Statistica Sinica Preprint No: SS-2022-0206	
Title	Asymmetric Estimation for Varying-Coefficient Additive Model with Functional Response in Reproducing Kernel Hilbert Space
Manuscript ID	SS-2022-0206
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0206
Complete List of Authors	Yi Liu, Wei Tu, Yanchun Bao, Bei Jiang and Linglong Kong
Corresponding Authors	Linglong Kong
E-mails	lkong@ualberta.ca

Asymmetric estimation for varying-coefficient additive model with functional response in reproducing kernel Hilbert space

Yi Liu[1], Wei Tu[2], Yanchun Bao[3], Bei Jiang[4], Linglong Kong[4]

[1] *York University,* [2] *Queen's University,* [3] *University of Essex and* [4] *Univertsity of Alberta*

Abstract: Function-on-scalar regression models are extensively utilized in applications involving longitudinal or functional responses. Prior literature has established the minimax optimal bounds for both mean and quantile regression. This paper explores expectile regression as a natural extension to mean regression, particularly for modeling potential heteroscedasticity in data. We propose an expectile function-on-scalar regression model that focuses on asymmetrical regression of functional responses based on scalar predictors. Employing the structure of Reproducing Kernel Hilbert Space (RKHS), we have developed a statistically efficient expectile estimator. This estimator comes with theoretical backing, derived from the minimax rates of convergence in both random and fixed design contexts. Our extensive simulations demonstrate the robust performance of the proposed methods across various settings. Additionally, we present an empirical analysis using quality of life data from a breast cancer clinical trial, showcasing the practical utility of our method.

Key words and phrases: additive model, functional regression, reproducing kernel Hilbert space, quality of life, cancer clinical trials.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0206

1. Introduction

Functional data analysis (FDA) has drawn significant attention from researchers from various fields due to its unique ability to deal with intrinsically infinite-dimensional data and processes defined on continuous domains. Data with such structure are widely available in medicine, biology, public health, environmental science and AI. Well-know monographs include Ramsay and Silverman (2005), Ferraty and Vieu (2006), Ramsay and Silverman (2007) and Zhang (2014). In this paper, we focus on the following varying-coefficient model with functional response:

$$y_i(s) = x_i^T \boldsymbol{\beta}(s) + \varepsilon_i(s),$$

where $y_i(s)$ is a functional response, x_i are associated p-dimensional covariates, $\boldsymbol{\beta}(s)$ is $p \times 1$ functional coefficient and $\varepsilon_i(s)$ is the measurement error.

Prior studies have yielded a vast literature on these varying-coefficient models. Pioneering works can trace back to Breiman and Friedman (1985) and regression on such models are systematically discussed in Hastie et al. (1993), Hastie and Tibshirani (1993) and Cleveland and Grosse (1991). More recently, Zhu et al. (2012) proposed a multivariate varying coefficient model and established a uniform convergence rate of the estimated

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0206

covariance function and its associated eigenvalue and eigenfunctions. On the other hand, driven by the need to extract more enriched information from the conditional response distribution, researchers are no longer satisfied with focusing on a single value (mean) of responses. To move beyond characterizing one single value (mean or median), one possible and intuitive generalization is quantiles, the inverse of the cumulative distribution function.

According to Stigler (1984), the history of interest in regression of median can trace back to 1760, and quantile regression hasn't been reintroduced until Koenker and Bassett Jr (1978). Starting from Koenker and Hallock (2001) extended smoothing spline method to quantiles, the last decades have seen extraordinary advances in quantile regression in the context of functional data analysis. Li et al. (2007) proposed an RKHS approach to estimate conditional quantile functions. Kato et al. (2012) studied quantile regression for scalar dependent variable and functional covariate by the plug-in method. Similar functional covariate case is also studied in Chen and Müller (2012) and Cardot et al. (2005). On the other hand, Function-on-scalar regression Functional responses regressed on scalar predictors have been studied by Reiss et al. (2010), Chen et al. (2016), Goldsmith et al. (2015) and Goldsmith and Kitago (2016).

