| Statistica Sinica Preprint No: SS-2021-0138 | |
|---|---|
| **Title** | Functional Response Quantile Regression Model |
| **Manuscript ID** | SS-2021-0138 |
| **URL** | http://www.stat.sinica.edu.tw/statistica/ |
| **DOI** | 10.5705/ss.202021-0138 |
| **Complete List of Authors** | Xingcai Zhou,<br>Dehan Kong,<br>A. B. Kashlak,<br>Linglong Kong,<br>R. Karunamuni and<br>Hongtu Zhu |
| **Corresponding Author** | Linglong Kong |
| **E-mail** | lkong@ualberta.ca |

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

Statistica Sinica

# Functional Response Quantile Regression Model

Xingcai Zhou[1], Dehan Kong[2], A. B. Kashlak[3], Linglong Kong[3],

R. Karunamuni[3], and Hongtu Zhu[4]

[1]*School of Statistics and Data Science, Nanjing Audit University, Nanjing 211815, China*

[2]*Department of Statistical Sciences, University of Toronto, Toronto, Ontario M5S 3G3, Canada*

[3]*Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1, Canada*

[4]*Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27516, U.S.A.*

*Abstract:* We propose a new functional response quantile regression model, and develop a data-driven estimation procedure to estimate the quantile regression processes based on a local linear approximation. Theoretically, we obtain the global uniform Bahadur representation of the estimator with respect to the time/location and the quantile level, and show that the estimator converges weakly to a two-parameter continuous Gaussian process. We then derive the asymptotic bias and mean integrated squared error of the smoothed individual functions and their uniform convergence rates under given quantile levels. Based

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

on the theoretical results, we introduce a global test for the coefficient functions and discuss how to construct simultaneous confidence bands. We evaluate our method using simulations and by applying it to diffusion tensor imaging data and ADHD-200 functional magnetic resonance imaging data.

*Key words and phrases:* Functional data, Global test statistic, Simultaneous confidence band, Weak convergence.

## 1. Introduction

Functional data analysis deals with data in the form of functions, images, and shapes, as well as more general objects (Wang et al., 2016). Functional regression models are widely used to model functional data, and include the functional linear regression (Ramsay and Dalzell, 1991; Ramsay and Silverman, 2005; Yao et al., 2005b) and functional response regression model (Ramsay and Silverman, 2005). The classical functional linear regression describes the relationship between a scalar response and a functional predictor. In contrast, a functional response regression characterizes the relation between a functional response and scalar predictors.

The functional response regression model is defined as independent realizations of an underlying stochastic process

$$y_i(s) = x_i^{\mathrm{T}}\beta(s) + \eta_i(s) \quad (i = 1, \ldots, n), \tag{1.1}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

where $y_i(s)$ denotes a functional response for the $i$th subject, $x_i$ is its associated $p$-dimensional covariates of interest, $\beta(s) = \{\beta_1(s), \ldots, \beta_p(s)\}^{\mathrm{T}}$ is a $p \times 1$ unknown smooth function of $s \in \mathcal{I}$, and $\eta_i(s)$ includes individual variation used to characterize the within-curve dependence. This functional response model can be used in neuroimaging applications, where researchers use clinical, genetic, and neuropsychological assessment to predict the time/location of varying brain signals. This model is also closely connected to various varying-coefficient models (Hastie and Tibshirani, 1993; Shen and Faraway, 2004; Zhang and Chen, 2007; Zhang, 2011; Zhu et al., 2012). Typically, brain signals are distorted by artifacts and noise, and outliers occur frequently. Studies have found that outliers may affect statistical results and conclusions significantly (Krauledat et al., 2007; Garrido et al., 2013). Quantile regression (Koenker and Bassett, 1978) has emerged as an important statistical methodology that provides robust statistical results when outliers exist. It allows scientists to make a statistical inference on the entire conditional distribution by estimating a collection of conditional quantiles, and does not require specifying an error distribution. Therefore, it is used widely in disciplines such as biology, medicine, finance, and economics, especially to model complex data such as longitudinal and functional data. A comprehensive survey of quantile regression can be found in

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

Koenker (2005).

To this end, we propose the following *functional response quantile regression model* (FRQR):

$$y_i(s) = x_i^{\mathrm{T}}\beta(s,\tau) + \eta_i(s,\tau), \qquad (i = 1, \ldots, n), \tag{1.2}$$

where $\tau \in (0,1)$ is a given quantile level, and $\eta_i(s,\tau)$ is a stochastic process with covariance function $\gamma_\eta(s,t)$ at each $\tau$. Without loss of generality, we assume that the $\tau$th quantile of $\eta_i(s,\tau)$ is equal to zero, that is, $F^{-1}_{\eta_i(s,\tau)}(\tau) = 0$, where $F^{-1}_{\eta_i(s,\tau)}$ is the quantile function of $F_{\eta_i(s,\tau)}$, and $F_{\eta(s,\tau)}$ denotes the distribution function of $\eta_i(s,\tau)$, for any $s$ and a given $\tau$. Thus, the $\tau$th conditional quantile of $y_i(s)$ given $x_i$ can be written as $Q_{y_i(s)}(\tau \mid x_i, s) = x_i^{\mathrm{T}}\beta(s,\tau)$, which is characterized by the only parameter $\beta(s,\tau)$. In fact, if $F^{-1}_{\eta_i(s,\tau)}(\tau) \neq 0$, then $(x_i, y_i(s))$ obeys the conditional quantile restriction $Q_{y_i(s)}(\tau \mid x_i, s) = F^{-1}_{\eta_i(s,\tau)}(\tau) + x_i^{\mathrm{T}}\beta(s,\tau)$. In a quantile regression, each conditional quantile is important for characterizing the distribution of the response. For a fixed $s$, the model (1.2) becomes the usual quantile regression (Koenker, 2005). Therefore, it is identifiable. In addition, the covariance structure of $y_i(\cdot)$ at the $\tau$th quantile level is characterized by the covariance function of $\eta_i(\cdot,\tau)$. We introduce an efficient estimation procedure for $\beta(s,\tau)$ based on a local linear approximation, derive the global uniform Bahadur representation of $\hat{\beta}(s,\tau)$, and establish

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

the weak convergence results of $\hat{\beta}(s,\tau)$. Based on these results, we propose statistical inference procedures, including a global test statistic and simultaneous confidence bands (SCBs). The procedure considers the within-curve dependence structure and the estimation errors of the underlying unknown data distributions.

The FRQR model (1.2) is receiving increasing attention. Liu et al. (2020a) performed a quantile regression on a functional response in a Bayesian framework, and applied it to mass spectrometry proteomics data, Yang et al. (2020) proposed a Bayesian quantile functional regression, using the quantile functions as functional data, to model the entire marginal distribution of the pixel intensities of tumor images, Zhang et al. (2021) developed a novel spatial FRQR model to characterize the conditional distribution of an image response on the whole spatial domain. Liu et al. (2020b) use a quantile regression to provide a comprehensive understanding of how scalar predictors influence the conditional distribution of a functional response, and perform a statistical inference using asymptotic SCBs. Our work differs substantially from Liu et al. (2020b) in the following ways. First, Liu et al. (2020b) perform a quantile regression separately at each sampling location to obtain a pointwise estimator of the coefficient functions, and then construct SCBs based on linear interpolation (LI). In contrast, we ap-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

ply kernel smoothing (KS) to estimate the regression coefficient functions, directly perform the global test for the linear hypotheses of the coefficient functions, and construct an SCB for each coefficient function. The estimation method of Liu et al. (2020b) is based on two points on both sides of the point of interest, and so does not use the full information from its neighbors. Thus, the SCB obtained using the LI method may need to be corrected in practice. Second, Liu et al. (2020b) do not consider a global hypothesis test. Third, the theoretical development in our work differs from that of Liu et al. (2020b), who give a uniform Bahadur representation of the pointwise estimator on a discrete sampling grid, and derive a strong Gaussian approximation of the LI estimator over the location $s \in \mathcal{I}$ for the fixed quantile level $\tau \in (0, 1)$. In contrast, we obtain a global uniform Bahadur representation of the KS estimator, and establish a strong two-parameter Gaussian approximation of the estimator with respect to $s \in \mathcal{I}$ and $\tau \in (0, 1)$. Thus, our theoretical development is more challenging than that of Liu et al. (2020b).

This study makes the following contributions to the literature. First, to the best of our knowledge, this is the first study to consider a functional response in a quantile regression. Second, we develop novel theory, including the global uniform Bahadur representation and the weak convergence

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

of the estimator of the varying-coefficient function, using advanced empirical process methods. To validate our inference procedure, we derive the null distribution of the global test statistic and introduce a wild bootstrap testing procedure with theoretical guarantees. Third, our approach reveals interesting findings in two applications to diffusion tensor imaging data and ADHD-200 functional magnetic resonance imaging data.

## 2. Estimation procedure

### 2.1 Assumptions

Throughout this paper we assume $s \in [0, 1]$, but our results can be easily extended to the unit square, cube, or higher dimensions. We assume that $\eta_i(s, \tau)$ are independent and identical copies of the stochastic process $\eta(s, \tau)$ with covariance function $\gamma_\eta(s, t)$ at each $\tau$, and with the $\tau$th quantile equal to zero. For imaging data, it is typical for the functional response $y_i(s)$ to be measured at the same location for all subjects. Therefore, $y_i(s)$ is measured at the same $m$ location points $0 = s_1 \leq s_2 \leq \cdots \leq s_m = 1$, for all $i = 1, \ldots, n$. We also assume $\beta(\cdot, \tau)$ is twice continuously differentiable.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138



## 2.2 Quantile regression estimation

In model (1.2), the conditional quantile function of $y_i(s)$ given $x_i$ can be expressed as $Q_{y_i(s)}(\tau \mid x_i, s) = x_i^{\mathrm{T}}\beta(s,\tau)$, for $i = 1,\ldots,n$, each $s \in \mathcal{I}$, and any $\tau \in (0,1)$. The main parameter of interest is $\beta(s,\tau)$, which we estimate using a local polynomial regression (Fan and Gijbels, 1996). In particular, $\beta(s_j,\tau)$ can be locally approximated by a linear function $\beta(s_j,\tau) \approx \beta(s,\tau) + \dot{\beta}(s,\tau)(s_j - s) \equiv b_1 + b_2(s_j - s)$, where $\dot{\beta}(s,\tau) = \{d\beta_1(s,\tau)/ds, \cdots, d\beta_p(s,\tau)/ds\}^{\mathrm{T}}$. Let $z_j(s) = z_{h_1 j}(s) = \{1, (s_j - s)/h_1\}^{\mathrm{T}}$, $z_{x,ij}(s) = \{x_i^{\mathrm{T}}, h_1^{-1}(s_j - s)x_i^{\mathrm{T}}\}^{\mathrm{T}} = z_j(s) \otimes x_i$, where $\otimes$ is the Kronecker product, $b = (b_1^{\mathrm{T}}, h_1 b_2^{\mathrm{T}})^{\mathrm{T}}$, $K(s)$ is a kernel function, $K_{h_1 j}(s) = h_1^{-1}K_j(s)$ with $K_j(s) = K\{(s_j - s)/h_1\}$, and $h_1$ is a bandwidth. Then, the estimator $\hat{\beta}(s,\tau)$ can be obtained by minimizing the following locally weighted quantile regression loss function:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \rho_\tau \left\{y_i(s_j) - z_{x,ij}(s)b\right\} K_j(s) \tag{2.1}$$

where $\rho_\tau(u) = u\{\tau - I(u < 0)\}$ is a check function (Koenker and Bassett, 1978). Thus, we have

$$\hat{\beta}(s,\tau) = \{(1,0) \otimes I_p\}\hat{b}.$$

For the choice of $h_1$, we introduce an automatic selection procedure in Section 2.4.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138



## 2.3 Data-driven estimation of density and distribution functions

We need to estimate the unknown error densities of $\eta(s)$ and the unknown bivariate cumulative distribution of the individual effects $\eta(s)$ and $\eta(t)$, for each $s, t \in [0, 1]$. To do so, we use residual-based empirical distributions. It is useful to estimate the distribution of the error using the empirical distribution of the residuals, and then to use this estimated error distribution to develop tests for the model assumptions (Akritas and Keilegom, 2001; Müller et al., 2007; Neumeyer and Keilegom, 2010).

