{\beta})$, respectively. Similarly, let $\mathcal{S}(\boldsymbol{\beta})$ and $\mathcal{S}_\lambda(\boldsymbol{\beta})$ be the Fréchet derivatives of $l(\boldsymbol{\beta})$ and $\iota_\lambda(\boldsymbol{\beta})$, respectively. Let $D$ be the Fréchet derivative operator and $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3 \in \mathcal{H}^m(\triangle)$ be any direction. Define the derivative of the check loss function as $\dot{\rho}_\tau(u) = \tau - \mathbf{1}(u \leq 0)$. Then, we have

$$\mathcal{S}_n(\boldsymbol{\beta})\boldsymbol{\beta}_1 = \frac{1}{n}\sum_{i=1}^{n}\left\{\int_\triangle -\mathbf{X}_i^\top\boldsymbol{\beta}_1(\tau)\tau + \mathbf{X}_i^\top\boldsymbol{\beta}_1(\tau)\mathbf{1}\{Y_i \leq \mathbf{X}_i^\top\boldsymbol{\beta}(\tau)\}\, d\tau\right\},$$

$$\mathcal{S}_{n,\lambda}(\boldsymbol{\beta})\boldsymbol{\beta}_1 = \mathcal{S}_n(\boldsymbol{\beta})\boldsymbol{\beta}_1 + \lambda J(\boldsymbol{\beta}_1, \boldsymbol{\beta}),$$

$$\mathcal{S}(\boldsymbol{\beta})\boldsymbol{\beta}_1 = E\left[\int_\triangle -\tau\mathbf{X}^\top\boldsymbol{\beta}_1(\tau) + \mathbf{X}^\top\boldsymbol{\beta}_1(\tau)F_{Y|\mathbf{X}}\{\mathbf{X}^\top\boldsymbol{\beta}(\tau)\}\, d\tau\right],$$

$$\mathcal{S}_\lambda(\boldsymbol{\beta})\boldsymbol{\beta}_1 = \mathcal{S}(\boldsymbol{\beta})\boldsymbol{\beta}_1 + \lambda J(\boldsymbol{\beta}_1, \boldsymbol{\beta}),$$

$$D\mathcal{S}(\boldsymbol{\beta})\boldsymbol{\beta}_1\boldsymbol{\beta}_2 = \int_\triangle \boldsymbol{\beta}_1(\tau)^\top E_{\mathbf{X}}[f_{Y|\mathbf{X}}\{\mathbf{X}^\top\boldsymbol{\beta}(\tau)\}\mathbf{X}\mathbf{X}^\top]\boldsymbol{\beta}_2(\tau)\, d\tau,$$

$$D\mathcal{S}_\lambda(\boldsymbol{\beta})\boldsymbol{\beta}_1\boldsymbol{\beta}_2 = D\mathcal{S}(\boldsymbol{\beta})\boldsymbol{\beta}_1\boldsymbol{\beta}_2 + \lambda J(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2),$$

$$D^2\mathcal{S}(\boldsymbol{\beta})\boldsymbol{\beta}_1\boldsymbol{\beta}_2\boldsymbol{\beta}_3 = \int_\triangle \boldsymbol{\beta}_1(\tau)^\top E_{\mathbf{X}}[f_{Y|\mathbf{X}}^{(1)}\{\mathbf{X}^\top\boldsymbol{\beta}(\tau)\}\mathbf{X}\mathbf{X}^\top\boldsymbol{\beta}_2(\tau)\mathbf{X}^\top]\boldsymbol{\beta}_3(\tau)\, d\tau,$$

$$D^2\mathcal{S}_\lambda(\boldsymbol{\beta})\boldsymbol{\beta}_1\boldsymbol{\beta}_2\boldsymbol{\beta}_3 = D^2\mathcal{S}(\boldsymbol{\beta})\boldsymbol{\beta}_1\boldsymbol{\beta}_2\boldsymbol{\beta}_3.$$

Motivated by Cheng and Shang (2015) and Shang and Cheng (2015), we define the inner product for any $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathcal{H}^m(\triangle)$ as the second derivative of $\iota_\lambda(\boldsymbol{\beta})$ at $\boldsymbol{\beta}_0$:

$$\langle \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \rangle_\lambda = D\mathcal{S}_\lambda(\boldsymbol{\beta}_0)\boldsymbol{\beta}_1\boldsymbol{\beta}_2.$$

The corresponding norm is denoted as $||\cdot||_\lambda$. Then $\mathcal{H}^m(\triangle)$ is a VRKHS with an inner product $\langle \cdot, \cdot \rangle_\lambda$. For simplicity, define a linear nonnegative definite and self-adjoint operator $W_\lambda$ and a bilinear operator $V(\cdot, \cdot)$ in $\mathcal{H}^{(m)}$ as $\langle W_\lambda \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \rangle_\lambda = \lambda J(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ and $V(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \int_\triangle \boldsymbol{\beta}_1(\tau)^\top E_{\mathbf{X}}[\mathbf{X}\mathbf{X}^\top f_{Y|\mathbf{X}}\{\mathbf{X}^\top \boldsymbol{\beta}_0(\tau)\}]\boldsymbol{\beta}_2(\tau)d\tau$, respectively. Then,

$$\langle \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \rangle_\lambda = V(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) + \langle W_\lambda \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \rangle_\lambda.$$

Let $K(\cdot, \cdot)$ be the reproducing kernel matrix of $\mathcal{H}^m(\triangle)$ defined on $\triangle \times \triangle$, namely, for any $\boldsymbol{\beta}$ in $\mathcal{H}^m(\triangle)$, $\mathbf{c} \in \mathbb{R}^p$ and any $\tau_1, \tau_2$ in $\triangle$, we have $(K_{\tau_1}\mathbf{c})(\tau_2) = K(\tau_1, \tau_2)\mathbf{c}$ and $\langle K_{\tau_1} c, \boldsymbol{\beta} \rangle_\lambda = \boldsymbol{\beta}(\tau_1)^\top c$. In addition, denote $\mathbf{h}_j = (h_{1j}, h_{2j}, \ldots, h_{pj})^\top \in \mathcal{H}^m(\triangle)$ and $\gamma_j \in \mathbb{R}, j = 0, 1, 2, \ldots,$ as the eigenfunctions and eigenvalues of $\mathcal{H}^{(m)}(\triangle)$, respectively, which satisfies that $V(\mathbf{h}_i, \mathbf{h}_j) = \delta_{ij}$, $J(\mathbf{h}_i, \mathbf{h}_j) = \gamma_j \delta_{ij}$, where $\delta_{ij}$ is a Kronecker delta; that is, $\delta_{ij} = 1$ when $i = j$, and $\delta_{ij} = 0$ otherwise. Direct calculations yield that, for any $\boldsymbol{\beta} \in \mathcal{H}^m(\triangle)$ and $\tau \in \triangle$, we have $\|\boldsymbol{\beta}\|_\lambda^2 = \sum_{j=0}^\infty V(\boldsymbol{\beta}, \mathbf{h}_j)^2(1 + \lambda\gamma_j)$, $K_\tau(\cdot) = \sum_{j=0}^\infty \mathbf{h}_j(\tau)\mathbf{h}_j(\cdot)^\top/(1 + \lambda\gamma_j)$, and $(W_\lambda \mathbf{h}_j)(\cdot) = (\lambda\gamma_j)/(1 + \lambda\gamma_j)\mathbf{h}_j(\cdot)$.

To illustrate the existence of the eigensystem, we first decompose $E_{\mathbf{X}}[\mathbf{X}\mathbf{X}^\top f_{Y|\mathbf{X}}\{\mathbf{X}^\top \boldsymbol{\beta}_0(\tau)\}] = L(\tau)^\top D(\tau) L(\tau)$, where $D(\tau) = (D_j(\tau))_{j=1,2,\ldots,p}$ is a diagonal matrix, and $L(\tau)^\top L(\tau) = L(\tau)L(\tau)^\top = I$ with $I$ being the identity matrix. Denote $\tilde{\boldsymbol{\beta}} = L\boldsymbol{\beta} \equiv (\tilde{\beta}_1, \tilde{\beta}_2, \ldots, \tilde{\beta}_p)^\top$.

Then, we can rewrite

$$\begin{aligned}\langle \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \rangle_\lambda &= \int_\triangle \boldsymbol{\beta}_1(\tau)^\top E_{\mathbf{X}}[\mathbf{X}\mathbf{X}^\top f_{Y|\mathbf{X}}\{\mathbf{X}^\top \boldsymbol{\beta}_0(\tau)\}]\boldsymbol{\beta}_2(\tau)\, d\tau + \lambda J(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \\ &= \sum_{j=1}^p \left\{ \int_\triangle D_j(\tau)\tilde{\beta}_{1j}(\tau)\tilde{\beta}_{2j}(\tau)\, d\tau + \lambda J(\tilde{\beta}_{1j}, \tilde{\beta}_{2j}) \right\} \\ &\equiv \sum_{j=1}^p \left\{ V_j(\tilde{\beta}_j, \tilde{\beta}_j) + W_j(\tilde{\beta}_j, \tilde{\beta}_j) \right\}.\end{aligned}$$

It follows from Shang and Cheng (2013) that, there exists the eigensystem $\{\tilde{h}_{j,l}, \tilde{\rho}_{j,l}\}, l = 0, 1, 2, \ldots$, that $V_j(\tilde{h}_{j,k}, \tilde{h}_{j,l}) = \delta_{kl}, W_j(\tilde{h}_{j,k}, \tilde{h}_{j,l}) = \lambda \tilde{\rho}_{j,k}\delta_{kl}$, where $\tilde{\rho}_{j,k} \asymp k^{2m}$ and $\sup_{j,k} \|\tilde{h}_{j,k}\|_\infty < \infty$. Denote $\tilde{\mathbf{h}}_l = (\tilde{h}_{1,l}, \tilde{h}_{2,l}, \ldots, \tilde{h}_{p,l})^\top$, then $\mathbf{h}_l = L^\top \tilde{\mathbf{h}}_l / p$. Then it is not hard to obtain that $\gamma_l = \sum_{j=1}^p \tilde{\rho}_{j,l}/p$.