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0206

In this paper, we propose an expectile function-on-scalar regression model, which represents a shift from traditional symmetric estimation methods primarily focused on modeling the central tendency of data, such as the median or mean. Asymmetric estimation, by contrast, targets either the higher or lower end of the response distribution. This approach is especially beneficial in capturing data aspects not adequately represented by central tendency measures. For example, while symmetric estimation might evaluate the median lifespan of a demographic group, asymmetric models excel in highlighting the life spans of individuals in less favorable conditions—outcomes that may elude explanation solely by covariates. In economic analysis, asymmetric estimation sheds light on financial returns under varying conditions, thereby aiding in the formulation of either conservative or aggressive strategic decisions.

One possible approach of asymmetric estimation is quantile regression, as explored in recent works by Wang (2016) and Zhang et al. (2021), Despite the natural interpretability and robustness of quantile regression, it encounters several challenges. Quantile regression is based on minimizing the check loss function, which is not strongly convex and lacks differentiability at the origin. These properties necessitate additional efforts to address problems stemming from the non-smooth nature of the check loss function.

Furthermore, the robustness of quantiles, while advantageous, also implies a diminished sensitivity to extreme observations and losses.

In response to these limitations, expectile regression was introduced as an alternative by Aigner et al. (1976) and Newey and Powell (1987). This regression technique extends beyond the scope of traditional quantile approaches, offering a nuanced view of the data distribution's tails. Our study leverages this methodology to explore functional responses against scalar predictors. Specifically, our focus is on the scenario where each component of the true parameter $\boldsymbol{\beta}_0$ is defined within a compact subset $\mathcal{T}$ of $\mathbb{R}$ and is part of an RKHS $\mathcal{H}$ with a continuous kernel. In this work, we simplify the discussion by assuming $\mathcal{T} = [0, 1]$. We observe the response functions y_i through sampling points, meaning our dataset comprises pairs $(x_i, y_i(\mathcal{T}_{ij}))$ for each subject i at locations $\mathcal{T}_{ij}, i = 1, \ldots, n$ and $j = 1, \ldots, m$. Two design approaches are considered: a random design, where $\mathcal{T}_{ij}$ are uniformly and independently sampled from $\mathcal{T}$, and a fixed design, where each subject's response functions are observed at identical locations ($\mathcal{T}_{1j} = \ldots = \mathcal{T}_{nj} := \mathcal{T}_j$). We assume that $0 = \mathcal{T}_1 < \cdots < \mathcal{T}_m = 1$. The primary objective of this paper is to construct efficient functional estimators within this framework and demonstrate their rate optimality.

The rest of this paper is organized as follows: In Section 2, we give our construction of Penalized RKHS estimator and establish the optimal convergence rate of the error in estimating $\boldsymbol{\beta}$ under both fixed and random designs. In Section 3, extensive numerical studies were presented to illustrate the effectiveness of the proposed estimator. Section 4 provides a data analysis using quality of life data from a breast cancer clinical trial. Section 5 concludes the paper.

2. Methodology

2.1 Penalized RKHS estimator

In this paper, we assume that each component of $\boldsymbol{\beta}$ resides in the Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}$. Consequently, any vector $\boldsymbol{\beta} = (\boldsymbol{\beta}^1, \cdots, \boldsymbol{\beta}^p)^T$ forms another RKHS, denoted as $\oplus_{i=1}^p \mathcal{H}$ or $\mathcal{H}^p$. The inner product in $\mathcal{H}^p$ is defined as $\langle (\boldsymbol{\alpha}^1, \cdots, \boldsymbol{\alpha}^p)^T, (\boldsymbol{\beta}^1, \cdots, \boldsymbol{\beta}^p)^T \rangle_{\oplus_{i=1}^p \mathcal{H}} = \sum_{i=1}^p \langle \boldsymbol{\alpha}^i, \boldsymbol{\beta}^i \rangle_{\mathcal{H}}$. Accordingly, the induced norm $\|\boldsymbol{\beta}\|_{\mathcal{H}^p}$ is given by $\sqrt{\sum_{i=1}^p \|\boldsymbol{\beta}^i\|_{\mathcal{H}}^2}$.

Similar notations are applied to L^2 spaces. We define the integral over the product space $\mathcal{T}^p$ as $\int_{\mathcal{T}^p} f(t)dt = \int_{\mathcal{T}} \mathbb{1}^T f(t)dt$. The notation $\|f\|_2^2$ signifies $\int_{\mathcal{T}} [f(t)]^2 dt$ for real-valued functions and $\int_{\mathcal{T}^p} [f(t)]^2 dt$ for functions valued in $\mathbb{R}^p$.