Define the bivariate cumulative distribution of the individual effects $\eta(s)$ and $\eta(t)$ as $F_\eta(a_1, a_2, s, t; \tau, \iota) = \text{pr}\{\eta(s, \tau) < a_1, \eta(t, \iota) < a_2\}$, and denote $f_\eta(a_1, a_2, s, t; \tau, \iota) = \frac{\partial^2 F_\eta(a_1, a_2, s, t; \tau, \iota)}{\partial a_1 \partial a_2}$. We need to estimate $f_\eta(0, s; \tau)$ and $\hat{F}_\eta(0, 0, s, t; \tau, \iota)$, for $s \neq t$, for statistical inference.

First, we can estimate $f_\eta(a, s; \tau)$ using kernel methods, such as

$$\hat{f}_\eta(a, s; \tau) = n^{-1} \sum_{i=1}^{n} K_{h_2}\{a - \hat{\eta}_i(s, \tau)\},$$

with the residuals $\hat{\eta}_i(s, \tau) = y_i(s) - x_i^{\mathrm{T}} \hat{\beta}(s, \tau)$. Thus, we get $\hat{f}_\eta(0, s; \tau)$.

Second, we can estimate the multivariate cumulative distribution function $F_\eta(a_1, a_2, s, t; \tau, \iota)$ using a kernel estimation such as

$$\hat{F}_\eta(a_1, a_2, s, t; \tau, \iota) = \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} \hat{f}_\eta(b_1, b_2, s, t; \tau, \iota) db_1 db_2,$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138



for $s \neq t$, with

$$\hat{f}_\eta(b_1, b_2, s, t; \tau, \iota) = n^{-1} \sum_{i=1}^{n} K_{h_3}\{b_1 - \hat{\eta}_i(s, \tau)\} K_{h_4}\{b_2 - \hat{\eta}_i(t, \iota)\}.$$

This yields $\hat{F}_\eta(0, 0, s, t; \tau, \iota)$.

For statistical inference, we need to estimate the unknown error density. Popular ways of doing so include the difference quotient method (Hendricks and Koenker, 1992; Wang et al., 2009) and the Nadaraya–Watson or local linear kernel method (Fan et al., 1996). We adopt Nadaraya–Watson-type estimators for the densities of the error processes $f_\eta(\cdot, s; \tau)$ and $f_\eta(\cdot, \cdot, s, t; \tau, \iota)$.

## 2.4 Automatic bandwidth selection

Bandwidth selection is critical in local smoothing. To implement our estimation method, we need to choose appropriate bandwidths $h_l$ $(l = 1, \ldots, 4)$, in order to obtain adequate estimators $\hat{\beta}(s)$, $\hat{f}_\eta(a, s; \tau)$, and $\hat{f}_\eta(b_1, b_2, s, t; \tau, \iota)$. Asymptotically, we require that $h_l$ $(l = 1, \ldots, 4)$ satisfies conditions (C7)–(C9), given in Section 4. However, it is difficult to use these conditions in practice. A simple and efficient of obtaining these bandwidths is to use cross-validation (CV) based on data. For the Nadaraya–Watson estimator of the random error density, bandwidth selection often minimizes the least integrated squared CV score (Fan and Yim, 2004; Hall et al., 2004). We

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138



minimize the sum of the integrated squared CV score to choose the bandwidth for the Nadaraya–Watson estimator of the stochastic process density function.

For $h_1$, we minimize the CV score

$$\text{CV}(h_1) = \sum_{i=1}^{n}\sum_{j=1}^{m} \rho_\tau\{y_i(s_j) - x_i^{\mathrm{T}}\hat{\beta}(s_j,\tau,h_1)^{(-i)}\},$$

where $\hat{\beta}(s_j,\tau,h_1)^{(-i)}$ is the locally weighted quantile estimator of $\beta(s,\tau)$, with the bandwidth $h_1$ based on deleting the $i$th subject from the data.

Following the heuristic suggestions of Rice and Silverman (1991), $h_l$ ($l = 2,3,4$) can be chosen using CV. Fan and Yim (2004) and Hall et al. (2004) proposed using CV for nonparametric conditional density estimators. Here, we adopt their method for our estimations of the density functions. For an estimator $\hat{f}_\eta(a,s_l;\tau)$ of $f_\eta(a,s_l;\tau)$, for $s_l \in \mathcal{S}$ and $\tau \in \mathcal{I}$, define the integrated squared error

$$\begin{aligned} I(s_l,\tau) &= \int \left\{\hat{f}_\eta(a,s_l;\tau) - f_\eta(a,s_l;\tau)\right\}^2 da \\ &= \int \hat{f}_\eta^2(a,s_l;\tau)da - 2\int \hat{f}_\eta(a,s_l;\tau)f_\eta(a,s_l;\tau)da + \int f_\eta^2(a,s_l;\tau)da \\ &= I_1(s_l,\tau) - I_2(s_l,\tau) + I_3(s_l,\tau). \end{aligned}$$

Note that $I_3(s_l,\tau)$ does not depend on the bandwidth $h_2$, and $f_\eta(a,s_l;\tau)$ in

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138



$I_2(s_l,\tau)$ is unknown. Thus, the CV estimator of

$$\hat{I}_2(s_l,\tau)=\frac{2}{n}\sum_{i=1}^{n}\hat{f}_\eta^{(-i)}\{\hat{\eta}_i(s_l,\tau),s_l;\tau\}=\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}K_{h_2}\{\hat{\eta}_i(s_l,\tau)-\hat{\eta}_j(s_l,\tau)\},$$

where $\hat{f}_\eta^{(-i)}\{\hat{\eta}_i(s_l,\tau),s_l,\tau\}$ denotes the estimator $f_\eta(a,s_l;\tau)$ based on the data with the observations of the $i$th subject left out at $s_l$.

For $h_3$ and $h_4$, minimize the sum of the integrated squared CV score:

$$\begin{aligned}&\mathrm{SCV}(h_3,h_4)\\&=\sum_{k,l=1}^{m}\left[\iint\hat{f}_\eta^2(b_1,b_2,s_k,s_l;\tau,\iota)db_1db_2-\frac{2}{n}\sum_{i=1}^{n}\hat{f}_\eta^{(-i)}\{\hat{\eta}_i(s_k,\tau),\hat{\eta}_i(s_l,\iota),s_k,s_l;\tau,\iota\}\right]\end{aligned}$$

where $\hat{f}_\eta^{(-i)}\{\hat{\eta}_i(s_k,\tau),\hat{\eta}_i(s_l,\iota),s_k,s_l;\tau,\iota\}=(n-1)^{-1}\sum_{j=1,j\neq i}^{n}K_{h_3}\{\hat{\eta}_i(s_k,\tau)-\hat{\eta}_j(s_k,\iota)\}K_{h_4}\{\hat{\eta}_i(s_l,\tau)-\hat{\eta}_j(s_l,\iota)\}$. In practice, we assume that $h_3=h_4$ to save computational time.

The above bandwidth choice is common in kernel estimation. Based on a simple grid search, we can adaptively select the bandwidths $h_1$, $h_2$, and $h_3$ one by one using the above CV methods. In a simulation study, we found that the bandwidth $h_1$ is slightly sensitive to the estimators, and $h_l$ $(l=2,3)$ are not sensitive to statistical inference. On the whole, the results based on a CV selection in our simulation study and real-data analysis are satisfactory.

Note that the theory developed here does not support a procedure in which the tuning parameters are selected adaptively.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

## 3. Inference procedure

### 3.1 Global test statistic

In this subsection, we propose a global test statistic for the general hypothesis testing problem

$$H_0 : C\beta(s,\tau) = c(s,\tau) \text{ for all } s \text{ and } \tau \quad \text{versus} \quad H_1 : C\beta(s,\tau) \neq c(s,\tau), \tag{3.1}$$

where $C$ is a given $r \times p$ full-rank matrix, and $c(s,\tau) = \{c_1(s,\tau), \ldots, c_r(s,\tau)\}^{\mathrm{T}}$ is a given vector of functions. Denote $e_{l,p}$ to be the $p$-dimensional unit vector with the $l$th component equal to one, and all others equal to zero. If we take $C = e_{l,p}^{\mathrm{T}} = (0, \ldots, 0, 1, 0, \ldots, 0)$ and $c(s,\tau) = 0$, this tests the significance of the $l$th covariate effect on the functional responses; if we take $C = (e_{k,p} - e_{l,p})^{\mathrm{T}}, k \neq l$, and $c(s,\tau) = 0$, it tests for an equality $\beta_k(s,\tau) = \beta_l(s,\tau)$.

From Theorem 1, $n^{1/2}[\hat{\beta}(s,\tau) - \beta(s,\tau) - \mathrm{Bias}\{\hat{\beta}(s,\tau)\}]$ converges weakly to a centered Gaussian process $\mathbb{G}(s,\tau)$ with covariance $\Sigma(s,t;\tau,\iota) = E\{\mathbb{G}(s,\tau)\mathbb{G}^{\mathrm{T}}(t,\iota)\}$, indexed by $s \in [0,1]$ and $\tau \in (0,1)$. Let the residual process $d(s,\tau) = n^{1/2}[C[\hat{\beta}(s,\tau) - \mathrm{Bias}\{\hat{\beta}(s,\tau)\}] - c(s,\tau)]$, and the normalized version $d_c(s,\tau) = \{C\hat{\Sigma}(s,s;\tau,\tau)C^{\mathrm{T}}\}^{-1/2} d(s,\tau)$, where $\hat{\Sigma}(\cdot,\cdot;\cdot,\cdot)$ is an estimator of $\Sigma(\cdot,\cdot;\cdot,\cdot)$, $\hat{\Sigma}(s,t;\tau,\iota) = (1,0)\hat{A}^{-1}(s,\tau)\hat{G}(s,t;\tau,\iota)\hat{A}^{-1}(t,\iota)(1,0)^{\mathrm{T}}\,\hat{\Omega}_x^{-1}$, $\hat{\Omega}_x = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^{\mathrm{T}}$,

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138



$\hat{A}(s,\tau) = \frac{1}{m}\sum_{j=1}^{m} K_{h_1 j}(s)\hat{f}_\eta(0,s_j;\tau)z_j(s)z_j^{\mathrm{T}}(s)$, and the $2\times 2$ matrix $\hat{G}(s,t;\tau,\iota)$ has entries

$$\hat{G}_{kl}(s,t;\tau,\iota) = \begin{cases} \hat{\mu}_k(K;s,h_1)\hat{\mu}_l(K;t,h_1)(\tau\wedge\iota-\tau\iota), & if \quad s=t, \\ \hat{\mu}_k(K;s,h_1)\hat{\mu}_l(K;t,h_1)\{\hat{F}_\eta(0,0,s,t;\tau,\iota)-\tau\iota\}, & if \quad s\neq t, \end{cases}$$

$$(k,l=0,1),$$

$$\hat{\mu}_k(K;s,h_1) = m^{-1}\sum_{j=1}^{m} K_{h_1 j}(s)\left(\frac{s_j-s}{h_1}\right)^k \qquad (k=0,1,2).$$

From the weak convergence and uniform consistency of $\hat{\Sigma}(s,t;\tau,\iota)$ in $[0,1]^2\times(0,1)^2$, $d_c(s,\tau)$ converges weakly to $\mathbb{G}_c(s,\tau) = \{C\Sigma(s,s;\tau,\tau)C^{\mathrm{T}}\}^{-1/2}C\mathbb{G}(s,\tau)$, by Slutsky's theorem (Kosorok, 2008), where $\mathbb{G}_c(s,\tau)$ is a centered Gaussian process.