Direct calculations yield the following proposition.

**Proposition 3.1** *$D\mathcal{S}_\lambda(\beta_0) = id$, where id is the identity operator in $\mathcal{H}^m(\triangle)$.*

This result plays an important role in deriving the VFBR of the proposed estimator.

## 3.2 Main Results

We denote $h = \lambda^{1/(2m)}$, a positive definite matrix $A$ as $A > 0$, and two positive sequences $a_n$ and $b_n$ as $a_n \asymp b_n$ if $\lim_{n\to\infty}(a_n/b_n) = c > 0$. For any vector $\mathbf{z}$, we have $\mathbf{z}^{\otimes 2} = \mathbf{z}\mathbf{z}^\top$. To establish the asymptotic properties of the proposed estimator, we need the following regularity conditions:

(A1) $\Sigma(\tau) = L(\tau)^\top D(\tau) L(\tau)$ is positive definite for all $\tau \in \triangle$ and $E_{\mathbf{X}}(\mathbf{X}\mathbf{X}^\top)$ is positive definite.

(A2) There exist positive constants $c_1$ and $c_2$ that

$$P(c_1^{-1} < \inf_{\tau\in\triangle} |f_{Y|\mathbf{X}}\{\mathbf{X}^\top \boldsymbol{\beta}(\tau)\}| \le \sup_{\tau\in\triangle} |f_{Y|\mathbf{X}}\{\mathbf{X}^\top \boldsymbol{\beta}(\tau)\}| < c_1) = 1$$

and

$$P(c_2^{-1} < \inf_{\tau\in\triangle} |f^{(1)}_{Y|\mathbf{X}}\{\mathbf{X}^\top \boldsymbol{\beta}(\tau)\}| \le \sup_{\tau\in\triangle} |f^{(1)}_{Y|\mathbf{X}}\{\mathbf{X}^\top \boldsymbol{\beta}(\tau)\}| < c_2) = 1$$

for $\boldsymbol{\beta}$ in a closed neighborhood of $\boldsymbol{\beta}_0$.

(A3) There exists a sequence of real vector valued functions $\mathbf{h}_j \in \mathcal{H}^m(\triangle)$ with $j = 0, 1, 2, \ldots,$ satisfying $\sup_j \|\mathbf{h}_j\|_\infty < \infty$ and a nondecreasing real sequence $\gamma_j \asymp j^{2m}$, such that $V(\mathbf{h}_i, \mathbf{h}_j) = \delta_{ij}$, $J(\mathbf{h}_i, \mathbf{h}_j) = \gamma_j \delta_{ij}$, where $\delta_{ij}$ is a Kronecker delta; that is, $\delta_{ij} = 1$ when $i = j$, and $\delta_{ij} = 0$ otherwise. Furthermore, any $\boldsymbol{\beta} \in H^m(\triangle)$ admits the Fourier expansion $\boldsymbol{\beta} = \sum_{j=0}^\infty V(\boldsymbol{\beta}, \mathbf{h}_j)\mathbf{h}_j$ under the $\langle\cdot,\cdot\rangle_\lambda$.

**Remark 1** *Assumptions (A1) and (A2) are standard regularity conditions in the quantile regression literature. Assumption (A3) is similar to Assumption (A.3) in Shang and Cheng (2013).*

Before stating the VFBR, we give the convergence rate of the regularized estimator.

**Theorem 3.1** *(Convergence Rate). Suppose the Assumptions (A1)-(A3) hold. If* $h = o(1), nh^2 \to \infty$, *then* $\hat{\boldsymbol{\beta}}_{n,\lambda}$ *is the unique estimate for* $\boldsymbol{\beta}_0$ *and* $\|\hat{\boldsymbol{\beta}}_{n,\lambda} - \boldsymbol{\beta}_0\|_\lambda = O_p(r_n)$, *where* $r_n = (nh)^{-1/2} + h^m$.

**Theorem 3.2** *(VFBR). Suppose that Assumptions (A1)-(A3) hold. If* $h = o(1), nh^2 \to \infty$, *we have* $\|\hat{\boldsymbol{\beta}}_{n,\lambda} - \boldsymbol{\beta}_0 + \mathcal{S}_{n,\lambda}(\boldsymbol{\beta}_0)\|_\lambda = O_p(\alpha_n)$, *where* $\alpha_n = h^{-1/2} r_n^2$.

Theorem 3.2 indicates that the "bias" of estimator $\hat{\boldsymbol{\beta}}_{n,\lambda}$ can be approximated by $\mathcal{S}_{n,\lambda}(\boldsymbol{\beta}_0)$, an i.i.d sum of zero mean random processes. In fact, Theorem 3.2 is crucial to deriving the asymptotic normality of the proposed estimator.

**Theorem 3.3** *Suppose that Assumptions (A1)-(A3) hold. If* $h = o(1), nh^3 \to \infty, nh^{2m} \to 0$, *and*

$$0 < hE\left[\int_{\triangle} -K_{\tau_0}(\tau)\mathbf{X}_i\tau + K_{\tau_0}(\tau)\mathbf{X}_i\mathbf{1}\{Y_i \leq \mathbf{X}_i^\top \boldsymbol{\beta}_0(\tau)\}\, d\tau\right]^{\otimes 2} \to \sigma_{\tau_0} > 0 \text{ as } h \to 0,$$

*then for any* $\tau_0 \in \mathbb{I}$, *we have*

$$\sqrt{nh}\{\hat{\boldsymbol{\beta}}_{n,\lambda}(\tau_0) - \boldsymbol{\beta}_0(\tau_0)\} \xrightarrow{d} N(0, \sigma_{\tau_0}),$$

*where* $\xrightarrow{d}$ *means convergence in distribution.*

Theorem 3.3 represents that the bias of our regularized estimate can be approximated by the normal distribution. This lays the theoretical foundation for statistical inference about quantile regression.

In the proposed estimation method, a tuning parameter $\lambda$ is involved where $\lambda$ controls the degree of penalization on the roughness of estimator $\hat{\beta}_{n,\lambda}(\tau)$. For penalized nonparametric mean regression, Craven and Wahba (1978) proposed the practical and effective Generalized Cross-Validation (GCV) criterion for smoothing splines, which enjoys the theoretical optimality in the sense that the resulting estimate minimizes the true empirical loss function asymptotically. Recently, researchers have adapted classic GCV to modern nonparametric regressions and inferences, e.g., distributed GCV for divide-and-Conquer kernel ridge regression in Xu et al. (2018, 2019) and GCV for nonparametric testing procedure based on quantized samples in Liu et al. (2019). The proposed estimation targets for the quantile regression, which is different form the nonparametric mean regression in some essential aspects including the model setup and objective functions. Therefore, the classic GCV cannot be directly used, but it maybe adapted to the present new model with modifications. In our estimation, we apply the cross-validated log likelihood method (CVL) suggested by Verweij and Van Houwelingen (1993) for choosing $\lambda$ and leave the adaptation of GCV as future studies.

# 4 Inference Procedure

For inferences on $\boldsymbol{\beta}_0(\tau)$, we aim to construct confidence interval and to conduct hypothesis testing. However, the results of Theorems 3.2 and 3.3 indicate that the estimation of the covariance involves finding the unknown reproducing kernel function. This requires solving $q$ ordinary differential equation (ODE) functions, which is very difficult and time consuming and the solution is not stable, see Shang and Cheng (2013). In this section, we propose to use a resampling approach that generalizes the minimand perturbing technique proposed by Rao and Zhao (1992) and Jin et al. (2001) to the situation with functional estimands.

Let $\epsilon_1, \ldots, \epsilon_n$ be independent nonnegative random variables from a known distribution with mean 1 and variance 1, such as exponential(1); see Peng and Huang (2008). Then, we consider a stochastic perturbation of $\iota_{n,\lambda}(\boldsymbol{\beta})$, denoted as $\tilde{\iota}_{n,\lambda}(\boldsymbol{\beta})$:

$$\tilde{\iota}_{n,\lambda}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \epsilon_i \int_{\triangle} \rho_\tau\{Y_i - \mathbf{X}_i^\top \boldsymbol{\beta}(\tau)\}\, d\tau + \frac{\lambda}{2} J(\boldsymbol{\beta}, \boldsymbol{\beta}).$$

Define

$$\bar{\boldsymbol{\beta}}_{n,\lambda} = \arg \min_{\boldsymbol{\beta} \in \mathcal{H}^m(\triangle)} \iota_{n,\lambda}(\boldsymbol{\beta}).$$

With the data fixed at the observed values, we repeatedly generate the random variables $\{\epsilon_1, \ldots, \epsilon_n\}$ for $B$ times and obtain a large number of realizations of $\bar{\boldsymbol{\beta}}_{n,\lambda}(\tau)$, denoted by $\left\{\bar{\boldsymbol{\beta}}^{(r)}_{n,\lambda}(\tau)\right\}_{r=1}^{B}$.

**Theorem 4.1** *Assume that the conditions in Theorem 3 hold. Then*

$$\|\bar{\boldsymbol{\beta}}_{n,\lambda} - \boldsymbol{\beta}_0 + \tilde{\mathcal{S}}_{n,\lambda}(\boldsymbol{\beta}_0)\|_\lambda = O_p(\alpha_n),$$

*where $\tilde{\mathcal{S}}_{n,\lambda}(\boldsymbol{\beta}_0)$ is the first Fréchet derivative of $\tilde{\iota}_{n,\lambda}(\boldsymbol{\beta})$ at the true value, and*

$$\|\bar{\boldsymbol{\beta}}_{n,\lambda} - \hat{\boldsymbol{\beta}}_{n,\lambda} + \tilde{\mathcal{S}}_{n,\lambda}(\boldsymbol{\beta}_0) - \mathcal{S}_{n,\lambda}(\boldsymbol{\beta}_0)\|_\lambda = O_p(\alpha_n).$$

The proof of the first part of Theorem 4.1 is very similar to that of Theorem 3.2, while the second part can be directly obtained via the Triangle inequality of the inner product. Thus we omitted the proof of Theorem 4.1. Based on the linear approximation of the difference between $\bar{\boldsymbol{\beta}}_{n,\lambda}$ and $\hat{\boldsymbol{\beta}}_{n,\lambda}$, we can get Kolmogorov-Smirnov distance between $(nh)^{1/2}\left\{\bar{\boldsymbol{\beta}}_{n,\lambda}(\tau) - \hat{\boldsymbol{\beta}}_{n,\lambda}(\tau)\right\}$ and $(nh)^{1/2}\left\{\hat{\boldsymbol{\beta}}_{n,\lambda}(\tau) - \boldsymbol{\beta}_0(\tau)\right\}$.