By Mercer's theorem Riesz and Nagy (1955), for each $\mathcal{H}$, we can find corresponding eigen-functions $\{\phi_k\}_{k\geq 1} \subset \mathcal{H}$ and eigen-values $\{\nu_k\}_{k\geq 1} \subset \mathbb{R}$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0206

such that any function in $\mathcal{H}$ has the following eigen decomposition:

$$f = \sum_{v \geq 1} f_v \phi_v,$$

satisfying,

$$\|f\|_2^2 = \sum_{v \geq 1} f_v^2 \text{ and } \|f\|_{\mathcal{H}}^2 = \sum_{v \geq 1} \frac{f_v^2}{\nu_v}.$$

The eigen-structures play important roles in determining the nature of convergence for functional linear regression problems. A detailed discussion of eigen-structures is beyond the scope of this paper. Interested readers may refer to Cai et al. (2006) and Hall et al. (2007) for more information.

The eigen-structures of $\mathcal{H}^p$ can be derived from the eigen-structures of $\mathcal{H}$. To see this, first, let us define δ_{ij} as the Kronecker delta, which is 1 if $i = j$ and zero otherwise. Now, define

$$\begin{aligned} \varphi_{ip+j-p} : \mathcal{T} &\longrightarrow \mathbb{R}^p \\ \Psi \quad &\quad\ \ \Psi \\ x &\longmapsto \phi_i(x)(\delta_{1j}, \cdots, \delta_{pj}) \end{aligned},$$

and

$$\rho_{ip+j-p} = \nu_i \text{ for } i \geq 1 \text{ and } 1 \leq j \leq p.$$

It's easy to verify that $\{\varphi_n\}_{n \geq 1}$ and $\{\rho_n\}_{n \geq 1}$ are eigen-functions and eigen-values of $\mathcal{H}^p$.

We propose the following Penalized expectile RKHS estimator:

Definition 1. *(Penalized expectile RKHS estimator) Given space $\mathcal{H}$, regulation parameter λ sampling points $\{\mathcal{T}_{ij}\}$ and observations $\{x_i\}$, $\{y_i(\mathcal{T}_{ij})\}$,* *then Penalized expectile RKHS estimator is defined by:*

$$\hat{\boldsymbol{\beta}}_\lambda = \underset{\hat{\boldsymbol{\beta}} \in \mathcal{H}^p}{\arg\min} \left\{ \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \rho \Big[y_i(\mathcal{T}_{ij}) - x_i^T \hat{\boldsymbol{\beta}}(\mathcal{T}_{ij}) \Big] + \lambda \|\hat{\boldsymbol{\beta}}\|^2_{\mathcal{H}^p} \right\}.$$

where $\rho(z) = \rho^e_\tau(z) = |\tau - \mathrm{I}(z < 0)| \cdot z^2$. We omit the expectile level later τ for simplicity.

2.2 Methods for numerical estimation

The estimator introduced in Definition 1 can be efficiently computed using the gradient descent method. Define $\mathbb{S} = \{S_1, \cdots, S_{|\mathbb{S}|}\}$ as the set of all sample points $\{\mathcal{T}_{ij}\}$, where $|\mathbb{S}|$ denotes the number of elements in $\mathbb{S}$. It is crucial to note that $|\mathbb{S}| \le mn$, which takes into account potential duplications in $\{\mathcal{T}_{ij}\}$. According to the representation theorem Wahba (1990), the optimal $\hat{\boldsymbol{\beta}}$ is confined to a finite-dimensional subspace of $\mathcal{H}^p$, simplifying computation. The estimator $\widehat{\boldsymbol{\beta}}^i$ is expressed as

$$\widehat{\boldsymbol{\beta}}^i = \sum_{j=1}^{|\mathbb{S}|} b_{ij} K(\cdot, S_j),$$

for $i = 1, \ldots, p$, where K is the kernel of the RKHS and b_{ij} are coefficients.