The global test statistic for the linear hypothesis $H_0$ is defined as follows: $T_n = \int_0^1\int_0^1 d_c^{\mathrm{T}}(s,\tau)d_c(s,\tau)dsd\tau$. Let $\tilde{T}_n = \int_0^1\int_0^1 \mathbb{G}_c^{\mathrm{T}}(s,\tau)\mathbb{G}_c(s,\tau)dsd\tau$. We show in Theorem 3 that $T_n$ converges weakly to $\tilde{T}_n$ .

Because Bias$\{\hat{\beta}(s,\tau)\}$ is unknown, we can estimate it by

$$\begin{aligned} &\widehat{\mathrm{Bias}}\{\hat{\beta}(s,\tau)\} \\ &= -(I_p,0)\left\{\hat{A}^{-1}(s,\tau)\otimes\hat{\Omega}_x^{-1}\right\}(nm)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m} K_{h_1 j}(s)z_{x,ij}(s)\hat{e}_{ij}(s,\tau), \end{aligned} \tag{3.2}$$

where $\hat{e}_{ij}(s,\tau) = -2^{-1}x_i^{\mathrm{T}}\hat{\ddot{\beta}}(s,\tau)(s_j-s)^2 - 6^{-1}x_i^{\mathrm{T}}\widehat{\dddot{\beta}}(s,\tau)(s_j-s)^3$, in which

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138



$\hat{\dot{\beta}}(s,\tau)$ and $\hat{\ddot{\beta}}(s,\tau)$ are obtained using a local cubic fit with a selected bandwidth.

In practice, it is difficult to obtain the percentiles of $T_n$ directly, even if its null distribution is known. Therefore, we use a wild bootstrap method to approximate the critical values of $T_n$. In particular, we fit model (1.2) under $H_0$ to obtain $\hat{\beta}^*(s_j,\tau)$ and $\hat{\eta}^*_{i,0}(s_j,\tau)$, for $i=1,\ldots,n$, $j=1,\ldots,m$. Then, we generate a random sample $\tilde{\varsigma}_i(s_j)$ from a standard normal distribution, for $i=1,\ldots,n$, $j=1,\ldots,m$, and compute $\tilde{y}_i(s_j)=x_i^{\mathrm{T}}\hat{\beta}^*(s_j,\tau)+\tilde{\varsigma}_i(s_j)\eta^*_{i,0}(s_j,\tau)$. From $\tilde{y}_i(s_j)$, we reobtain $\tilde{\beta}(s,\tau)$, $\mathrm{Bias}\{\tilde{\beta}(s,\tau)\}$, and $d^{(B)}(s,\tau)=n^{1/2}[C[\tilde{\beta}(s,\tau)-\mathrm{Bias}\{\tilde{\beta}(s,\tau)\}]-c(s,\tau)]$ by using the locally weighted quantile regression loss function (2.1). As shown in Theorem 1, $\mathrm{Bias}\{\tilde{\beta}(s,\tau)\}$ is asymptotically negligible; thus, we drop the term $\mathrm{Bias}\{\tilde{\beta}(s,\tau)\}$ for computational efficiency. Specifically, we calculate $\tilde{T}_n=\int_0^1\int_0^1 d_c(s,\tau)^{\mathrm{T}}\,d_c(s,\tau)dsd\tau$, where $d_c(s,\tau)=\{C\tilde{\Sigma}(s,s;\tau,\tau)C^{\mathrm{T}}\}^{-1/2}d(s,\tau)$. Repeat $M$ times to have $\{\tilde{T}_{n,B},B=1,\ldots,M\}$, and compute the $p$-value as $M^{-1}\sum_{B=1}^{M}I(\tilde{T}_{n,B}\geq T_n)$. If the $p$-value is smaller than a given significance level $\alpha$, say 0.05, then we reject $H_0$.

Although we theoretically obtain a global test statistic $T_n$ over $s\in[0,1]$ and $\tau\in(0,1)$, in practice, the required double integration for $T_n$ is computationally expensive. Thus, we are often more interested in the

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138



function of $s$ for a given $\tau$. Therefore, we consider the global test statistic $T_n^* = \int_0^1 d_c^{\mathrm{T}}(s,\tau) d_c(s,\tau) ds$ of $\{\hat{\beta}(s,\tau), s \in [0,1]\}$ for a given $\tau \in (0,1)$. The null distribution of $T_n^*$ is given in Theorem 3 (b). A similar wild bootstrap procedure can be implemented to improve computational efficiency.

### 3.2 SCBs

For a given quantile $\tau$ and a preassigned significance level $\alpha$, we construct SCBs of $\beta_l(s,\tau)$ for each $\tau$, that is, we find $\hat{\beta}_l^{L,\alpha}(s,\tau)$ and $\hat{\beta}_l^{U,\alpha}(s,\tau)$, such that

$$\mathrm{pr}\{\hat{\beta}_l^{L,\alpha}(s,\tau) < \beta_l(s,\tau) < \hat{\beta}_l^{U,\alpha}(s,\tau), s \in [0,1], \tau \in (0,1)\} = 1-\alpha \quad (l = 1, \cdots, p).$$

By the weak convergence of $\hat{\beta}(s,\tau)$, we have that $\sup_{s\in[0,1],\tau\in(0,1)} |n^{1/2}[\hat{\beta}_l(s,\tau) - \beta_l(s,\tau) - \mathrm{Bias}\{\hat{\beta}_l(s,\tau)\}]|$ converges weakly to $\sup_{s\in[0,1],\tau\in(0,1)} |\mathbb{G}_l(s,\tau)|$. Define $C_l(\alpha)$ such that $\mathrm{pr}\{\sup_{s\in[0,1],\tau\in(0,1)} |\mathbb{G}_l(s,\tau)| \le C_l(\alpha)\} = 1-\alpha$. The $1-\alpha$ SCB for $\beta_l(s,\tau)$ can be written as $[\hat{\beta}_l(s,\tau) - \mathrm{Bias}\{\hat{\beta}_l(s,\tau)\} - C_l(\alpha)n^{-1/2}, \hat{\beta}_l(s,\tau) - \mathrm{Bias}\{\hat{\beta}_l(s,\tau)\} + C_l(\alpha)n^{-1/2}]$, where $\mathrm{Bias}\{\hat{\beta}_l(s,\tau)\}$ is given by (3.2). Here, we can drop the term $\mathrm{Bias}\{\hat{\beta}_l(s,\tau)\}$ for computational efficiency.

Now, we approximate $C_l(\alpha)$ using an efficient resampling approach, in line with Kosorok (2003), Zhu et al. (2007), and Zhu et al. (2012). We estimate $\hat{\eta}_i(s_j) = y_i(s_j) - x_i^{\mathrm{T}}\hat{\beta}(s_j,\tau)$, for $i = 1, \cdots, n$ and $j = 1, \cdots, m$. For $B = 1, \cdots, M$, we generate independent samples $\{\varsigma_i^{(B)}, i = 1, \cdots, n\}$ from

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

$N(0,1)$ and compute the stochastic process $\mathbb{G}^{(B)}(s,\tau) = -n^{-1/2}m^{-1}(I_p,0)\check{S}_{nX}^{-1}(s,\tau)\sum_{i=1}^{n}$ $\varsigma_i^{(B)}\sum_{j=1}^{m}K_{h_1 j}(s)z_{x,ij}(s)\hat{w}_{ij}^{*}(s,\tau)$, where $\check{S}_{nX}(s,\tau) = (nm)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}K_{h_1 j}(s)$ $\hat{f}_\eta(0,s_j;\tau)z_{x,ij}(s)z_{x,ij}^{\mathrm{T}}(s)$ and $\hat{w}_{ij}^{*}(s,\tau) = I\{\hat{\eta}_i(s_j,\tau)\le 0\} - \tau$. We calculate $\sup_{s\in[0,1],\tau\in(0,1)}|e_{l,p}\mathbb{G}^{(B)}(s,\tau)|$, for all $B$, and then adopt its $1-\alpha$ empirical percentile as an estimate of $C_l(\alpha)$.

We are also interested in constructing SCBs for $\beta(s,\tau)$ for a given $\tau$. In this case, we need to change the previous inference procedure by replacing $\sup_{s\in[0,1],\tau\in(0,1)}|e_{l,p}\mathbb{G}^{(B)}(s,\tau)|$ with $\sup_{s\in[0,1]}|e_{l,p}\mathbb{G}^{(B)}(s,\tau)|$.

## 4. Theoretical properties

In this section, we provide theoretical guarantees for the estimators and inference procedures developed in Sections 2 and 3. We first introduce some notation. For any smooth functions $f(s,\tau)$ and $g(s,t;\tau,\iota)$, let $\dot{f}(s,\tau) = df(s,\tau)/ds$, $\ddot{f}(s,\tau) = d^2f(s,\tau)/ds^2$, $\dddot{f}(s,\tau) = d^3f(s,\tau)/ds^3$, and $g^{(a,b)}(s,t) = \partial^{a+b}g(s,t;\tau,\iota)/\partial^a s\partial^b t$, where $a$ and $b$ are any nonnegative integers. For a

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

square matrix $A$, $\det(A)$ denote the determinant of $A$. Define

$$\begin{aligned}
\mu_r(K) &= \int s^r K(s) ds, \\
\mu_r\left[K^\star\{(s-t)/h\}\right] &= \int u^r K(u) K\left(u + \frac{s-t}{h}\right) du, \\
\nu_r(K) &= \int s^r K^2(s) ds, \\
\mu_r(K; s, h) &= \int_0^1 h^{-r} (u-s)^r K_h(u-s) du,
\end{aligned}$$

where $r$ is any nonnegative integer.