**Theorem 4.2** *Assume that the conditions in Theorem 4.1 hold. Then, we have*

$$\sup_{\mathbf{v} \in \mathbb{R}^p, \tau \in \Delta} \left| P^*\left((nh)^{1/2}\left\{\bar{\boldsymbol{\beta}}_{n,\lambda}(\tau) - \hat{\boldsymbol{\beta}}_{n,\lambda}(\tau)\right\} \le \mathbf{v}\right) - P\left((nh)^{1/2}\left\{\hat{\boldsymbol{\beta}}_{n,\lambda}(\tau) - \boldsymbol{\beta}_0(\tau)\right\} \le \mathbf{v}\right)\right| \to 0$$

*with probability approaching one, where $P^*$ is the conditional distribution given the observed data.*

Theorem 4.2 shows that the conditional distribution of $(nh)^{1/2}\left\{\bar{\boldsymbol{\beta}}_{n,\lambda}(\tau)-\hat{\boldsymbol{\beta}}_{n,\lambda}(\tau)\right\}$ given the observed data is asymptotically the same as the unconditional distribution of

$$(nh)^{1/2}\left\{\hat{\boldsymbol{\beta}}_{n,\lambda}(\tau)-\boldsymbol{\beta}_0(\tau)\right\}$$

for all $\tau\in\Delta$. This result validates the resampling-based inference. Specifically, the covariance matrix of $\hat{\boldsymbol{\beta}}_{n,\lambda}$ can be estimated by the sample covariance matrix constructed from $\left\{\bar{\boldsymbol{\beta}}^r_{n,\lambda}(\tau)\right\}_{r=1}^B$. Estimating the distribution of $(nh)^{1/2}\left\{\hat{\boldsymbol{\beta}}_{n,\lambda}(\tau)-\boldsymbol{\beta}_0(\tau)\right\}$ based on the distribution of $(nh)^{1/2}\left\{\bar{\boldsymbol{\beta}}_{n,\lambda}(\tau)-\hat{\boldsymbol{\beta}}_{n,\lambda}(\tau)\right\}$ leads to the construction of $(1-\alpha)100\%$ confidence region for $\boldsymbol{\beta}_0(\tau)$.

Next, we consider testing whether $\boldsymbol{\beta}_{0,\mathcal{M}}(\tau)\equiv\mathbf{0}\quad\text{or}\quad\mathbf{c}$ for $\tau\in\Delta$, and $\mathcal{M}\subset\{1,2,\ldots,p\}$. In general, we test $H_0:\mathbf{H}(\tau)^\top\boldsymbol{\beta}_0(\tau)=H_0(\tau),\tau\in\Delta$. Motivated by Peng and Huang (2008), we use the integral test statistic for $H_0$. Define $\mathcal{T}_1=(nh)^{1/2}\int_\Delta\left\{\mathbf{H}(s)^\top\hat{\boldsymbol{\beta}}_{n,\lambda}(s)-H_0(s)\right\}\Lambda(s)\,ds$, and

$$\mathcal{T}_1^*=(nh)^{1/2}\int_\Delta\left\{\mathbf{H}(s)^\top\bar{\boldsymbol{\beta}}_{n,\lambda}(s)-\mathbf{H}(s)^\top\hat{\boldsymbol{\beta}}_{n,\lambda}(s)\right\}\Lambda(s)ds,$$

where $\Lambda(s)$ is a nonnegative weight function. Theorem 4.2 implies that the limiting conditional distribution of $\mathcal{T}_1^*$ given the observed data is the same as the limiting unconditional distribution of $\mathcal{T}_1$ under $H_0$. Thus, the rejection region of $H_0$ with significance level $\alpha$ is $\mathcal{T}_1>c_{1-\alpha/2}$ or $\mathcal{T}_1<c_{\alpha/2}$, where $c_{\alpha/2}$ and $c_{1-\alpha/2}$ are the $(\alpha/2)$-th and $(1-\alpha/2)$-th empirical percentiles of the $\mathcal{T}_1^*$. We can show that $\mathcal{T}_1$ is a consistent test when $\int_\Delta\mathbf{H}(s)^\top\boldsymbol{\beta}_0(s)\Lambda_0(s)ds\neq\int_\Delta H_0(s)\Lambda_0(s)ds$. As pointed out by Peng and Huang (2008), $\Lambda(\tau)$ can be appropriately chosen to accentuate the deviation from $H_0$ so as to ensure good power with realistic sample sizes.

# 5 Simulation Studies

Extensive simulation studies were conducted to evaluate the finite-sample performance of the proposed method. We generated data from

$$Y_i = \boldsymbol{\theta}_0^\top \boldsymbol{Z_i} + \boldsymbol{\gamma}_0^\top \boldsymbol{Z_i} \times \epsilon_i, \qquad i = 1, 2, ..., n,$$

where $\boldsymbol{\theta}_0$ and $\boldsymbol{\gamma}_0$ are $(p+1)$-dimensional vectors, $\boldsymbol{Z_i} = (1, \boldsymbol{X}_i^\top)^\top$, $\boldsymbol{X}_i$ follow a multivariate uniform distribution over $[0,1]^p$ with independent entries, and $\epsilon_i$ are independently and identically distributed (i.i.d.) random errors. Two distributions of $\epsilon$ were considered: (i) $\epsilon \sim 0.1 \times \mathcal{N}(0,1)$; and (ii) $\epsilon \sim 0.1 \times t(3)$. For a specific $\tau \in (0,1)$, the conditional quantile function of $Y$ given $\boldsymbol{X} = \boldsymbol{x}$ is

$$Q_{Y|\boldsymbol{X}}(\tau|\boldsymbol{x}) = (1, \boldsymbol{x}^\top)^\top \boldsymbol{\theta}_0 + (1, \boldsymbol{x}^\top)^\top \boldsymbol{\gamma}_0 \times F_\epsilon^{-1}(\tau),$$

where $F_\epsilon^{-1}(\cdot)$ is the inverse function of the cumulative distribution function of $\epsilon$. For a given $\tau \in (0,1)$, $\boldsymbol{\beta}_0(\tau) = (\beta_{01}(\tau), \ldots, \beta_{0p}(\tau))^\top$ with the $j$-th component $\beta_{0j}(\tau) = \theta_{0j} + \gamma_{0j} \times F_\epsilon^{-1}(\tau)$ for $j = 1, \ldots, p$. We investigated two settings for parameters: (i) $\boldsymbol{\theta}_0 = (1, -2)^\top$, $\boldsymbol{\gamma}_0 = (1, 0.5)^\top$; and (ii) $\boldsymbol{\theta}_0 = (1, -2, 3, -4, 5)^\top$, $\boldsymbol{\gamma}_0 = (1, 0.5, 1, 0.5, 1)^\top$.

We estimated $\boldsymbol{\beta}_0(\tau)$ in a component-wise manner by Fourier basis functions with $L_2$ penalty on the second derivatives. The regularization parameter $\lambda$ was chosen by the cross-validated log likelihood method (CVL) suggested by Verweij and Van Houwelingen (1993) on even grids of $[0,1]$ with 1001 lattice points. Simulation results from the proposed method were compared to those from the divide-and-conquer and quantile projection method (DAC-QP) introduced by Volgushev, Chao and Cheng (2019). For each setting, we set the sample size $n = 400$ and repeated the Monte Carlo simulation 500 times. Moreover, the estimated standard error (ESE) and coverage probability (CP) of the proposed estimator were calculated based on 300 bootstrap samples.

Tables 1-3 and Figures 1-5 summarize numerical results from different settings. Figures 1 and 2 plot the estimation results of the proposed estimator under normally distributed errors in case (i). It can be observed that the point estimates are in close proximity of the true values, the ESE and sample standard error (SE) are similar, and CP fluctuates near 95% for $\tau \in [0.1, 0.9]$. Figures 3 and 4 display performance comparisons of the proposed method to the DAC-QP method. Evidently, the proposed estimator yields more reasonable CP and deviates less from the true values as $\tau$ approaches 0 or 1. Moreover, Tables 1 and 2 show the estimation results at quantile levels $\tau = 0.3, 0.5, 0.9$, including the average bias, SE, and CP. In contrast to the DAC-QP method, the proposed method yields comparable bias and SE while achieving more reasonable CP in all simulation scenarios.