The loss function is formulated as

$$L(\boldsymbol{\beta}) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \rho \Big[y_i(\mathcal{T}_{ij}) - x_i^T \boldsymbol{\beta}(\mathcal{T}_{ij}) \Big] + \lambda \|\boldsymbol{\beta}\|^2_{\mathcal{H}^p}.$$

which can be rewritten as

$$\frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \rho \left[y_i(\mathcal{T}_{ij}) - \sum_{k=1}^{m} \sum_{r=1}^{|\mathbb{S}|} x_{ik}\, K(\mathcal{T}_{ij}, S_r)\, b_{kr} \right] + \lambda \sum_{j=1}^{p} \sum_{k=1}^{|\mathbb{S}|} \sum_{r=1}^{|\mathbb{S}|} b_{jk} b_{jr}\, K(S_r, S_k).$$

The first term of this function consists of a sum of convex functions ρ, while the second term is a positive definite quadratic form. The gradient of the loss function can be efficiently computed as

$$\frac{\partial L}{\partial b_{kr}} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} -x_{ik}K(\mathcal{T}_{ij}, S_r)\rho'\Bigg[y_i(\mathcal{T}_{ij}) - \sum_{k=1}^{m}\sum_{r=1}^{|\mathbb{S}|} x_{ik}\, K(\mathcal{T}_{ij}, S_r)\, b_{kr}\Bigg]$$
$$+ 2\lambda \sum_{t=1}^{|\mathbb{S}|} b_{kt}\, K(S_r, S_t).$$

This ensures that the overall function exhibits strong convexity with respect to b_{ij}, facilitating efficient computation of the estimator $\boldsymbol{\beta}$. The optimal $\hat{\boldsymbol{\beta}}$ can be effectively determined using gradient descent or Nesterov's accelerated gradient method Nesterov (1983).

The parameter λ plays a crucial role in regulating the complexity of the function $\boldsymbol{\beta}$ by penalizing its complexity as measured by the RKHS norm, thus preventing overfitting. While our theory provides guidelines for choosing λ, in practical applications, we recommend determining the suitable λ through cross-validation, as the exact smoothness condition might be unknown. This balanced approach between data fitting and smoothness ensures optimal performance of the estimator.

2.3 Optimal Rate of Convergence

In this section, we establish guarantees for the optimality of the Penalized expectile RKHS estimator by deriving the minimax rate of convergence for estimating $\boldsymbol{\beta}$. We begin by introducing the following assumptions:

(A1) The eigenvalues of $E\left(xx^T\right)$ are respectively bounded from below and above by a constant c and $1/c$.

(A2) Each $\epsilon_i(\mathcal{T}_{ij})$ has a zero τ expectile and a variance of $\sigma_0^2 < +\infty$.

Intuitively, condition (A1) ensures that the true value of $\boldsymbol{\beta}$ will have a non-negligible, yet bounded, impact on the observations, while condition (A2) guarantees that the observations have a well-defined mean and a finite variance. Under the random design, our results are formulated as follows.

Theorem 1. *Fix $\tau \in (0,1)$. Assume that each component of the true parameter $\boldsymbol{\beta}_0$ resides in a compact subset $\mathcal{B}_k$ of an RKHS $\mathcal{H}$, with eigenvalues ν_k of $\mathcal{H}$ satisfying $\nu_k \approx k^{-2r}$ for some $r > \frac{1}{2}$. Given that (A1) and (A2) hold, then under a random design:*

$$\lim_{a\to 0} \lim_{m,n\to\infty} \inf_{\hat{\boldsymbol{\beta}}} \sup_{\boldsymbol{\beta}\in\mathcal{B}_k^p} P\left(\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|_2^2 > a\left[(nm)^{-\frac{2r}{2r+1}} + \frac{1}{n}\right]\right) = 1.$$

where the infimum is taken over all possible estimators $\boldsymbol{\beta}_0$ based on the observations. The expression $\nu_k \approx k^{-2r}$ implies that there exist constants $0 < c_0 < c_1$ such that $c_0 k^{-2r} \leq \nu_k \leq c_1 k^{-2r}$.