Denote

$$\begin{aligned}
\Phi(s,h) &= \begin{Bmatrix} \mu_0(K;s,h) & \mu_1(K;s,h) \\ \mu_1(K;s,h) & \mu_2(K;s,h) \end{Bmatrix}, \\
\tilde{\mu}(K,s,h_1) &= \frac{\mu_2^2(K,s,h_1) - \mu_0(K,s,h_1)\mu_3(K,s,h_1)}{\mu_0(K,s,h_1)\mu_2(K,s,h_1) - \mu_1^2(K,s,h_1)}.
\end{aligned}$$

In this section, we present the conditions needed for the main theorems 1, 3, and 4. The following regularity conditions are sufficient for the asymptotic properties, although they might not be the weakest possible. Moreover, we do not distinguish between the differentiation and continuation at the boundary points and those in the interior of $[0, 1]$. For example, a function is continuous at the boundary of $[0, 1]$, which means that the function is left continuous at zero and right continuous at one.

We require the following technical conditions:

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

(C1). The sequence $\{\eta_i(s,\tau) : s \in [0,1]\}$ is a stochastic process with the $\tau$th quantile conditional on $(x_i, s)$ equal to zero.

(C2). The grid points $\mathcal{S}$ are prefixed according to $\pi(s)$ such that $\int_0^{s_j} \pi(s)ds = j/m$, for $1 \le j \le m$. Moreover, $\pi(s) > 0$, for $s \in [0,1]$, and $\pi(s)$ has a continuous second-order derivative on $(0,1)$.

(C3). The distribution function $F_\eta(a, s; \tau)$ is Lipschitz continuous in $a$, for each $s$ and $\tau$, with continuous density $f_\eta(a, s; \tau)$ uniformly bounded away from zero and $\infty$ that has a continuous second-order derivative at $s$ and is Lipschitz continuous in $\tau$. The joint distribution function $F_\eta(a, b, s, t; \tau, \iota)$ is continuous at $a$ and $b$, and has continuous second-order derivatives at $s$ and $t$, with continuous density $f_\eta(a, b, s, t; \tau, \iota)$ that has continuous second-order derivatives at $s$ and $t$.

(C4). $\beta_l(s,\tau)$ have continuous second-order derivatives with respect to $s$ for each $\tau$ and all $l = 1, \cdots, p$. Furthermore $\partial^2\beta_l(s,\tau)/\partial s^2$ are Lipschitz continuous over $\{(s,\tau) : (t-s)/h_1 \in \text{supp}\{K(\cdot)\}, s, t \in \mathcal{I}, \tau \in (0,1)\}$ for all $l = 1, \cdots, p$.

(C5). The covariates $x_i$ are independently and identically distributed (i.i.d.) with $\|x_i\| < \infty$, and $E(x_i x_i^{\mathrm{T}}) = \Omega_x$, which is positive definite.

(C6). The kernel function $K(\cdot)$ in Section 2.2 of the main text is a symmetric Lipschitz continuous density function with compact support $[-1, 1]$.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

Moreover, $\inf_{h\in(0,h_0],s\in[0,1]} \det\{\Phi(s,h)\} > 0$ for a small scalar $h_0 > 0$.

(C7) The bandwidth $h_1$ satisfies (i) $h_1 \to 0$, $mh_1 \to \infty$ and (ii) $m/(nh_1^3) = O(1)$.

(C8) The density $f_\eta(a, s; \tau)$ has bounded continuous third-order partial derivatives with respect to $a$ for all $s \in [0, 1]$ and $\tau \in (0, 1)$. The kernel function $K(\cdot)$ in $K_{h_2}(\cdot)$ is third-order continuously differentiable. The bandwidth $h_2$ satisfies $nh_2^3 \to \infty$, $nh_2^6 \to 0$, and $h_1^{-1}h_2^{2/3} \to \infty$.

(C9) $f_\eta(a, b, s, t; \tau, \iota)$ has bounded continuous second-order partial derivatives and mixed derivatives with respect to $a$ and $b$ for all $s, t \in [0, 1]$ and given $\tau, \iota \in (0, 1)$. The kernel function $K(\cdot)$ in $K_{h_3}(\cdot)$ and $K_{h_4}(\cdot)$ is Lipschitz continuous. The bandwidths $h_3$ and $h_4$ satisfy $h_1h_l^{-1} \to 0$ and $nh_l^4 \to \infty$, for $l = 3, 4$.

(C10) For given $\tau$ and $\iota$, the function $\tilde{\gamma}(s, t) = \gamma(s, s; \tau, \tau)^{-1/2}\gamma(s, t; \tau, \iota)$ $\gamma(t, t; \iota, \iota)^{-1/2}$, $s, t \in [0, 1]$, has a finite trace, that is, $\mathrm{tr}(\tilde{\gamma}) = \int_0^1 \tilde{\gamma}(s, s)ds < \infty$, where $\gamma(s, t; \tau, \iota)$ is defined in Theorem 1 (b).

**Remark 1.** Condition (C1) is a general assumption for quantile regression. Conditions (C2) and (C4)–(C7) are used by Zhu et al. (2012). The conditions of the distribution/density function are necessary for quantile regression (Koenker, 2005; Cai and Xu, 2009; Wang et al., 2009; Chen and Müller, 2012; Kato, 2012). Condition (C3) is a very general assumption for

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

the distribution/density function, and is specific to a quantile regression for functional response data. Conditions (C8)–(C9) are used to establish the weak consistency of the error densities of $\eta(s,\tau)$, and the unknown bivariate cumulative distribution of the individual effects $\eta(s,\tau)$ and $\eta(t,\iota)$, for each $s,t \in [0,1]$ and $\tau,\iota \in (0,1)$. To obtain a $\chi^2$-type mixture distribution of $T_n$ in (3.6), we need the finite trace of $\tilde{\gamma}(s,t)$ in Condition (C10), which ensures that $\tilde{\gamma}(s,t)$ has the singular value decomposition defined in (4.1). Condition (C10) is similar to the conditions used in Theorem 7 of Zhang and Chen (2007).

**Remark 2.** Conditions (C2) and (C7) are weak conditions on the random grid points $\mathcal{S} = \{s_j : 1,\ldots,m\}$. Our proof can be easily extended to the case of fixed grid points.

We first present the global uniform Bahadur representation and weak convergence results of our estimate $\hat{\beta}(s,\tau)$ in the following theorem. The proof of this theorem is deferred to the Supplementary Material.

**Theorem 1.** *Suppose that Conditions (C1)–(C7) hold. Then, the following results hold:*

*(a) (Global uniform Bahadur representation)*

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

$$\sqrt{n}\left[\hat{\beta}(s,\tau)-\beta(s,\tau)-\frac{1}{2}h_1^2\tilde{\mu}(K,s,h_1)\ddot{\beta}(s,\tau)\{1+o_p(1)\}\right]$$
$$=\pi(s)^{-1}f_\eta^{-1}(0,s;\tau)(I_p,0)\left\{\Phi^{-1}(s,h_1)\otimes\Omega_x^{-1}\right\}$$
$$\times\left\{\frac{1}{\sqrt{n}m}\sum_{i=1}^{n}\sum_{j=1}^{m}\left[\tau-I\{\eta_i(s_j,\tau)\le 0\}\right]z_{x,ij}(s)K_{h_1j}(s)\right\}.$$

*(b) (Weak convergence)*

$$\sqrt{n}\left[\hat{\beta}(s,\tau)-\beta(s,\tau)-\frac{1}{2}h_1^2\tilde{\mu}(K,s,h_1)\ddot{\beta}(s,\tau)\{1+o_p(1)\}:s\in[0,1],\tau\in(0,1)\right]$$

*converges weakly to a two-parameter centered Gaussian process* $\mathbb{G}(\cdot,\cdot)$ *with covariance matrix function* $\Sigma(s,t;\tau,\iota)=E\{\mathbb{G}(s,\tau)\mathbb{G}^{\mathrm{T}}(t,\iota)\}=\gamma(s,t;\tau,\iota)\Omega_x^{-1}$, *where*

$$\gamma(s,t;\tau,\iota)=(1,0)A^{-1}(s,\tau)G(s,t;\tau,\iota)A^{-1}(t,\iota)(1,0)^{\mathrm{T}},$$

$$A(s,\tau)=f_\eta(0,s;\tau)\Phi(s,h_1),$$

$$G(s,t;\tau,\iota)=\begin{Bmatrix} G_{00}(s,t;\tau,\iota) & G_{01}(s,t;\tau,\iota)\\ G_{10}(s,t;\tau,\iota) & G_{11}(s,t;\tau,\iota)\end{Bmatrix},$$

*with*

$$G_{kl}(s,t;\tau,\iota)=\begin{cases}\mu_k(K;s,h_1)\mu_l(K;s,h_1)(\tau\wedge\iota-\tau\iota), & if\quad s=t,\\ \mu_k(K;s,h_1)\mu_l(K;t,h_1)\left\{F_\eta(0,0,s,t;\tau,\iota)-\tau\iota\right\}, & if\quad s\neq t,\end{cases}$$

$(k,l=0,1).$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

**Remark 3.** Theorem 1 (a) gives a global uniform Bahadur representation of the local quantile regression estimator $\{\hat{\beta}(s,\tau) : s \in [0,1], \tau \in (0,1)\}$. Theorem 1 (b) establishes the weak convergence of $\{\hat{\beta}(s,\tau) : s \in [0,1], \tau \in (0,1)\}$. For Theorem 1 (b), if $s \in (0,1)$, then $\sqrt{n}\left[\hat{\beta}(s) - \beta(s) - 0.5h_1^2\mu_2(K)\ddot{\beta}(s)\{1+o_p(1)\} : s,\tau \in (0,1)\right]$ converges weakly to a centered Gaussian process $\mathbb{G}(\cdot,\cdot)$ with covariance matrix $\Sigma(s,t;\tau,\iota) = \frac{F_\eta(0,0,s,t;\tau,\iota)-\tau\iota}{f_\eta(0,s;\tau)f_\eta(0,t;\iota)}\Omega_x^{-1}$.

**Theorem 2.** *(a) Under Conditions (C1)–(C8), we have for $s \in [0,1]$ and $\tau \in (0,1)$,*

$$\sup_a |\hat{f}_\eta(a,s;\tau) - f_\eta(a,s;\tau)|$$
$$= O_p\left\{h_2^2 + h_1^2 + (nh_2)^{-1/2}|\log h_2|^{1/2} + n^{-1/2} + h_2^{-4}\left(n^{-3/2} + h_1^6\right)\right\}.$$

*(b) Under Conditions (C1)–(C8) and (C9), we have for $s,t \in [0,1]$ and $\tau,\iota \in (0,1)$,*

$$\sup_{b_1,b_2} |\hat{f}_\eta(b_1,b_2,s,t;\tau,\iota) - f_\eta(b_1,b_2,s,t;\tau,\iota)|$$
$$= O_p\left\{h_3^2 + h_4^2 + \left(\frac{\log n}{nh_3h_4}\right)^{1/2}\right\} + O_p\left\{\left(h_1^2 + n^{-1/2}\right)\left(h_3^{-2} + h_4^{-2}\right)\right\}.$$

Theorem 2 implies that the estimators of the "error" process densities $f_\eta(\cdot,s)$ and $f_\eta(\cdot,\cdot,s,t)$ converge uniformly to their true densities, and gives their rates of convergence. From Theorem 2, $\sup_{a,b\in[0,1]} |\hat{F}_\eta(a,b,s,t) - F_\eta(a,b,s,t)| = o_p(1)$, for each $s,t \in [0,1]$, which is used in the inference procedure.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

Define $\tilde{\gamma}(s,t) = \gamma(s,s;\tau,\tau)^{-1/2}\gamma(s,t;\tau,\iota)\gamma(t,t;\iota,\iota)^{-1/2}, s,t \in [0,1]$. By Condition (C10), $\int_0^1 \int_0^1 \tilde{\gamma}^2(s,t)dsdt < \infty$. By the Cauchy–Schwarz inequality, the function $\tilde{\gamma}(s,t)$ has the eigen decomposition

$$\tilde{\gamma}(s,t) = \sum_{l=1}^{l_0} \kappa_l \psi_l(s)\psi_l(t),$$

where the $\kappa_l$ are the eigenvalues, in decreasing order, $\psi_l$ are the associated orthonormal eigenfunctions of $\tilde{\gamma}(s,t)$, and $l_0$ is the number of positive eigenvalues.