Next, we conduct hypothesis testing to evaluate the empirical size and power associated with the proposed tests. Under parameter setting (i), the true quantile regression coefficient is

$$\boldsymbol{\beta}_0(\tau) = \begin{pmatrix} \beta_{01}(\tau) \\ \\ \beta_{02}(\tau) \end{pmatrix} = \begin{pmatrix} 1 + F_\epsilon^{-1}(\tau) \\ \\ -2 + 0.5F_\epsilon^{-1}(\tau) \end{pmatrix}, \text{ for } \tau \in (0, 1).$$

Consider the null hypothesis $H_0: \;\; \mathbf{H}(\tau)^\top \boldsymbol{\beta}_0(\tau) = H_0(\tau, c)$, where $\mathbf{H}(\tau) = (1, 0)^\top$, $H_0(\tau, c) = 1 + c + F_\epsilon^{-1}(\tau)$ for $\tau \in \Delta$ and $c \in [-0.5, 0.5]$. We considered a weight function $\Lambda_0(\tau) \equiv 1$ for $\tau \in \Delta$, and restricted $\Delta = [0.05, 0.95]$ to exclude the extreme quantiles. Based on theoretical results derived in Section 4, we computed the test statistic $\mathcal{T}_1$ and estimated its $(\alpha/2)$-th quantile, denoted by $c_{\alpha/2}$, and $(1-\alpha/2)$-th quantile, denoted by $c_{1-\alpha/2}$. The quantiles $c_{\alpha/2}$ and $c_{1-\alpha/2}$ are taken as the $(\alpha/2)$-th and $(1-\alpha/2)$-th empirical percentiles of 300 resampled $\mathcal{T}_1^*$. We set $\alpha = 0.05$ and reject $H_0$ if $\mathcal{T}_1 < c_{\alpha/2}$ or $\mathcal{T}_1 > c_{1-\alpha/2}$. For different values of $c$, we calculated the empirical probability for rejecting $H_0$ over 500 replications, which is interpreted as the empirical size of the test when $c = 0$ and the empirical power otherwise.

Under parameter setting (ii), the true quantile regression coefficient is

$$\boldsymbol{\beta}_0(\tau) = \begin{pmatrix} \beta_{01}(\tau) \\ \beta_{02}(\tau) \\ \beta_{03}(\tau) \\ \beta_{04}(\tau) \\ \beta_{05}(\tau) \end{pmatrix} = \begin{pmatrix} 1 + F_\epsilon^{-1}(\tau) \\ -2 + 0.5F_\epsilon^{-1}(\tau) \\ 3 + F_\epsilon^{-1}(\tau) \\ -4 + 0.5F_\epsilon^{-1}(\tau) \\ 5 + F_\epsilon^{-1}(\tau) \end{pmatrix}, \quad \text{for } \tau \in (0, 1).$$

Again, consider the null hypothesis: $\mathbf{H}(\tau)^\top \boldsymbol{\beta}_0(\tau) = H_0(\tau, c)$, where $\mathbf{H}(\tau) = (0, 0, 0, 0, 1)^\top$ and $H_0(\tau, c) = 5 + c + F_\epsilon^{-1}(\tau)$ for $\tau \in \Delta$ and $c \in [-0.5, 0.5]$. The empirical size and power were calculated in the same way as setting (i).

Figure 5 and Table 3 present the simulation results for hypothesis testing. It can be observed that when $c = 0$, the empirical size is close to 5% under all scenarios, and the empirical power rapidly approaches 1 as $c$ deviates away from 0. These demonstrate a good performance of the proposed test.

# 6 Application

We applied the proposed QRP method to analyze two datasets including the Taipei real estate valuation dataset and the world happiness report data.

## 6.1 Application to Real Estate Dataset

The Taipei real estate dataset included housing data from two districts in Taipei City and two districts in New Taipei City during the period from June 2012 to May 2013 and is publicly available from the Ministry of Interior of Taiwan (Yeh and Hsu, 2018). The study surveyed 414 properties and reported housing prices as 10,000 NT$/3.3 $\mathrm{m}^2$. Six appraisal factors were chosen as independent variables for this analysis, including transaction date, distance to the nearest MRT station in meters, number of convenience stores within walking distance, house age in years, and geographic coordinates (latitude and longitude). In the estimation procedure, the regularization parameter $\lambda$ was chosen by the CVL method (Verweij and Van Houwelingen, 1993) on even grids of $[0, 1]$ with 1001 lattice points and the 95% confidence intervals were constructed based on 300 bootstrapping samples. Figure 6 shows the point estimates and 95% confidence bands of the intercept and six covariate effects on housing price. Table 4 summarizes estimated covariate effects and standard errors at quantiles $\tau = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ for the six covariates, where all the covariates were shown to be significant at 95% confidence level.

It can be seen that the house age related to the living quality of the indoor living circle negatively affects the price of the house, while for expensive houses the negative effect of house age on the price is lighter than that of cheap houses. Besides, the closer the distance to the nearest MRT station, the higher the house price. And the negative effect of distance to MRT on the price is increasing as the house price increases. The convenience stores are all over the urban and country and bring people living convenience. Therefore, the number of convenience stores in the living circle on foot affects positively on the house price across all price levels of the houses. The latitude and longitude of house determines the distance to the downtown and hence may affect the time and money cost to go shopping and to office. The estimation results for latitude and longitude of house reflects that fact that the north-east area is closer to the downtown of Taipei city, and its house price per unit area is much higher than that in the south-west area. The transaction date is positively related to the house price across all quantile levels, which reflects the fact that the price index of New Taipei City has a very big rising through the studied period. The importance of the transaction date reflects that the house price is not only affected by the house but also the market.

## 6.2 Application to World Happiness Report Data

The World Happiness Report 2019 was released by the United Nations for an event celebrating the International Day of Happiness. The report surveyed people from 156 countries on various aspects related to happiness levels. For this application, we chose the happiness score from the Gallup World Poll as the response variable. The score was based on the Cantril ladder, which asked respondents to rank their current lives on a scale from 0 to 10, with 0 as the worst possible life and 10 as the best possible life. The explanatory variables include economic production, social support, life expectancy, freedom, perceptions of corruption, and generosity, contributions to making life evaluations higher in each country. In the estimation procedure, the regularization parameter $\lambda$ was chosen by the CVL method (Verweij and Van Houwelingen, 1993) on even grids of $[0, 1]$ with 1001 lattice points. The 95% confidence bands were obtained based on 300 bootstrap samples. Figure 7 plots the point estimates and the 95% confidence bands of the six covariate effects. A decreasing trend on the effect from social support and an increasing trend on the effect from freedom are observed from lower to higher quantiles.

# 7 Concluding Remarks

To make statistical inference on functional regression coefficients in the globally concerned quantile regression model, we have developed a novel estimation approach through integrating quantile regression loss function and smoothing techniques with penalization. It seems that we are the first to design a globally concerned quantile method based on a penalized integrated quantile regression loss function. The nonsmoothness of the loss function poses new challenges for computation and establishment of theoretical results. The asymptotic properties have been derived through the vector-valued functional Bahadur representation of the proposed estimators in a reproducing kernel Hilbert space. As demonstrated in our simulation studies and real data analysis, the proposed approaches perform well.

Further research is to extend the proposed methods to semiparametric quantile regression with survival data such as right-censored data and interval-censored data.

## Supplementary Materials

All proofs are provided in the Supplementary Material.

## References


Bondell, H. D., Reich, B. J. and Wang, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika* **97**, 825–838.

Briollais, L. and Durrieu, G. (2014). Application of quantile regression to recent genetic and -omic studies. *Hum. Genet.* **133**, 951–966.

Cade, B. S. and Noon, B. R. (2003). A gentle introduction to quantile regression for ecologists. *Front. Ecol. Environ.* **1**, 412–420.

Chao, S.-K., Volgushev, S. and Cheng, G. (2017). Quantile processes for semi and nonparametric regression. *Electron. J. Stat.* **11**, 3272–3331.

Cheng, G. and Shang, Z. (2015). Joint asymptotics for semi-nonparametric regression models with partially linear structure. *Ann. Statist.* **43**, 1351–1390.

Chung, Y. and Dunson, D. B. (2009). Nonparametric Bayes conditional distribution modeling with variable selection. *J. Am. Stat. Assoc.* **104**, 1646–1660.

Craven, P. and Wahba, G. (1978). Smoothing noisy data with spline functions. *Numerische mathematik.* **31**, 377–403.

Dunson, D. B. and Taylor, J. A. (2005). Approximate Bayesian inference for quantiles. *J. Nonparametr. Statist.* **17**, 385–400.

Feng, Y., Chen, Y. and He, X. (2015). Bayesian quantile regression with approximate likelihood. *Bernoulli* **21**, 832–850.

Gorfine, M., Goldberg, Y. and Ritov, Y. (2017). A quantile regression model for failure-time data with time-dependent covariates. *Biostatistics* **18**, 132–146.

He, X. (1997). Quantile curves without crossing. *Am. Statist.* **51**, 186–192.

He, X., Pan, X., Tan, K. and Zhou, W. (2020). Smoothed Quantile Regression with Large-Scale Inference.

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

He, X. and Zhu, L. X. (2003). A lack-of-fit test for quantile regression. *J. Am. Stat. Assoc.* **98**, 1013–1022.

Jiang, F., Cheng, Q., Yin, G. and Shen, H. (2020). Functional censored quantile regression. *J. Am. Stat. Assoc.* **115**, 931–944.

Jiang, L., Wang, H. J. and Bondell H. D. (2013). Interquantile shrinkage in regression models. *J. Comput. Graph. Stat.* **22**, 970–986.

Jin, Z., Ying, Z. and Wei, L. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88**, 381–390.

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

Kato, K. (2011). Group Lasso for high dimensional sparse quantile regression models. arXiv preprint arXiv: 1103.1458

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

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

Koenker, R., Chernozhukov, V., He, X. and Peng, L. (Eds.) (2017). Handbook of Quantile Regression. Boca Raton, Florida: Chapman and Hall/CRC.

Koenker, R. and Geling, O. (2001). Reappraising medfly longevity: A quantile regression survival analysis. *J. Am. Stat. Assoc.* **96**, 458–468.

Koenker, R. and Hallock, K. F. (2001). Quantile regression. *J. Econ. Perspect* **15**, 143–156.

Koenker, R., Ng, P. and Portnoy, S. (1994). Quantile smoothing splines. *Biometrika* **81**(4), 673-680.

Koenker, R. and Xiao, Z. (2002). Inference on the quantile regression process. *Econometrica* **70**, 1583–1612.

Liu, R., Xu, G. and Shang, Z. (2019). Optimal Nonparametric Inference under Quantization. *arXiv preprint arXiv:1901.08571.*

Müller, P. and Quintana, F. A. (2004). Nonparametric Bayesian data analysis. *Stat. Sci.* **19**, 95–110.

Peng, L. and Fine, J. P. (2009). Competing risks quantile regression. *J. Am. Stat. Assoc.* **104**, 1440–1453.

Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *J. Am. Stat. Assoc.* **103**, 637–649.

Portnoy, S. (2003). Censored regression quantiles. *J. Am. Stat. Assoc.* **98**, 1001–1012.

Portnoy, S. and Koenker, R. (1989). Adaptive L-estimation for linear models. *Ann. Statist.* **40**, 1714–1736.

Portnoy, S. and Koenker, R. (1997). The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators (with discussion). *Stat. Sci.* **12**, 279–300.

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

Rao, C. and Zhao, L. (1992). Approximation to the distribution of M-estimates in linear models by randomly weighted bootstrap. *Sankhya* **54**, 323–331

Reich, B. J., Fuentes, M. and Dunson, D. B. (2011). Bayesian spatial quantile regression. *J. Am. Stat. Assoc. 106*, 6–20.

Shang, Z. and Cheng, G. (2013). Local and global asymptotic inference in smoothing spline models. *Ann. Statist.* **41**(5), 2608–2638.

Shang, Z. and Cheng, G. (2015). Nonparametric inference in generalized functional linear models. *Ann. Statist.* **43**, 1742–1773.

Verweij, P. J. and Van Houwelingen, H. C. (1993). Cross-validation in survival analysis. *Stat. Med.* **12**, 2305–2314.

Volgushev, S., Chao, S.-K. and Cheng, G. (2019). Distributed inference for quantile regression processes. *Ann. Statist.* **47**, 1634–1662.

Wang, H. J. and Li, D. (2013). Estimation of extreme conditional quantiles through power transformation. *J. Am. Stat. Assoc.* **108**, 1062–1074.

Wang, H. J., Li, D. and He, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *J. Am. Stat. Assoc.* **107**, 1453–1464.

Wang, H. J. and Wang, L. (2009). Locally weighted censored quantile regression. *J. Am. Stat. Assoc.* **104**, 1117–1128.

Wang, L., Wu, Y. and Li, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *J. Am. Stat. Assoc.* **107**, 214–222.

Wu, Y. and Yin, G. (2013). Cure rate quantile regression for censored data with a survival fraction. *J. Am. Stat. Assoc.* **108**, 1517–1531.

Xu, G., Shang, Z. and Cheng, G. (2018). Optimal tuning for divide-and-conquer kernel ridge regression with massive data. *Proceedings of the ICML.***80**, 5483–5491.

Xu, G., Shang, Z. and Cheng, G. (2019). Distributed Generalized Cross-Validation for Divide-and-Conquer Kernel Ridge Regression and Its Asymptotic Optimality. *Journal of Computational and Graphical Statistics.***28**, 891–908.

Yang, X., Narisetty, N. N. and He, X. (2018). A new approach to censored quantile regression estimation. *J. Comput. Graph. Stat.* **27**, 417–425.

Yeh, I. C. and Hsu, T. K. (2018). Building real estate valuation models with comparative approach through case-based reasoning. *Applied Soft Computing.* **65**, 260–271.

Yu, K. and Jones, M. (1998). Local linear quantile regression. *J. Am. Stat. Assoc.* **93**, 228–237.

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

Zhou, K. Q. and Portnoy, S. (1998). Statistical inference on heteroscedastic models based on regression quantiles. *J. Nonparametr. Statist.* **9**, 239–260.

Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *Ann. Statist.* **36**, 1108–1126.

Table 1: Comparison results of our method and DAC-QP with $\boldsymbol{\theta}_0 = (1, -2)^\top$, $\boldsymbol{\gamma}_0 = (1, 0.5)^\top$. For each component of the estimate, the averaged bias (Bias), the sample standard error (SE), and the coverage probability (CP) of 95% CI are presented. The averaged statistics are calculated over 500 replications.

| Error | $\hat{\boldsymbol{\beta}}(\tau)$ | Method | $\tau = 0.3$ | | | $\tau = 0.5$ | | | $\tau = 0.9$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Bias | SE | CP | Bias | SE | CP | Bias | SE | CP |
| $0.1 \times \mathcal{N}(0,1)$ | $\hat{\beta}_1(\tau)$ | Proposed | 0.012 | 0.015 | 0.956 | 0.011 | 0.014 | 0.948 | 0.016 | 0.019 | 0.964 |
| | | DAC-QP | 0.012 | 0.015 | 0.918 | 0.011 | 0.014 | 0.944 | 0.015 | 0.019 | 0.928 |
| | $\hat{\beta}_2(\tau)$ | Proposed | 0.023 | 0.028 | 0.968 | 0.021 | 0.027 | 0.964 | 0.030 | 0.038 | 0.962 |
| | | DAC-QP | 0.022 | 0.027 | 0.936 | 0.021 | 0.026 | 0.944 | 0.029 | 0.037 | 0.928 |
| $0.1 \times t(3)$ | $\hat{\beta}_1(\tau)$ | Proposed | 0.015 | 0.017 | 0.932 | 0.013 | 0.016 | 0.948 | 0.044 | 0.033 | 0.914 |
| | | DAC-QP | 0.016 | 0.019 | 0.922 | 0.013 | 0.016 | 0.934 | 0.063 | 0.037 | 0.628 |
| | $\hat{\beta}_2(\tau)$ | Proposed | 0.027 | 0.034 | 0.946 | 0.025 | 0.031 | 0.940 | 0.054 | 0.065 | 0.944 |
| | | DAC-QP | 0.029 | 0.036 | 0.936 | 0.025 | 0.031 | 0.954 | 0.059 | 0.072 | 0.918 |

Notes: DAC-QP stands for the method in Volgushev, Chao and Cheng (2019).

Table 2: Comparison results of our method and DAC-QP with $\boldsymbol{\theta}_0 = (1, -2, 3, -4, 5)^\top$, $\boldsymbol{\gamma}_0 = (1, 0.5, 1, 0.5, 1)^\top$. For each component of the estimate, the averaged bias (Bias), the sample standard error (SE), and the coverage probability (CP) of 95% CI are presented. The averaged statistics are calculated over 500 replications.

| Error | $\hat{\boldsymbol{\beta}}(\tau)$ | Method | $\tau = 0.3$ | | | $\tau = 0.5$ | | | $\tau = 0.9$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Bias | SE | CP | Bias | SE | CP | Bias | SE | CP |
| $0.1 \times \mathcal{N}(0,1)$ | $\hat{\beta}_1(\tau)$ | Proposed | 0.040 | 0.051 | 0.952 | 0.041 | 0.051 | 0.966 | 0.058 | 0.071 | 0.954 |
| | | DAC-QP | 0.048 | 0.060 | 0.924 | 0.044 | 0.055 | 0.934 | 0.062 | 0.077 | 0.948 |
| | $\hat{\beta}_2(\tau)$ | Proposed | 0.043 | 0.054 | 0.968 | 0.043 | 0.053 | 0.960 | 0.059 | 0.072 | 0.958 |
| | | DAC-QP | 0.048 | 0.059 | 0.940 | 0.045 | 0.056 | 0.936 | 0.068 | 0.081 | 0.926 |
| | $\hat{\beta}_3(\tau)$ | Proposed | 0.043 | 0.055 | 0.956 | 0.045 | 0.055 | 0.956 | 0.060 | 0.075 | 0.954 |
| | | DAC-QP | 0.051 | 0.065 | 0.930 | 0.047 | 0.059 | 0.922 | 0.074 | 0.083 | 0.884 |
| | $\hat{\beta}_4(\tau)$ | Proposed | 0.046 | 0.058 | 0.960 | 0.042 | 0.054 | 0.962 | 0.062 | 0.077 | 0.950 |
| | | DAC-QP | 0.051 | 0.064 | 0.924 | 0.047 | 0.059 | 0.934 | 0.067 | 0.084 | 0.914 |
| | $\hat{\beta}_5(\tau)$ | Proposed | 0.044 | 0.055 | 0.962 | 0.042 | 0.054 | 0.954 | 0.062 | 0.077 | 0.946 |
| | | DAC-QP | 0.048 | 0.060 | 0.946 | 0.047 | 0.059 | 0.914 | 0.078 | 0.083 | 0.894 |
| $0.1 \times t(3)$ | $\hat{\beta}_1(\tau)$ | Proposed | 0.049 | 0.060 | 0.954 | 0.043 | 0.054 | 0.962 | 0.095 | 0.119 | 0.946 |
| | | DAC-QP | 0.070 | 0.082 | 0.914 | 0.056 | 0.071 | 0.932 | 0.170 | 0.168 | 0.818 |
| | $\hat{\beta}_2(\tau)$ | Proposed | 0.051 | 0.063 | 0.952 | 0.047 | 0.058 | 0.966 | 0.107 | 0.133 | 0.954 |
| | | DAC-QP | 0.068 | 0.087 | 0.920 | 0.059 | 0.073 | 0.942 | 0.121 | 0.156 | 0.916 |
| | $\hat{\beta}_3(\tau)$ | Proposed | 0.053 | 0.067 | 0.962 | 0.047 | 0.058 | 0.966 | 0.108 | 0.120 | 0.960 |
| | | DAC-QP | 0.066 | 0.081 | 0.958 | 0.058 | 0.072 | 0.940 | 0.117 | 0.150 | 0.930 |
| | $\hat{\beta}_4(\tau)$ | Proposed | 0.051 | 0.064 | 0.952 | 0.045 | 0.056 | 0.968 | 0.104 | 0.130 | 0.964 |
| | | DAC-QP | 0.066 | 0.083 | 0.926 | 0.059 | 0.073 | 0.934 | 0.119 | 0.152 | 0.930 |
| | $\hat{\beta}_5(\tau)$ | Proposed | 0.051 | 0.063 | 0.962 | 0.046 | 0.059 | 0.964 | 0.109 | 0.123 | 0.956 |
| | | DAC-QP | 0.068 | 0.086 | 0.938 | 0.058 | 0.071 | 0.954 | 0.119 | 0.157 | 0.904 |



Notes: DAC-QP stands for the method in Volgushev, Chao and Cheng (2019).