Theorem 2. *Fix $\tau \in (0,1)$. Suppose each component of the true parameter $\boldsymbol{\beta}_0$ resides in a compact subset $\mathcal{B}_k$ of an RKHS $\mathcal{H}$ with eigenvalues $\{\nu_k\}$ of $\mathcal{H}$ satisfies $\nu_k \asymp k^{-2r}$ for some $r > \frac{1}{2}$. If (A1) and (A2) hold, then under random design:*

$$\lim_{a \to +\infty} \lim_{m,n \to \infty} \sup_{\boldsymbol{\beta} \in \mathcal{B}_k^p} P\left(\left\| \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right\|_2^2 > a\left[(nm)^{-2r/(2r+1)} + \frac{1}{n} \right] \right) = 0,$$

when $\hat{\boldsymbol{\beta}}$ is the Penalized expectile RKHS estimator with $\lambda \asymp (nm)^{-2r/(2r+1)}$.

The combination of these theorems demonstrates that the Penalized RKHS estimator is rate-optimal in this setting. The convergence rate depends on the RHKS itself (characterized by the parameter r) and the relative scale between m and n. When $m = O\left(n^{1/2r}\right)$, the term $1/n$ can be absorbed into $(nm)^{-2r/(2r+1)}$, and when $m = \omega\left(n^{1/2r}\right)$, the total error will be dominated by $1/n$ instead. This phase transition phenomenon is common in varying-coefficient functional regressions (Zhang et al. (2021), Cai and Weng (2016)).

Under the fixed design, we limit our discussion to the Sobolev spaces. The Sobolev space $\mathcal{W}_2^r([0,1])$ with order r is defined as

$$\mathcal{W}_2^r([0,1]) = \left\{ f : [0,1] \to \mathbb{R} \mid f^{(r)} \in \mathcal{L}_2 \right\}.$$

It is well known Aronszajn (1950) that the Sobolev space $\mathcal{W}_2^r([0,1])$ forms a reproducing kernel Hilbert space when endowed with the norm

$$\|f\|^2_{\mathcal{W}_2^r([0,1])} = \int_0^1 [f(t)]^2 + [f^{(r)}(t)]^2 \, dt.$$

Similar to the random design case, the combination of the theorems below establishes that the Penalized RKHS estimator achieves rate-optimality in the fixed design setting (see Theorems 3 and 4 below).

Theorem 3. *Fix $\tau \in (0,1)$. Suppose each component of the true parameter $\boldsymbol{\beta}_0$ resides in a compact subset $\mathcal{B}_k$ of a Sobolev space $\mathcal{H} = \mathcal{W}_2^r$. Then, under fixed design:*

$$\lim_{a \to 0} \lim_{m,n \to \infty} \inf_{\hat{\boldsymbol{\beta}}} \sup_{\boldsymbol{\beta} \in \mathcal{B}_k^p} P\left(\left\| \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right\|_2^2 > a \left(m^{-2r} + n^{-1} \right) \right) = 1,$$

where the infimum is taken over all possible estimators $\hat{\boldsymbol{\beta}}$ based on the observations.

Theorem 4. *Fix $\tau \in (0,1)$. Suppose each component of the true parameter $\boldsymbol{\beta}_0$ resides in a compact subset $\mathcal{B}_k$ of a Sobolev space $\mathcal{H} = \mathcal{W}_2^r$. If (A1) and (A2) hold and sample points $\mathbb{S} = \{\mathcal{T}_1, \ldots, \mathcal{T}_m\}$ satisfy $\mathcal{T}_{i+1} - \mathcal{T}_i \asymp \frac{1}{m}$, then under fixed design:*

$$\lim_{a \to +\infty} \lim_{m,n \to \infty} \sup_{\boldsymbol{\beta}_0 \in \mathcal{B}_k^p} P\left(\left\| \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right\|_2^2 > a \left(m^{-2r} + n^{-1} \right) \right) = 0,$$

when $\hat{\boldsymbol{\beta}}$ is the Penalized expectile RKHS estimator with $\lambda \lesssim m^{-2r} + n^{-1}$.

The optimal convergence rate under fixed design is interpretable. The m^{-2r} term can be attributed to the unavoidable error caused by discretization, and the n^{-1} term to the stochastic error.