We next present theorems about our inference procedure. The proofs of the theorems are deferred to the Supplementary Material.

**Theorem 3.** *(a) Under Conditions (C1)–(C9), we have*

$$T_n = \tilde{T}_n + o_p(1).$$

*(b) In addition, if Condition (C10) holds, then*

$$T_n^* \Rightarrow \sum_{l=1}^{l_0} \kappa_l \chi_l^2(r),$$

*where $\chi_l^2(r)$ is the $l$th random variable that follows a central $\chi^2$-distribution with $r$ degrees of freedom.*

**Theorem 4.** *Under Conditions (C1)–(C10), the bootstrapped process of $\{\mathbb{G}^{(B)}(s,\tau) : s \in [0,1], \tau \in (0,1)\}$ converges weakly to $\mathbb{G}(s,\tau)$ conditioning*

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

*on the data, where* $\mathbb{G}(s,\tau)$ *is a centered Gaussian process indexed by* $s \in [0,1]$ *and* $\tau \in (0,1)$.

**Remark 4.** Theorems 3 and 4 present theoretical support for the statistical inference procedure introduced in Section 3. Theorem 3 suggests that $T_n$ asymptotically follows a mixture of chi-square distributions. To use the asymptotic results for testing, we need to select an integer $\iota$ such that the eigenvalues $\kappa_l$ $(l = 1, \ldots, \iota)$ explain a sufficiently large portion of the total variation $tr(\tilde{\gamma})$. This threshold is often difficult to select in practice. Therefore, we use the bootstrap method to obtain approximate critical values of $T_n$, where Theorem 4 provides the theoretical guarantee.

**Remark 5.** The proofs of Theorems 3 and 4 rely on the uniform convergence rates of $\hat{f}_\eta(\cdot, s)$ and $\hat{f}_\eta(\cdot, \cdot, s, t)$. These results are developed in Theorem 2.

## 5. Simulation study

In this section, we conduct simulation studies to evaluate our estimation and inference procedures. In the estimations and the SCBs, we compare the finite simple performance of our KS wild bootstrap method and the LI random weighted bootstrap method of Liu et al. (2020b) (hereafter, denoted by KS and LI, respectively) in terms of their root mean integrated squared

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

errors, uniform coverage probabilities, and average coverage widths, as well as providing a graphical visualization.

The data are generated from the following heteroscedastic model $y_i(s_j) = x_{i1}\beta_1(s_j) + x_{i2}\beta_2(s_j) + x_{i3}\beta_3(s_j) + \eta_i(s_j,\tau)$, for $i = 1,\ldots,n, j = 1,\ldots,m$, where $x_{i1} = 1$, $(x_{i2}, x_{i3})^{\mathrm{T}} \sim N\{(0,0)^{\mathrm{T}}, \mathrm{diag}(1-1/\sqrt{2}, 1-1/\sqrt{2}) + 1/\sqrt{2}(1,1)^{\mathrm{T}}(1,1)\}$, $s_j \sim \mathrm{Uniform}[0,1]$. To generate $\eta_i(s_j,\tau)$, we first let $\eta_i(s_j) = v_i(s_j) + \varepsilon_i(s_j)$, where $\{v_i(s_1),\ldots,v_i(s_m)\}^{\mathrm{T}}$ follows a multivariate normal distribution with zero mean, and its covariance matrix has a first-order autoregressive correlation structure with $\mathrm{corr}\{v_i(s_j), v_i(s_l)\} = \gamma^{|j-l|}$ with strength $\gamma = 0.5$, and $\varepsilon_i(s_j) \sim N(0, 0.2)$. Because $\eta_i(s_1),\ldots,\eta_i(s_m)$ are dependent, we set $\eta_i(s_j,\tau) = \eta_i(s_j) - F^{-1}(\tau)$, with $F$ being the marginal density function of $\eta_i(s_j)$. Here, $F^{-1}(\tau)$ is subtracted from $\eta_i(s_j)$ to make the $\tau$th quantile of $\eta_i(s_j,\tau)$ zero for identifiability. We set $\beta_1(s) = s^2$, $\beta_2(s) = (1-s)^4$, and $\beta_3(s) = \exp(s) - 1$. In all our numerical studies, we use a Gaussian kernel and select the bandwidths using the procedure in Section 2.4.

We report the root mean integrated squared error, defined as

$$\mathrm{RMISE}_\tau = \left\{ m^{-1} \sum_{j=1}^{m} |\hat{\beta}_l(s_j,\tau) - \beta_l(s_j,\tau)|^2 \right\}^{1/2} \quad (l = 1,2,3),$$

where $m$ is the number of locations. The average root mean integrated squared errors over 500 Monte Carlo runs for $m = 50, 70, 90$ are included in Table 1. The results show that $\mathrm{RMISE}_\tau$ decreases as $m$ increases, which

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

validates the consistency results of the two estimation methods. KS outperforms LI, having a much smaller RMISE in most cases.

Second, we test the hypotheses $H_0 : \beta_1(s,\tau) = 0$ for all $s$ against $H_1 : \beta_1(s,\tau) \neq 0$ for at least one $s$, for fixed $\tau = 0.10, 0.25, 0.50, 0.75$ and 0.90. Here, we do not provide a comparison with LI, because Liu et al. (2020b) do not give a global hypothesis test. We set $\beta_2(s) = c(1-s)^4$, with $c = 0, 0.2, 0.3, \ldots, 3.0$, to examine the power of $T_n^*$ when $m = 50$ and $n = 100$. We use $B = 200$ bootstrap samples, and depict the power curves in Figure 1. The rejection rates for $T_n^*$ based on the wild bootstrap method are accurate for the different quantile levels at both significance levels ($\alpha = 0.05$ and 0.01). However, the power of the hypothesis tests at the extreme quantile levels ($\tau = 0.10$ and 0.90) is obviously weaker than that at moderate levels ($\tau = 0.25$, 0.50, and 0.75).

Last, we evaluate the coverage probabilities (CPs) and coverage widths (SWs) of the SCBs for $\beta_l(s)$ ($l = 1, 2, 3$) using the wild bootstrap procedure, and compare them with those of the LI method in Liu et al. (2020b). Both methods have theoretical guarantees. For the estimation of $C(\alpha)$, Liu et al. (2020b) adopt the weighted bootstrap method proposed in Belloni et al. (2019), and draw random weights from an exponential distribution with parameter one. In the simulation, we find that the $C(\alpha)$ obtained by

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

Table 1: Simulated results for the root mean integrated squared errors for the KS and LI methods.

| | | $m = 50$ | | | $m = 70$ | | | $m = 90$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Method | $\tau$ | $\beta_1(s)$ | $\beta_2(s)$ | $\beta_3(s)$ | $\beta_1(s)$ | $\beta_2(s)$ | $\beta_3(s)$ | $\beta_1(s)$ | $\beta_2(s)$ | $\beta_3(s)$ |
| KS | 0.10 | 0.0139 | 0.0295 | 0.0108 | 0.0110 | 0.0249 | 0.0094 | 0.0081 | 0.0186 | 0.0071 |
| | 0.25 | 0.0115 | 0.0291 | 0.0100 | 0.0118 | 0.0287 | 0.0077 | 0.0071 | 0.0173 | 0.0073 |
| | 0.50 | 0.0130 | 0.0279 | 0.0129 | 0.0092 | 0.0237 | 0.0115 | 0.0074 | 0.0172 | 0.0086 |
| | 0.75 | 0.0147 | 0.0294 | 0.0097 | 0.0140 | 0.0246 | 0.0131 | 0.0079 | 0.0174 | 0.0080 |
| | 0.90 | 0.0154 | 0.0308 | 0.0114 | 0.0135 | 0.0288 | 0.0096 | 0.0078 | 0.0183 | 0.0084 |
| LI | 0.10 | 0.0205 | 0.0283 | 0.0235 | 0.0187 | 0.0235 | 0.0200 | 0.0176 | 0.0225 | 0.0222 |
| | 0.25 | 0.0142 | 0.0235 | 0.0228 | 0.0144 | 0.0210 | 0.0163 | 0.0091 | 0.0167 | 0.0153 |
| | 0.50 | 0.0172 | 0.0195 | 0.0220 | 0.0142 | 0.0212 | 0.0194 | 0.0107 | 0.0156 | 0.0158 |
| | 0.75 | 0.0191 | 0.0256 | 0.0266 | 0.0151 | 0.0231 | 0.0225 | 0.0122 | 0.0169 | 0.0160 |
| | 0.90 | 0.0203 | 0.0246 | 0.0207 | 0.0264 | 0.0198 | 0.0253 | 0.0226 | 0.0215 | 0.0184 |