Table 3: Hypothesis testing results for different settings of $(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)$ and error distributions. Empirical sizes and powers with different $c$ over 500 replications with $\alpha = 0.05$ are presented.

| $(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)$ | $\epsilon$ | Size | Power | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $c = 0$ | $c = -0.3$ | $c = -0.2$ | $c = -0.1$ | $c = -0.05$ | $c = 0.05$ | $c = 0.1$ | $c = 0.2$ | $c = 0.3$ |
| (i) | $0.1 \times \mathcal{N}(0,1)$ | 0.050 | 1.000 | 1.000 | 1.000 | 0.994 | 0.974 | 1.000 | 1.000 | 1.000 |
| | $0.1 \times t(3)$ | 0.052 | 1.000 | 1.000 | 1.000 | 0.526 | 0.322 | 0.922 | 1.000 | 1.000 |
| (ii) | $0.1 \times \mathcal{N}(0,1)$ | 0.042 | 1.000 | 0.992 | 0.558 | 0.164 | 0.222 | 0.650 | 0.996 | 1.000 |
| | $0.1 \times t(3)$ | 0.054 | 0.990 | 0.796 | 0.252 | 0.076 | 0.178 | 0.410 | 0.908 | 1.000 |

Table 4: Estimates of the regression quantiles and standard errors (in parenthesis) at $\tau = 0.1, 0.2, \ldots, 0.9$ for the 6 predictors in the analysis of the real estate valuation dataset.

| $\tau$ | Transaction date | House age | Distance to MRT | Number of convenience stores | Latitute | Longitute |
|---|---|---|---|---|---|---|
| 0.1 | 1.6691(0.0088) | -0.3180(0.0347) | -0.0025(0.0006) | 1.4223(0.0716) | 200.6313(0.0020) | 108.2600(0.0094) |
| 0.2 | 1.7281(0.0124) | -0.2977(0.0352) | -0.0034(0.0006) | 1.3329(0.0970) | 185.2732(0.0022) | 75.6894(0.0105) |
| 0.3 | 0.4697(0.0150) | -0.3104(0.0358) | -0.0038(0.0005) | 1.2388(0.1137) | 198.0478(0.0022) | 27.8422(0.0100) |
| 0.4 | 0.8571(0.0164) | -0.3105(0.0354) | -0.0042(0.0005) | 1.2150(0.1224) | 201.2377(0.0022) | 8.0092(0.0102) |
| 0.5 | 2.8466(0.0168) | -0.3313(0.0357) | -0.0043(0.0005) | 1.1462(0.1254) | 236.4902(0.0023) | -18.5732(0.0106) |
| 0.6 | 4.3448(0.0165) | -0.3518(0.0374) | -0.0043(0.0004) | 1.0942(0.1249) | 295.1654(0.0023) | -58.3875(0.0106) |
| 0.7 | 5.2004(0.0155) | -0.3470(0.0410) | -0.0044(0.0004) | 1.0776(0.1208) | 326.2590(0.0022) | -91.4226(0.0103) |
| 0.8 | 8.4235(0.0133) | -0.3439(0.0460) | -0.0048(0.0003) | 0.9317(0.1100) | 289.5915(0.0023) | -97.4307(0.0106) |
| 0.9 | 7.6517(0.0101) | -0.3146(0.0669) | -0.0052(0.0003) | 1.0540(0.0916) | 265.0590(0.0022) | -89.0183(0.0102) |

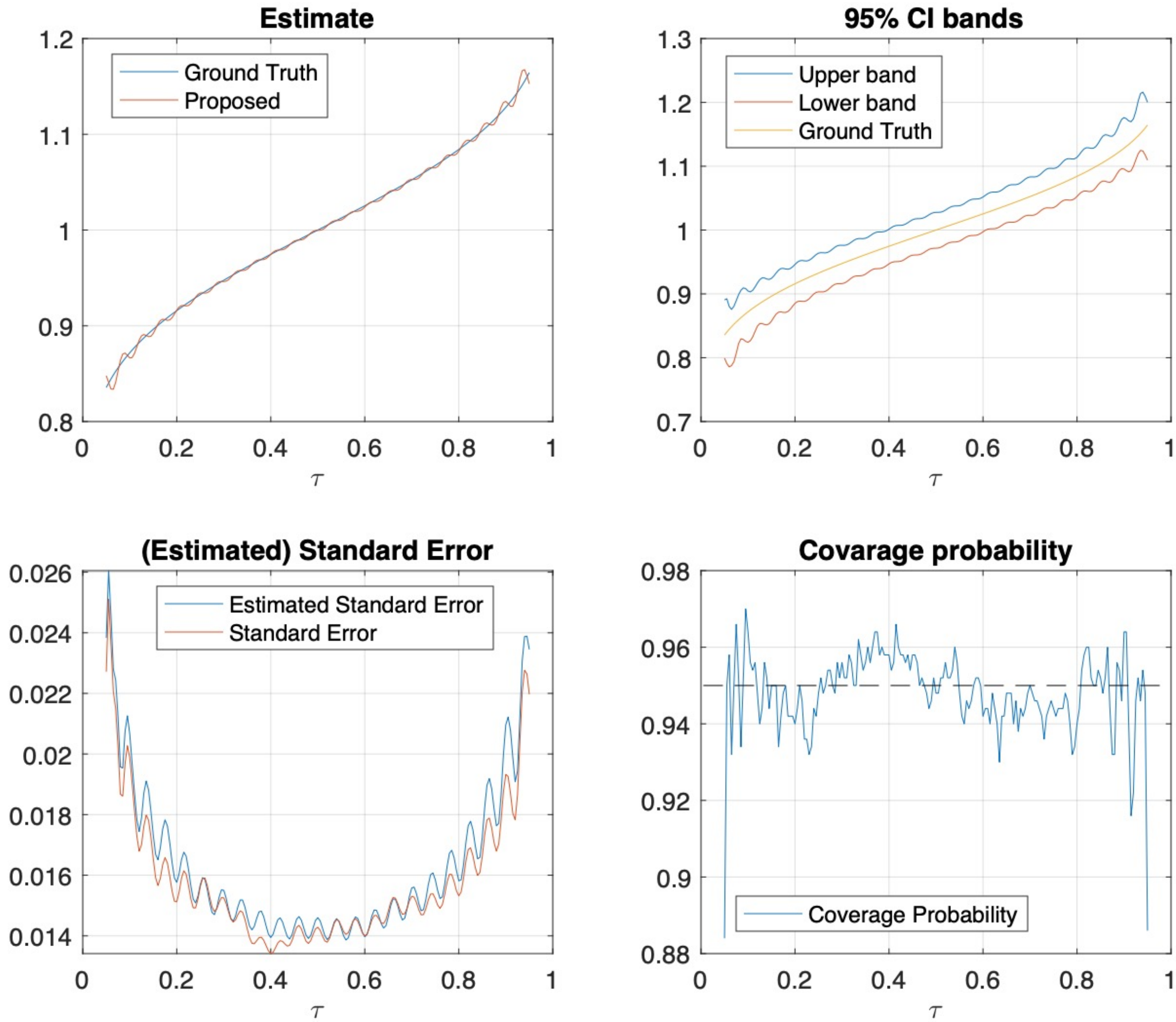


Figure 1: Estimation of $\beta_{01}(\tau)$ for our proposed method with $\boldsymbol{\theta}_0 = (1, -2)^\top,\ \boldsymbol{\gamma}_0 = (1, 0.5)^\top$. Upper left graph: the true $\beta_{01}(\tau)$ versus the averaged $\hat{\beta}_{n1,\lambda}(\tau)$. Upper right graph: the averaged estimated 95% CI bands. Lower left graph: pointwise standard errors versus the averaged estimated standard errors. Lower right graph: pointwise coverage probabilities, the horizontal dashed line (- -) is a reference line with height 0.95. All averaged statistics are calculated over 500 replications.