3. Simulated Numerical Analysis

In Section 2.2, we established the optimal rate of convergence for our proposed penalized expectile RKHS estimator. This section delves into evaluating the numerical performance of these estimators through extensive simulation studies. We employ the following function-on-scalar regression model to generate data for these simulations:

$$y_i\left(\mathcal{T}_j\right)=\sum_{k=1}^{4} x_{ik}\boldsymbol{\beta}_k\left(\mathcal{T}_j\right)+v_i\varepsilon_i\left(\mathcal{T}_j\right), i=1,\cdots,n;\ j=1,\cdots,m.$$

For each i in the range $1,\cdots,n$, the variables x_{i2}, x_{i3}, and x_{i4} are independently and identically distributed (i.i.d) as standard normal, uniform distribution over $[0,1]$, and Bernoulli distribution with parameter 0.5, respectively. The variable x_{i1} is set to 1 for all i, serving as the intercept. In the simulation studies, we investigate how various factors influence the performance of the proposed estimator.

1. number of samples: $n\in\{50,100,200\}$;

2. number of nodes on each sample: $m\in\{20,40,80\}$;

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0206

3. expectile level: $\tau \in \{0.1, 0.5, 0.9\}$;

4. design (fixed vs random, even vs quadratic, distribution of s): Fixed even distribution: $\mathcal{T}_j = \frac{j-1}{m-1}$, $j = 1, \cdots, m$; fixed quadratic distribution: $\sqrt{\mathcal{T}_j} = \frac{j-1}{m-1}$, $j = 1, \cdots, m$; random even distribution: $\mathcal{T}_j \sim Uniform(0,1)$, $j = 1, \cdots, m$; random quadratic distribution: $\sqrt{\mathcal{T}_j} \sim Uniform(0,1)$, $j = 1, \cdots, m$.

5. noise distribution: Independent normal: $\varepsilon_i(\mathcal{T}_j)$ are independent $N(0,1)$; dependent normal: $\varepsilon_i(\mathcal{T}_j)$ are $N(0,1)$ and $Cov(\varepsilon_i(\mathcal{T}_j), \varepsilon_i(\mathcal{T}_j)) = \exp(-|\mathcal{T}_j - \mathcal{T}_j|)$; independent lognormal: $\log \varepsilon_i(\mathcal{T}_j)$ are independent $N(0,1)$; dependent lognormal: $\log \varepsilon_i(\mathcal{T}_j)$ are $N(0,1)$ and $Cov(\log \varepsilon_i(\mathcal{T}_j), \log \varepsilon_i(\mathcal{T}_j)) = \exp(-|\mathcal{T}_j - \mathcal{T}_j|)$;

6. heteroscedasticity in noise: No heteroscedasticity: $v_i = 1 = x_{i1}$; With heteroscedasticity: $v_i = \sqrt{2/5}(1 + x_{i4}) = \sqrt{2/5}(x_{i1} + x_{i4})$. The constant are choosen such that $\mathbb{E}(v_i^2(s)) = 1$;

7. Kernel: Hyperbolic kernel, defined as $K(x, x') = \frac{\cosh(\min(x,x'))\cosh(1-\max(x,x'))}{\sinh(1.0)}$, as described in Berlinet and Thomas-Agnan (2011) as the kernel of standard Sobolev space with $r = 1$. The second is the kernel of the Laplace radial basis function,defined as $K(x, x') = \exp\left(-\frac{|x-x'|}{2}\right)$.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0206

For the choice of shape of $\boldsymbol{\beta}$, we set $\boldsymbol{\beta}_1(s) = 2sin(2\pi s), \boldsymbol{\beta}_2(s) = -2sin(4\pi s), \boldsymbol{\beta}_3(s) = -3s(1-s), \boldsymbol{\beta}_4(s) = 4s^2 sin(2\pi s)$.

Let ϵ_τ be τ expectile of ϵ, without heteroscedasticity:

$$\boldsymbol{\beta}_1^\tau = \boldsymbol{\beta}_1 + \epsilon_\tau, \boldsymbol{\beta}_2^\tau = \boldsymbol{\beta}_2, \boldsymbol{\beta}_3^\tau = \boldsymbol{\beta}_3, \boldsymbol{\beta}_4^\tau = \boldsymbol{\beta}_4;$$

With heteroscedasticity:

$$\boldsymbol{\beta}_1^\tau = \boldsymbol{\beta}_1 + \sqrt{2/5}\epsilon_\tau, \boldsymbol{\beta}_2^\tau = \boldsymbol{\beta}_2, \boldsymbol{\beta}_3^\tau = \boldsymbol{\beta}_3, \boldsymbol{\beta}_4^\tau = \boldsymbol{\beta}_4 + \sqrt{2/5}\epsilon_\tau.$$