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

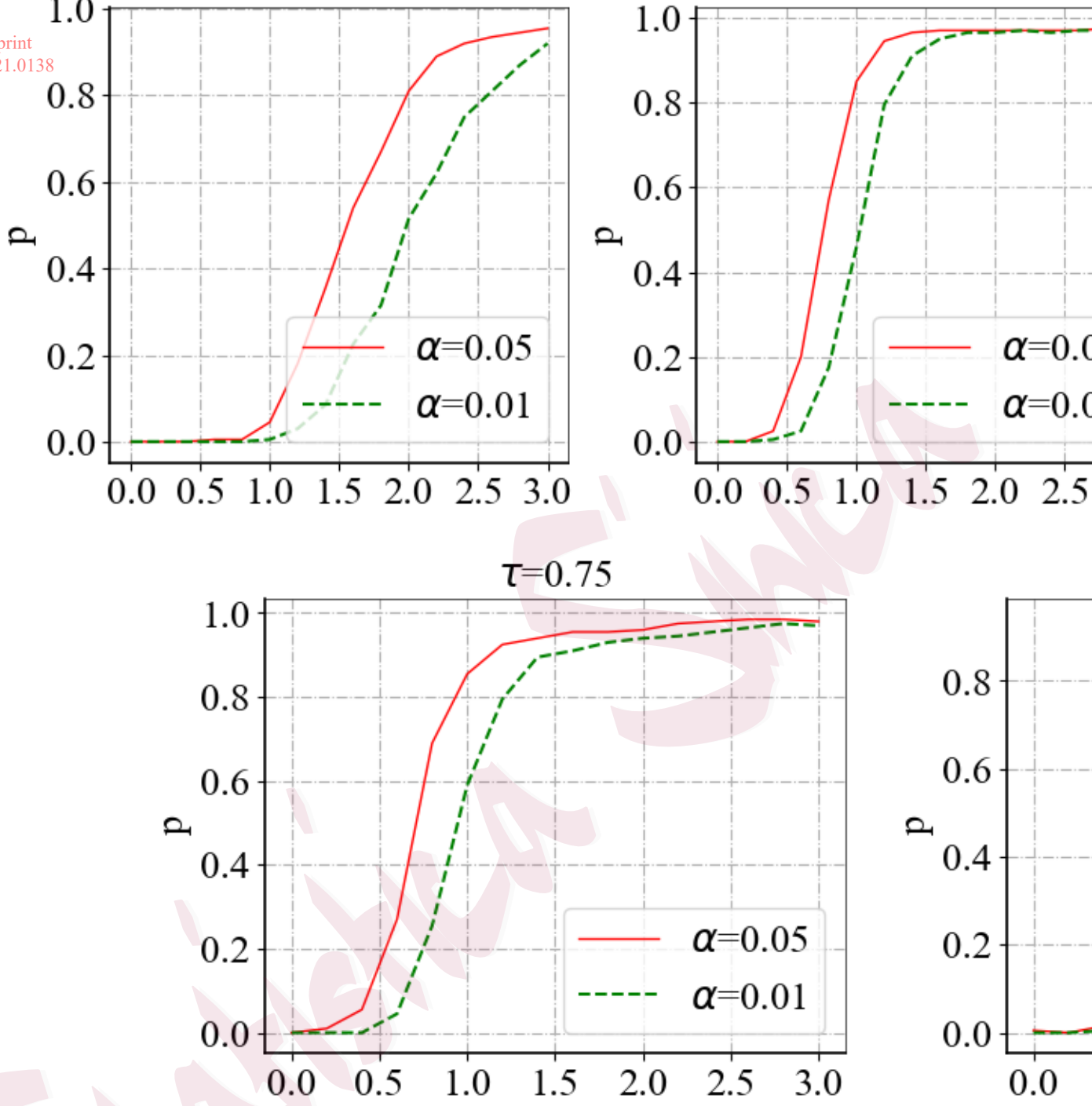


Figure 1: Plot of power curves. Rejection rate of $T_n^*$ based on wild bootstrap methods at $c = 0, 0.2, 0.4, \cdots, 3.0$ for different quantile levels at significance levels $\alpha = 0.05$ (real) and 0.01 (dash).

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

Liu et al. (2020b) are slightly large, resulting in empirical CPs that are significantly greater than the given nominal confidence levels and wider SCBs. We do not know why this is. The SCBs obtained using the LI method may need to be corrected because of linear interpolation. For comparison, we control the empirical CPs of the LI method to the given nominal confidence levels by removing some extreme values generated by the bootstrap method. We set $n = 500$ and $m = 50, 70, 90$ for $\tau = 0.10, 0.25, 0.50, 0.75$ and $0.90$, perform 500 runs, and take $B = 500$ bootstrap samples for each combination. We consider the confidence levels 95% and 99%, and list the simulated empirical coverage probabilities and coverage widths in Tables 2–3, yielding the following findings. (1) Our KS method has a much smaller RMISE than that of the LI method. (2) At all noise levels, the coverage probabilities are close to the nominal levels for both methods. (3) The coverage widths of the SCBs of our KS method become narrower as the number of grid points $m$ increases, whereas those obtained using the LI method become much wider. That is, we have tighter SCBs when there are more sampling locations. For example, when $m = 90$, our KS method clearly outperforms the LI method, because our method has tighter SCBs. (4) For each approach and each regression function, the coverage widths become narrower as the quantile level gets closer to the median. At the extreme quantile

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

levels $\tau = 0.10$ and 0.90, our KS method also performs well. However, the LI method has a very wide coverage width for the given confidence level. In addition, to illustrate the good performance of our KS method, we plot the SCBs of $\beta_i(s)$ $(l = 1, 2, 3)$ for fixed $\tau = 0.5$ at confidence levels 95% and 99% in Figure 2. The results again show that the widths of the SCBs become narrower as the number of grid points $m$ increases. Figure 3 compares the SCBs of the KS and the LI for $\beta_i(s)$ $(l = 1, 2, 3)$ at the extreme quantile level $\tau = 0.90$ and confidence level 95%. The figures for other cases are similar. In summary, our KS method clearly outperforms the LI method by having a much smaller RMISE and much tighter SCBs, especially for larger sample locations.

## 6. Diffusion tensor imaging data analysis

We analyze imaging data from a neonatal project on early brain development. The data set consists of 128 healthy infants, including 75 males and 53 females. The gestational ages of the infants range from 262 to 433 days. The diffusion tensor imaging and T1-weighted images were acquired for each subject. We have included the imaging preprocessing steps in the Supplementary Material.

We consider two diffusion properties: fractional anisotropy and mean

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

Table 2: Simulation results for the empirical coverage probability (CP) and coverage width (CW) of the 95% SCBs for the KS and LI methods. CWs are given in parentheses.

| | | $m = 50$ | | | $m = 70$ | | | $m = 90$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Method | $\tau$ | $\beta_1(s)$ | $\beta_2(s)$ | $\beta_3(s)$ | $\beta_1(s)$ | $\beta_2(s)$ | $\beta_3(s)$ | $\beta_1(s)$ | $\beta_2(s)$ | $\beta_3(s)$ |
| KS | 0.10 | 0.972 (0.581) | 0.978 (0.825) | 0.962 (0.829) | 0.964 (0.493) | 0.970 (0.701) | 0.962 (0.707) | 0.990 (0.481) | 0.986 (0.681) | 0.976 (0.683) |
| | 0.25 | 0.962 (0.518) | 0.968 (0.738) | 0.944 (0.735) | 0.958 (0.439) | 0.966 (0.624) | 0.956 (0.626) | 0.972 (0.426) | 0.968 (0.605) | 0.952 (0.608) |
| | 0.50 | 0.950 (0.497) | 0.950 (0.707) | 0.932 (0.706) | 0.954 (0.421) | 0.962 (0.601) | 0.950 (0.601) | 0.968 (0.408) | 0.968 (0.580) | 0.960 (0.583) |
| | 0.75 | 0.954 (0.519) | 0.954 (0.739) | 0.948 (0.739) | 0.930 (0.437) | 0.956 (0.623) | 0.942 (0.624) | 0.972 (0.427) | 0.972 (0.605) | 0.960 (0.608) |
| | 0.90 | 0.964 (0.587) | 0.980 (0.834) | 0.970 (0.837) | 0.956 (0.494) | 0.972 (0.701) | 0.970 (0.704) | 0.986 (0.480) | 0.980 (0.681) | 0.968 (0.684) |
| LI | 0.10 | 0.928 (0.602) | 0.980 (0.863) | 0.968 (0.862) | 0.938 (0.629) | 0.972 (0.898) | 0.970 (0.900) | 0.966 (0.650) | 0.990 (0.926) | 0.986 (0.928) |
| | 0.25 | 0.926 (0.449) | 0.946 (0.654) | 0.956 (0.652) | 0.952 (0.467) | 0.966 (0.679) | 0.974 (0.681) | 0.950 (0.479) | 0.972 (0.698) | 0.974 (0.701) |
| | 0.50 | 0.944 (0.406) | 0.962 (0.594) | 0.968 (0.592) | 0.956 (0.421) | 0.970 (0.617) | 0.968 (0.616) | 0.944 (0.433) | 0.966 (0.632) | 0.968 (0.633) |
| | 0.75 | 0.968 (0.450) | 0.960 (0.654) | 0.976 (0.653) | 0.936 (0.467) | 0.958 (0.680) | 0.958 (0.680) | 0.938 (0.478) | 0.964 (0.697) | 0.956 (0.699) |
| | 0.90 | 0.942 (0.603) | 0.970 (0.864) | 0.968 (0.864) | 0.946 (0.628) | 0.984 (0.901) | 0.974 (0.901) | 0.970 (0.649) | 0.982 (0.929) | 0.982 (0.930) |

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

Table 3: Simulation results for the empirical coverage probability (CP) and coverage width (CW) of the 99% SCBs for the KS and LI methods. CWs are given in parentheses.

| | | $m = 50$ | | | $m = 70$ | | | $m = 90$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Method | $\tau$ | $\beta_1(s)$ | $\beta_2(s)$ | $\beta_3(s)$ | $\beta_1(s)$ | $\beta_2(s)$ | $\beta_3(s)$ | $\beta_1(s)$ | $\beta_2(s)$ | $\beta_3(s)$ |
| KS | 0.10 | 0.994 (0.738) | 0.996 (1.049) | 0.990 (1.056) | 0.992 (0.622) | 0.992 (0.887) | 1.000 (0.892) | 0.998 (0.608) | 0.992 (0.863) | 0.992 (0.864) |
| | 0.25 | 0.986 (0.658) | 0.988 (0.937) | 0.978 (0.936) | 0.992 (0.554) | 0.990 (0.792) | 0.988 (0.794) | 0.990 (0.539) | 0.988 (0.767) | 0.996 (0.772) |
| | 0.50 | 0.986 (0.629) | 0.986 (0.898) | 0.978 (0.900) | 0.988 (0.533) | 0.986 (0.758) | 0.988 (0.761) | 0.996 (0.519) | 0.990 (0.735) | 0.992 (0.738) |
| | 0.75 | 0.992 (0.659) | 0.988 (0.942) | 0.984 (0.943) | 0.988 (0.555) | 0.988 (0.788) | 0.992 (0.789) | 0.992 (0.540) | 0.988 (0.763) | 0.992 (0.769) |
| | 0.90 | 0.992 (0.743) | 0.992 (1.056) | 0.994 (1.063) | 0.986 (0.626) | 0.994 (0.889) | 0.994 (0.891) | 0.994 (0.607) | 0.986 (0.861) | 0.990 (0.866) |
| LI | 0.10 | 0.984 (0.680) | 0.992 (0.966) | 0.990 (0.964) | 0.990 (0.708) | 0.996 (1.000) | 0.996 (1.001) | 0.996 (0.771) | 1.000 (1.080) | 0.998 (1.082) |
| | 0.25 | 0.972 (0.497) | 0.992 (0.726) | 0.986 (0.723) | 0.986 (0.514) | 0.992 (0.749) | 0.992 (0.751) | 0.996 (0.551) | 1.000 (0.803) | 0.998 (0.807) |
| | 0.50 | 0.980 (0.448) | 0.984 (0.658) | 0.998 (0.656) | 0.988 (0.462) | 0.992 (0.677) | 0.990 (0.678) | 0.994 (0.494) | 0.992 (0.725) | 0.996 (0.726) |
| | 0.75 | 0.988 (0.498) | 0.990 (0.726) | 0.996 (0.724) | 0.978 (0.514) | 0.992 (0.750) | 0.992 (0.751) | 0.994 (0.551) | 0.996 (0.802) | 0.996 (0.805) |
| | 0.90 | 0.984 (0.681) | 0.998 (0.968) | 0.998 (0.969) | 0.990 (0.707) | 0.998 (1.004) | 0.988 (1.004) | 0.996 (0.771) | 0.998 (1.083) | 1.000 (1.086) |