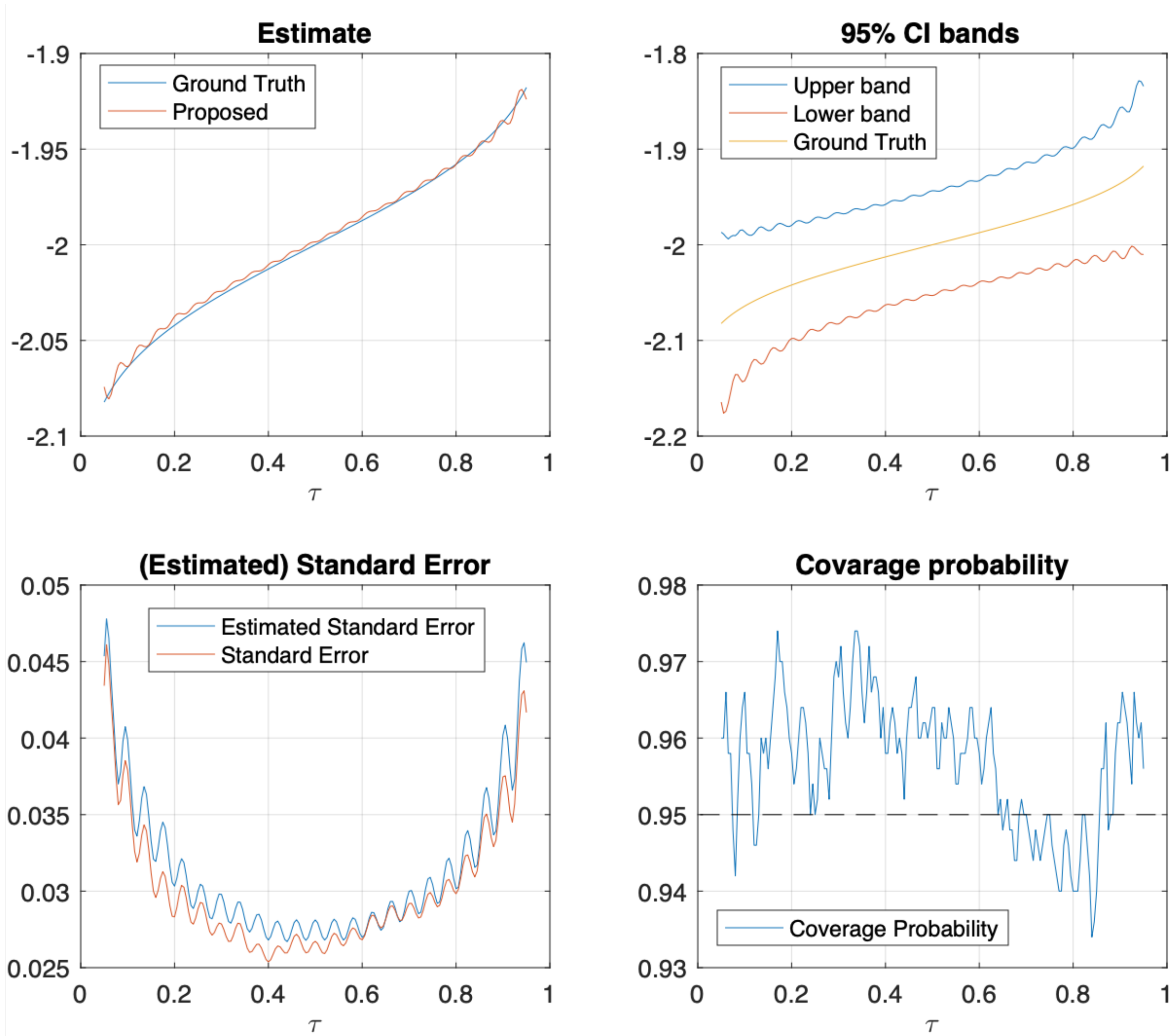


Figure 2: Estimation of $\beta_{02}(\tau)$ with our proposed method and $\boldsymbol{\theta}_0 = (1, -2)^\top$, $\boldsymbol{\gamma}_0 = (1, 0.5)^\top$. Upper left graph: the true $\beta_{02}(\tau)$ versus the averaged $\hat{\beta}_{n2,\lambda}(\tau)$. Upper right graph: the averaged estimated 95% CI bands. Lower left graph: pointwise standard errors versus the averaged estimated standard errors. Lower right graph: pointwise coverage probabilities, the horizontal dashed line (- -) is a reference line with height 0.95. All averaged statistics are calculated over 500 replications.

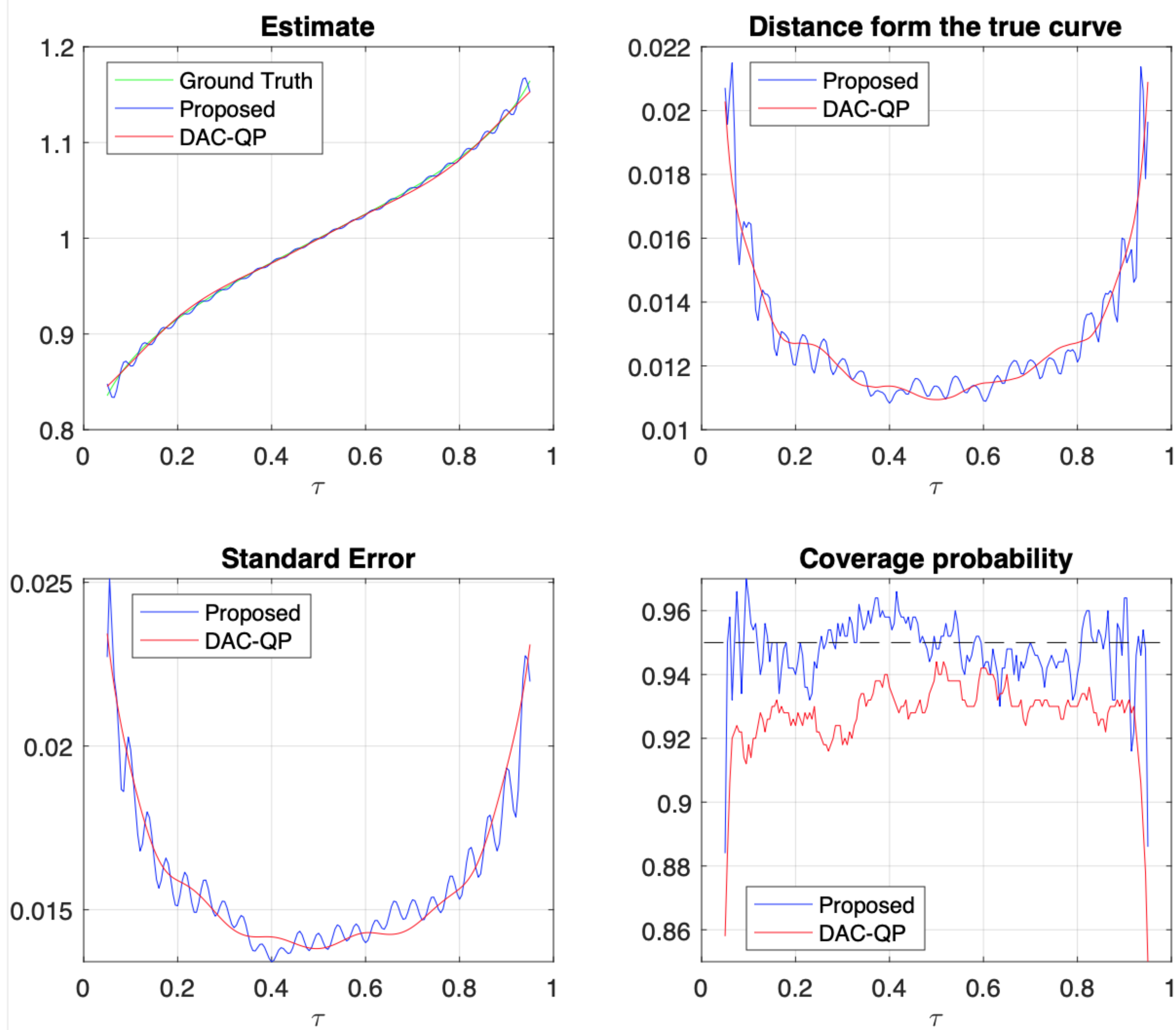


Figure 3: Estimations of $\beta_{01}(\tau)$ with our proposed method and DAC-QP and $\boldsymbol{\theta}_0 = (1, -2)^\top$, $\boldsymbol{\gamma}_0 = (1, 0.5)^\top$. Upper left graph: the true $\beta_{01}(\tau)$ versus the averaged $\hat{\beta}_1(\tau)$. Upper right graph: averaged pointwise distance between $\beta_{01}(\tau)$ and $\hat{\beta}_1(\tau)$. Lower left graph: pointwise standard errors. Lower right graph: pointwise coverage probabilities, the horizontal dashed line (- -) is a reference line with height 0.95. All averaged statistics are calculated over 500 replications.

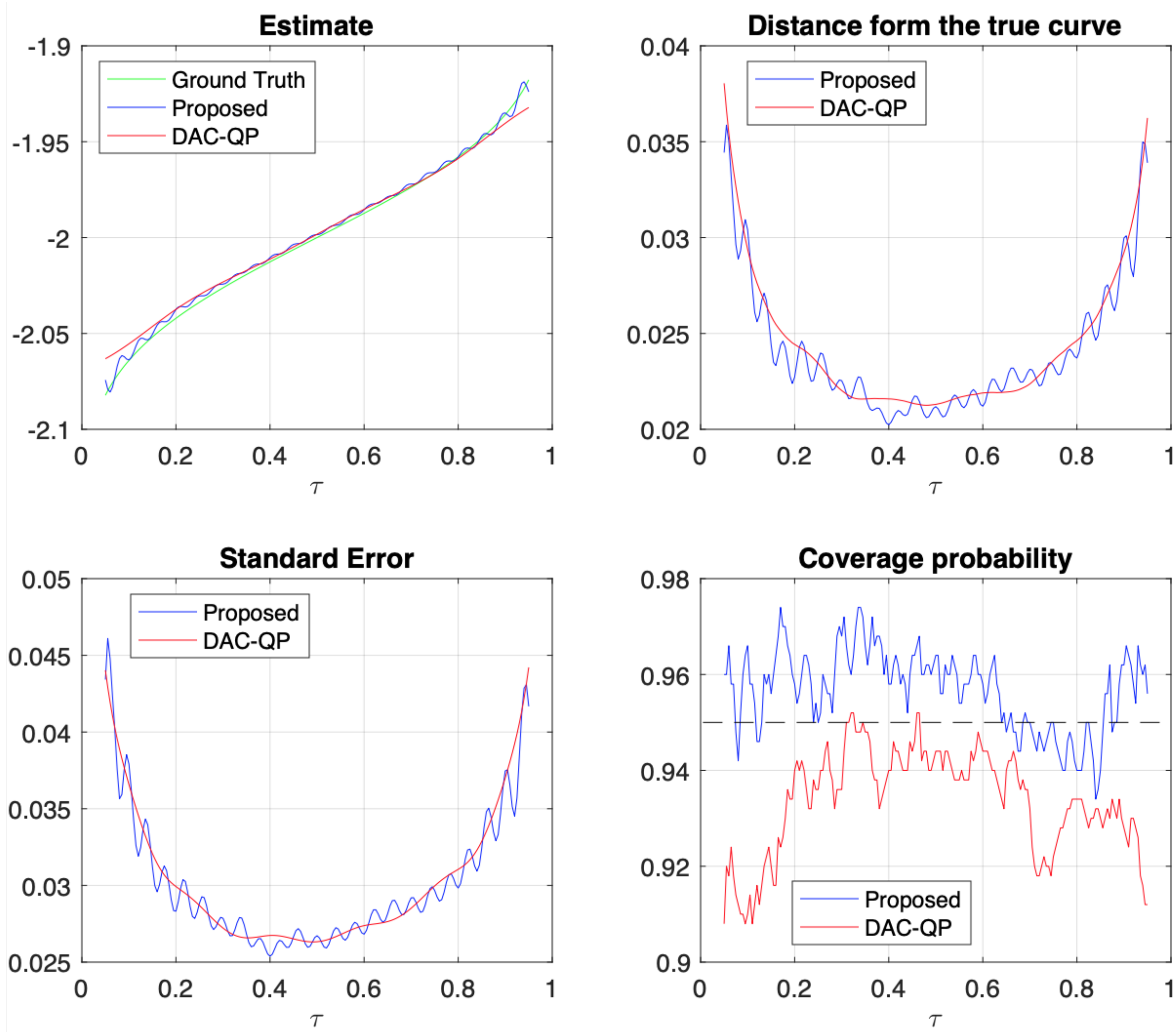


Figure 4: Estimations of $\beta_{02}(\tau)$ with our proposed method and DAC-QP and $\boldsymbol{\theta}_0 = (1, -2)^\top$, $\boldsymbol{\gamma}_0 = (1, 0.5)^\top$. Upper left graph: the true $\beta_{02}(\tau)$ versus the averaged $\hat{\beta}_2(\tau)$. Upper right graph: averaged pointwise distance between $\beta_{02}(\tau)$ and $\hat{\beta}_2(\tau)$. Lower left graph: pointwise standard errors. Lower right graph: pointwise coverage probabilities, the horizontal dashed line (- -) is a reference line with height 0.95. All averaged statistics are calculated over 500 replications.