To evaluate the performance of the proposed estimator, the mean squared error (MSE) of the coefficients was used:

$$\sum_{k=1}^{4} \int_{x=0}^{1} (\hat{\boldsymbol{\beta}}_k^\tau(s) - \boldsymbol{\beta}_k^\tau)^2(s) ds.$$

The following estimator was used to evaluate the above MSE on sampled data points:

$$\sqrt{\frac{\sum_{j=1}^{n} \sum_{k=1}^{3} (\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)^2(s_n)}{\sum_{j=1}^{n} \sum_{k=1}^{3} (\boldsymbol{\beta}_k)^2(s_n)}}.$$

Given the extensive number of factor combinations in our study (amounting to $2^6 3^3 = 1728$), it is impractical to detail all the simulation results in this section. Instead, we highlight a few representative cases to demonstrate the impact of each factor. We initially focus on a moderately complex scenario featuring an even design, independent noise without heteroscedasticity, a medium sample size (n=100), and a moderate number of nodes

doi:10.5705/ss.202022.0206

(m=40). Table 1 presents the median MSE for various values of τ, noise distributions, and sampling designs, based on 50 repeats for each setting. The primary conclusions drawn from Table 1 are as follows: 1) the proposed method exhibits reasonable effectiveness across all examined scenarios; 2) in this specific context, the performance is better with random design; 3) the scenario involving log-normal error distribution and $\tau = 0.9$ underperformed relative to others, likely due to the lower density of the log-normal distribution at higher expectiles.

Table 1: Median MSE under different τ, noise distribution and sampling design with even design, independent noise with no heteroscedasticity and medium sample size ($n = 100$)

τ	Noise Distribution	Fixed Design (MSE std)	Random Design (MSE std)
0.1	normal	0.0118 ± 0.0039	0.0071 ± 0.0020
0.5	normal	0.0071 ± 0.0015	0.0034 ± 0.0016
0.9	normal	0.0118 ± 0.0033	0.0064 ± 0.0020
0.1	lognormal	0.0073 ± 0.0019	0.0035 ± 0.0018
0.5	lognormal	0.0089 ± 0.0037	0.0067 ± 0.0035
0.9	lognormal	0.0457 ± 0.0263	0.0381 ± 0.0298

To manage the complexity arising from over several hundred factor combinations in our simulation study, we utilized regression analysis to efficiently condense and analyze the data. This approach allowed us to systematically evaluate the impact of various factors on the MSE of the proposed estimator within a single, comprehensive model. Focusing on a harder scenario with a lognormal error distribution and $\tau = 0.1$, we used log-transformed MSE as our response variable, with binary predictors including design type, sample size, number of nodes, and the presence of correlation and heteroscedasticity in noise. The experiments are repeated 50 times for each setting. The model, with an explanatory power of 75% (R^2), revealed significant insights into factor impacts, detailed in Table 2.

In this regression model, all factors were significant except for heteroscedasticity and the choice of kernel. Notably, positive coefficients indicate poorer performance of the proposed estimator under the corresponding scenario, while negative coefficients suggest better performance. As expected, larger values of m (number of nodes) and n (sample size) improved estimator performance, with random even design also yielding better results. Additionally, the presence of correlations in noise distribution posed challenges to the estimation process.

Table 3 provides a comparative analysis of the median MSE for dif-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0206

Table 2: Regression coefficients to illustrate the effect of each factor on the MSE of the proposed estimators

	coef	**p-value**	[**0.025**	**0.975**]
Intercept	−3.9322	0.000	−4.110	−3.755
Random design	−0.7985	0.000	−0.911	−0.686
Quadratic design	0.7893	0.000	0.677	0.901
$n = 100$	−0.3632	0.000	−0.501	−0.226
$n = 200$	−0.6583	0.000	−0.796	−0.521
$m = 40$	−0.6681	0.000	−0.806	−0.531
$m = 80$	−0.8150	0.000	−0.952	−0.678
Correlation	0.7909	0.000	0.679	0.903
Heteroscedasticity	−0.0831	0.146	−0.195	0.029
Hyperbolic Kernel	0.0884	0.122	−0.024	0.201