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

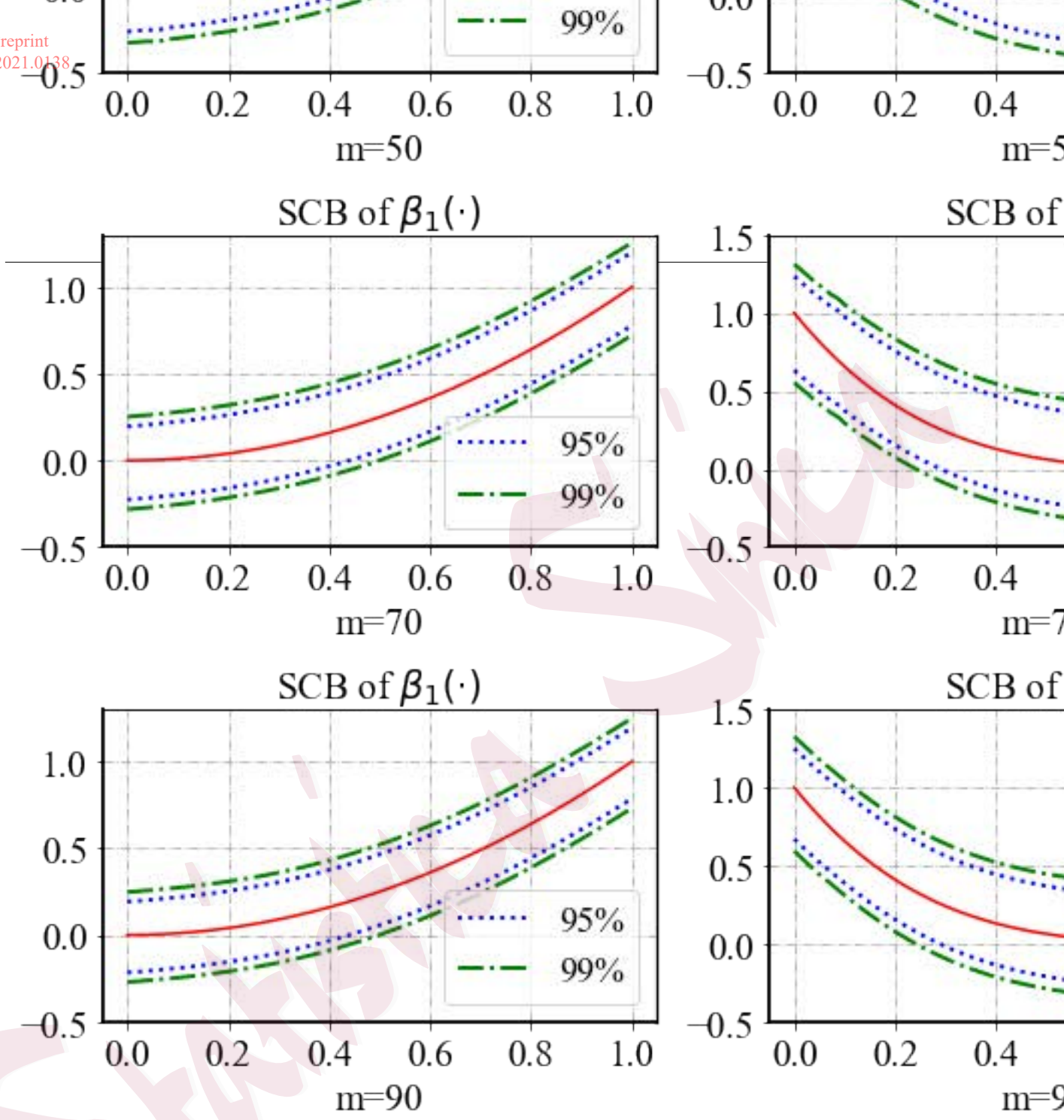


Figure 2: The SCB of three coefficient functions for $\tau = 0.5$ when the nominal levels are 95% and 99%, based on the KS method. The solid curves are the true coefficient functions, and the dotted curves and the dash-dot curves are the confidence bands of the nominal levels 95% and 99%, respectively.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

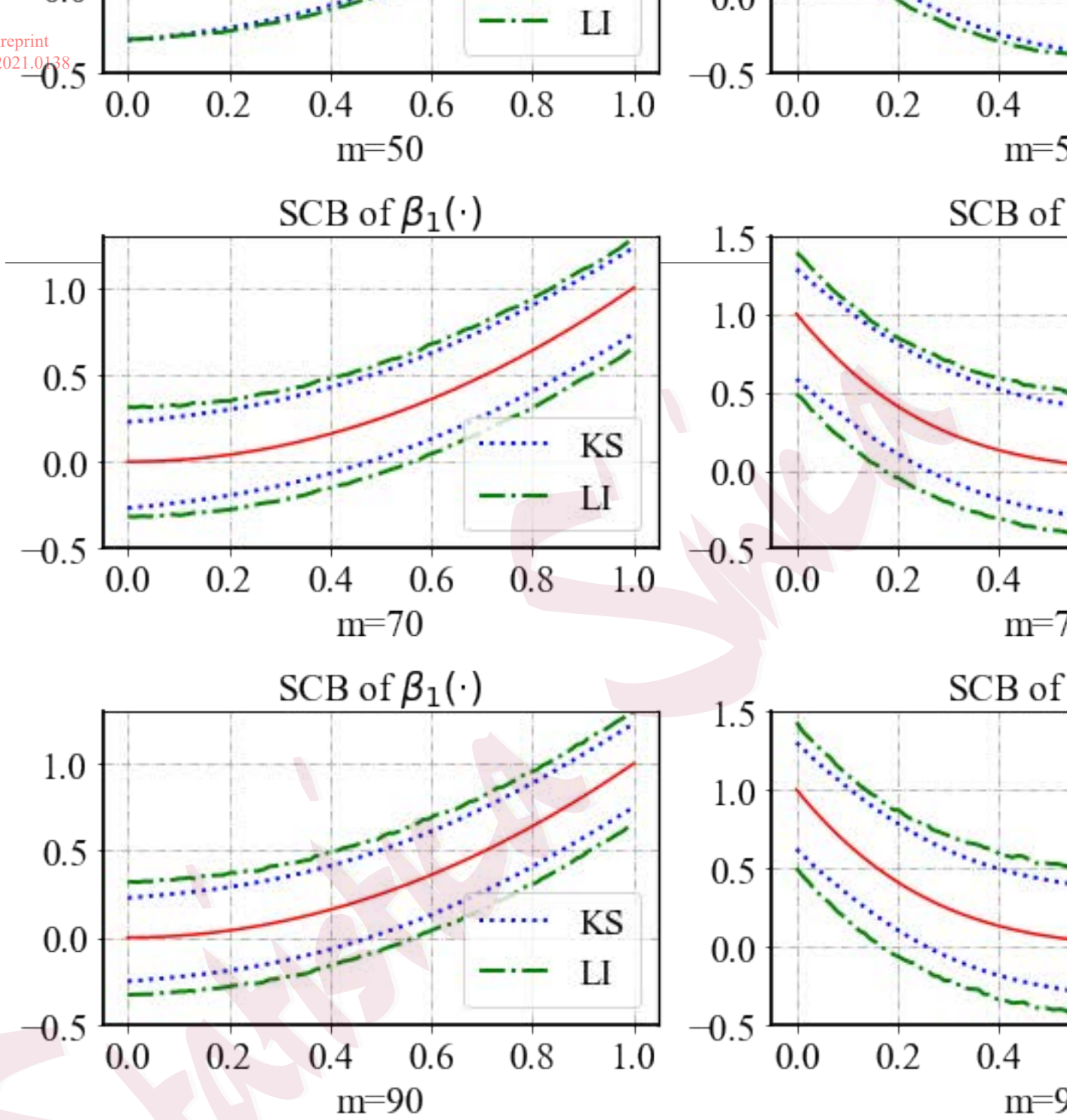


Figure 3: Comparisons of the SCBs of KS and LI for three coefficient functions at the quantile level $\tau = 0.90$ and nominal level 95%. The solid curves are the true coefficient functions, and the dotted curves and dash-dot curves are the confidence bands of the KS and the LI methods, respectively.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

diffusivity, which are measured at 45 grid points along the genu tract of the corpus callosum. They measure the inhomogeneous extent of local barriers to water diffusion and the average magnitude of local water diffusion, respectively. At different quantile levels $\tau$, we use magnetic resonance imaging scanning to explore the effects of an infant's gender and gestational age on the fractional anisotropy and mean diffusivity, and to delineate the tendency of fiber diffusion properties over time. For $\tau = 0.25, 0.5, 0.75$, we fit our model by taking the fractional anisotropy and mean diffusivity values as responses, and gender and age as covariates. We also include an intercept term. The age, fractional anisotropy, and mean diffusivity values are standardized before fitting the model (1.2). For each given $\tau$, we estimate the coefficient function $\beta(s) = \{\beta_0(s), \beta_1(s), \beta_2(s)\}^{\mathrm{T}}$, compute $T_n^*$ for each hypothesis test, and obtain the $p$-value by applying wild bootstrap procedures with $B = 500$ replications.

For the given quantile levels, Figure 4 plots the estimated coefficient functions corresponding to the intercept, gender, and age associated with fractional anisotropy [panels (a), (b), (c)] and mean diffusivity [panels (d), (e), (f)], and the $p$-values of the global test statistics under the null hypothesis $H_{0l} : \beta_l(s) = 0$ $(l = 1, 2, 3)$. The intercept functions $\beta_0(s)$ [panels (a) and (d)] give the overall tendency, and are significantly grid-point varying

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

at all quantiles considered at the 0.01 level. In addition, from panel (d), we see that the intercept function corresponding to the mean diffusivity in the lower quantile $\tau = 0.25$, that is, for the group of healthy infants with lower mean diffusivity, is negative, whereas that corresponding to the group of healthy infants with higher mean diffusivity is positive. These negative and positive effects would not have been revealed by methods focusing only on the conditional mean. For gender effects [panels (b) and (e)], the six $p$-values are all greater than 0.01, except for the mean diffusivity when $\tau = 0.5$. This is consistent with the findings in Zhu et al. (2012) that gender has a weakly significant effect on fractional anisotropy and mean diffusivity. Male infants have a relatively bigger average local water diffusion along the genu tract of the corpus callosum compared with that of female infants for first five grid points when $\tau = 0.5$. For the gestational age effects [panels (c) and (f)], the $p$-value is greater than 0.01 for the fractional anisotropy, but is less than 0.01 for the mean diffusivity response. This indicates that gestational age has a weakly significant effect on the fractional anisotropy along the genu tract of the corpus callosum, and that the mean diffusivity changes noticeably with gestational age.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