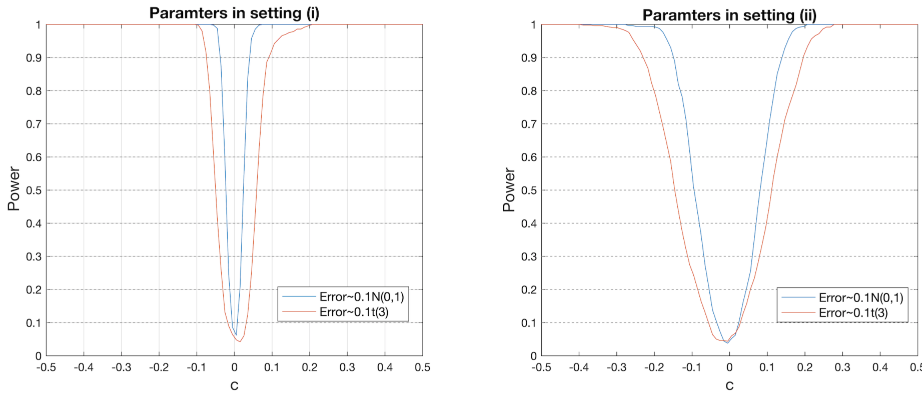


Figure 5: Empirical powers (over 500 replications) with $\alpha = 0.05$ and different settings of $(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)$ and error distributions across different $c$.

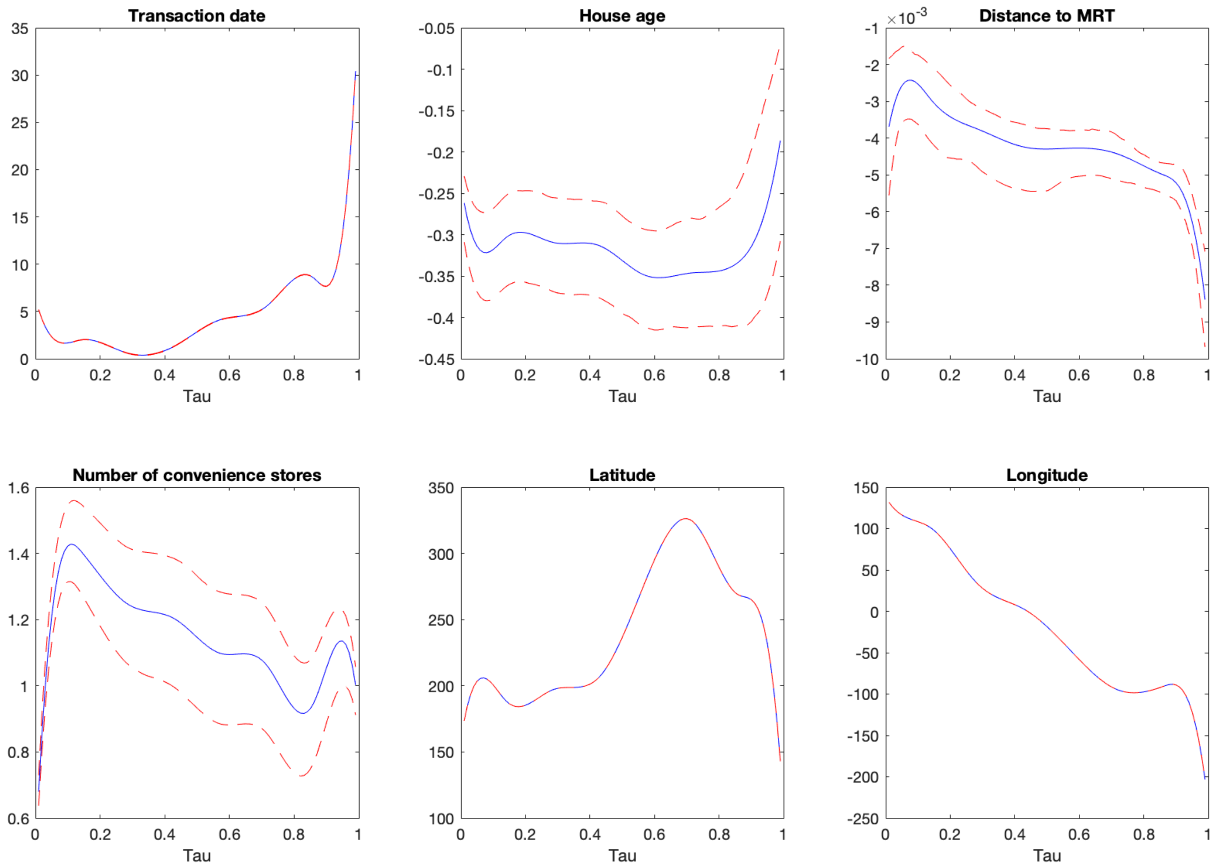


Figure 6: Estimates of the regression quantiles (in blue) and the 95% confidence bands (in red) for the 6 predictors in the analysis of the real estate valuation dataset.

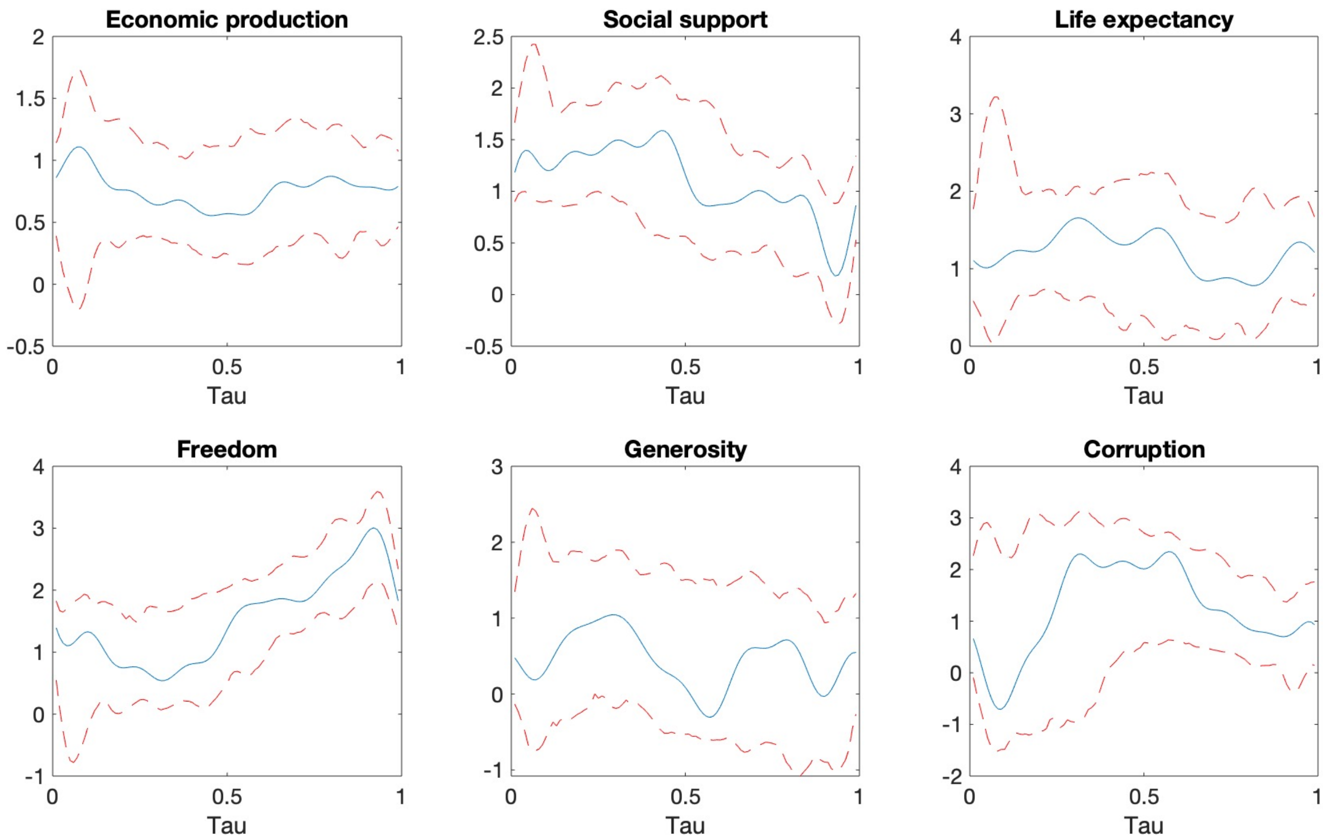


Figure 7: Estimates of the regression quantiles (in blue) and the 95% confidence bands (in red) for the 6 predictors in the analysis of the world happiness report 2019.