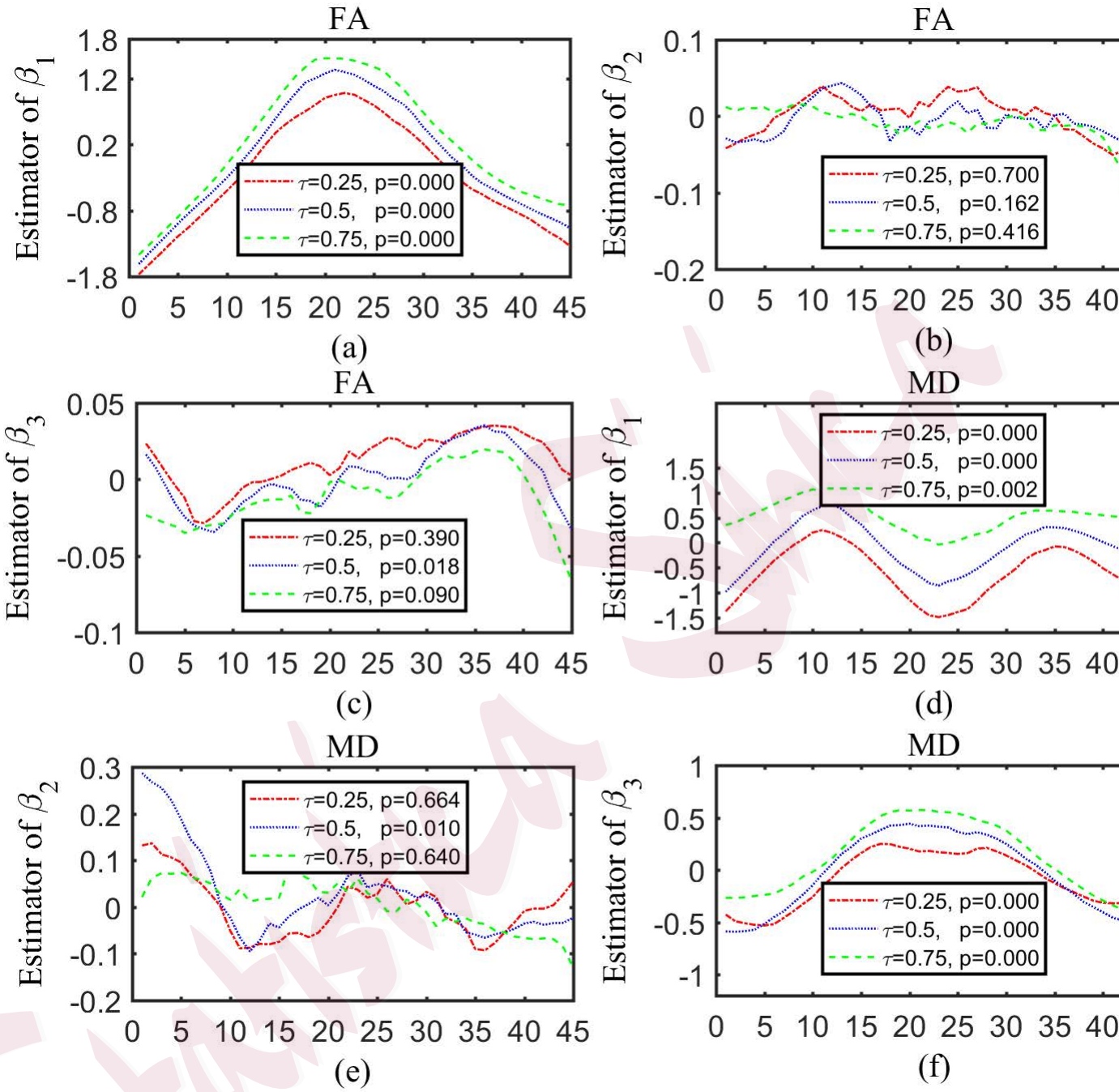


Figure 4: Plot of estimated coefficient functions and $p$-values of the global test statistics for the quantile levels $\tau = 0.25$, 0.5, and 0.75 for diffusion tensor imaging data. "FA" denotes fractional anisotropy, and "MD" denotes mean diffusivity.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

## 7. Discussion

We have proposed a functional response quantile regression model that explicitly characterizes the conditional distribution of a functional response given a set of scalar predictors. We have developed a global test statistic for linear hypotheses of the varying coefficient functions, and constructed an asymptotic SCB for each regression coefficient function. Simulations and a real-data analysis show that the FRQR characterizes the effect of scalar predictors on the functional responses at different quantile levels.

We focus on densely observed functional data in this study. However, in practice, functional data may be irregularly and sparsely observed. Our method is not directly applicable to this case because the strong approximation results used to construct the nonparametric confidence bands, commonly known as "Hungarian embedding," cannot be applied to irregular and sparse functional data. To the best of our knowledge, only pointwise confidence bands have been developed for irregular and sparse functional data (Yao et al., 2005a,b; Yao, 2007; Ma et al., 2012; Zheng et al., 2014; Şentürk and Müller, 2010). In future work, it would be interesting to construct SCBs for such data, and to develop the theory to support the procedure with tuning parameters selected adaptively.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138



## Supplementary Material

The online Supplementary Material includes all proofs and technical details.

## Acknowledgments

Dr. Zhou was supported by the National Social Science Fund of China (19BTJ034) and the National Natural Science Foundation of China (12171242, 11971235). We are grateful to the editor, associate editor, and anonymous referees for their many helpful comments and suggestions. Dr. Kong, Dr. Kashlak, Dr. Kong, and Dr. Karunamuni were supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). Dr. Kong was also supported by Canada Research Chair in Statistical Learning and Canada CIFAR AI Chair. Dr. Zhus work was partially supported by NIH grants R01MH086633 and R01MH116527.

## References

Akritas, M. and I. V. Keilegom (2001). Nonparametric estimation of the residual distribution. *Scand. J. Statist. 28*, 549–567.

Belloni, A., V. Chernozhukov, D. Chetverikov, and I. Fernández-Val (2019). Conditional quantile processes based on series or many regression. *Journal of Econometrics 213(1)*, 4–29.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138

# REFERENCES

Cai, Z. and X. Xu (2009). Nonparametric quantile estimations for dynamic smooth coefficient models. *J. Am. Statist. Assoc. 104*, 371–383.

Chen, K. and H. Müller (2012). Conditional quantile analysis when covariates are functions with application to growth data. *J. R. Statist. Soc. B 74*, 1–23.

Şentürk, D. and H.-G. Müller (2010). Functional varying coefficient models for longitudinal data. *Journal of the American Statistical Association 105*, 1256–1264.

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

Fan, J., Q. Yao, and H. Tong (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika 83*, 189–206.

Fan, J. and T. Yim (2004). A data-driven method for estimating conditional densities. *Biometrika 91*, 819–834.

Garrido, M. I., M. Sahani, and R. J. Dolan (2013). Outlier responses reflect sensitivity to statistical structure in the human brain. *PLoS Comp. Biology 9*(3), e1002999.

Hall, P., J. Racine, and Q. Li (2004). Cross-validation and the estimation of conditional probability densities. *J. Am. Statist. Assoc. 99*, 1015–1026.

Hastie, T. and R. Tibshirani (1993). Varying-coefficient models. *J. R. Statist. Soc. B 55*, 757–796.

Hendricks, W. and R. Koenker (1992). Hierarchical spline models for conditional quantiles and

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138



the demand for electricity. *J. Am. Statist. Assoc. 87*, 58–68.

Kato, K. (2012). Estimation in functional linear quantile regression. *Ann. Statist. 40*, 3108–3136.

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

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

Kosorok, M. (2008). *Introduction to Empricial Processes and Semiparametric Inference*. New York: Springer.

Kosorok, M. R. (2003). Bootstraps of sums of independent but not identically distributed stochastic processes. *J. Mult. Anal. 84*, 299–318.

Krauledat, M., G. Dornhege, B. Blankertz, and K.-R. Müller (2007). Robustifying eeg data analysis by removing outliers. *Chaos and Compl. Letters 2*(3), 259–274.

Liu, Y., M. Li, and J. Morris (2020a). Function-on-scalar quantile regression with application to mass spectrometry proteomics data. *Annal of Applied Statistics 14*, 521–541.

Liu, Y., M. Li, and J. Morris (2020b). On function-on-scalar quantile regression. *arXiv 2002.03355v1*.

Ma, S., L. Yang, and R. J. Carroll (2012). A simultaneous confidence band for sparse longitudinal regression. *Statist. Sinica 22*, 95–122.

Müller, U. U., A. Schick, and W. Wefelmeyer (2007). Estimating the error distribution function in semiparametric regression. *Statist. Decisions 25*, 1–18.

Neumeyer, N. and I. V. Keilegom (2010). Estimating the error distribution in nonparametric

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138



multiple regression with applications to model testing,. *J. Mult. Anal. 101*, 1067–1078.

Ramsay, J. O. and C. J. Dalzell (1991). Some tools for functional data analysis (with discussion). *J. R. Statist. Soc. B 53*, 539–572.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis (2nd ed)*. New York: Springer.

Rice, J. A. and B. W. Silverman (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Statist. Soc. B 53*, 233–243.

Shen, Q. and J. Faraway (2004). An $f$ test for linear models with functional responses. *Statist. Sinica 14*, 1239–1257.

Wang, H., Z. Zhu, and J. Zhou (2009). Quantile regression in partially linear varying coefficient models. *Ann. Statist. 37*, 3841–3866.

Wang, J.-L., J.-M. Chiou, and H.-G. Müller (2016). Functional data analysis. *Annual Review of Statistics and Its Application 3*, 257–295.

Yang, H., V. Baladandayuthapani, A. Rao, and J. Morris (2020). Quantile function on scalar regression analysis for distributional data. *Journal of the American Statistical Association 115*, 90–106.

Yao, F. (2007). Asymptotic distributions of nonparametric regression estimators for longitudinal or functional data. *J. Mult. Anal. 98*, 40–56.

Yao, F., H. G. Müller, and J. L. Wang (2005a). Functional data analysis for sparse longitudinal

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138



data. *J. Am. Statist. Assoc. 100*, 577–590.

Yao, F., H. G. Müller, and J. L. Wang (2005b). Functional linear regression analysis for longitudinal data. *Ann. Statist. 33*, 2873–2903.

Zhang, J. T. (2011). Statistical inference for linear models with functional responses. *Statist. Sinica 21*, 1431–1451.

Zhang, J. T. and J. Chen (2007). Statistical inferences for functional data. *Ann. Statist. 35*, 1052–1079.

Zhang, Z., X. Wang, L. Kong, and H. Zhu (2021). High-dimensional spatial quantile function-on-scalar regression. *Journal of the American Statistical Association doi.org/10.1080/01621459.2020.1870984*, 1–16.

Zheng, S., L. Yang, and W. Härdle (2014). A smooth simultaneous confidence corridor for the mean of sparse functional data. *Journal of the American Statistical Association 109*, 661–673.

Zhu, H. T., J. G. Ibrahtm, N. Tang, D. B. Rowe, X. Hao, R. Bansal, and B. S. Peterson (2007). A statistical analysis of brain morphology using wild bootstrapping. *IEEE Trans. Med. Imaging 26*, 954–966.

Zhu, H. T., R. Li, and L. Kong (2012). Multivariate varying coefficient model for functional responses. *Ann. Statist. 40*, 2634–2666.

Xingcai Zhou, School of Statistics and Data Science, Nanjing Audit University, Nanjing 211815,

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0138



China

E-mail: xczhou@nuaa.edu.cn

Dehan Kong, Department of Statistical Sciences, University of Toronto, Toronto, Ontario M5S 3G3, Canada

E-mail: kongdehan@utstat.toronto.edu

A. B. Kashlak, Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1, Canada

E-mail: kashlak@ualberta.ca

Linglong Kong, Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1, Canada

E-mail: lkong@ualberta.ca

R. Karunamuni, Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1, Canada

E-mail: r.j.karunamuni@ualberta.ca

Hongtu Zhu, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27516, U.S.A.

E-mail: hzhu@bios.unc.edu