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0206

ferent m and n values, conducted under an even design using hyperbolic kernel without heteroscedasticity or noise correlations. In fixed design settings, the number of nodes (m) was more influential than the sample size (n). In contrast, in random design scenarios, where all sampling points are pooled for estimation, the effects of m and n were more balanced. Table 4 offers similar comparisons but includes noise correlations. The results align with the noncorrelated scenarios, yet the performance generally deteriorates, particularly for $n = 50$, due to the effective reduction in sample size caused by correlations.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0206

Table 3: Median MSE under even design without correlations.

n	m	Fixed Design (MSE $\pm$ std)	Random Design (MSE $\pm$ std)
50	20	0.0319 ± 0.0073	0.0061 ± 0.0028
50	40	0.0094 ± 0.0026	0.0048 ± 0.0022
50	80	0.0048 ± 0.0014	0.0043 ± 0.0015
100	20	0.0304 ± 0.0051	0.0037 ± 0.0016
100	40	0.0073 ± 0.0019	0.0035 ± 0.0018
100	80	0.0031 ± 0.0008	0.0027 ± 0.0011
200	20	0.0281 ± 0.0035	0.0034 ± 0.0017
200	40	0.0030 ± 0.0009	0.0026 ± 0.0008
200	80	0.0029 ± 0.0006	0.0022 ± 0.0010

Table 4: Median MSE under even design with correlations.

n	m	Fixed Design (MSE $\pm$ std)	Random Design (MSE $\pm$ std)
50	20	0.0556 ± 0.0266	0.0276 ± 0.0292
50	40	0.0353 ± 0.0366	0.0175 ± 0.0282
50	80	0.0234 ± 0.0247	0.0171 ± 0.0292
100	20	0.0393 ± 0.0094	0.0122 ± 0.0091
100	40	0.0154 ± 0.0082	0.0120 ± 0.0091
100	80	0.0080 ± 0.0087	0.0117 ± 0.0102
200	20	0.0335 ± 0.0077	0.0083 ± 0.0061
200	40	0.0063 ± 0.0035	0.0062 ± 0.0046
200	80	0.0062 ± 0.0040	0.0055 ± 0.0037

4. Data Analysis

In this section, we apply the proposed method to the QoL data from a breast cancer clinical trial. MA.5 (Levine et al. (1998)) was a randomized trial comparing a combination chemotherapy regimen CEF (cyclophosphamide, epirubicin, and fluorouracil) with the standard of care, CMF (cyclophosphamide, methotrexate, and fluorouracil), in premenopausal women with node-positive breast cancer, conducted by the Canadian Cancer Trials Group (CCTG). The trial demonstrated the superiority of CEF over CMF in both disease-free and overall survival, and was one of the early clinical trials to incorporate quality of life data collection into its design. The Breast Cancer Questionnaire (Levine et al. (1988)) was administered at baseline, then monthly for the first six months, and subsequently every three months for up to two years. In the primary publication of MA.5, the mean BCQ summary score trajectories were presented by treatment arm, revealing a clear interaction between time and treatment. However, no formal statistical analysis was conducted. The MA.5 QoL data have been analyzed using quantile regression and a generalized partially linear model (Zheng et al. (2017), Lv et al. (2019)), with a focus on missing data. In this analysis, we explore the use of expectile function-on-scalar regression in this setting.

During the clinical trial, 716 premenopausal women with early-stage breast cancer participated, with 356 randomized to receive CEF treatment and 360 to CMF treatment. The QoL of patients was assessed using the self-administered BCQ, comprising 30 questions scored from one to seven, where the lowest score indicates the worst possible outcome and the highest, the best possible outcome. The BCQ measures various dimensions of QoL and was administered to each patient on the first day of each treatment cycle, and then every three months after completing six treatment cycles, until the end of the second year. The maximum number of visits was 14, with a minimum of one for those who participated in this study. Figure 1 presents a scatterplot of the average BCQ score change from baseline to year 2 for each individual. The solid blue line represents the mean change for the CEF group, and the dashed red line, the CMF group. The left panel includes all trial participants, while the right panel only those followed until year two post-randomization. The increasing trend in BCQ scores is partly due to data missing from disease progression or death of patients on trial.

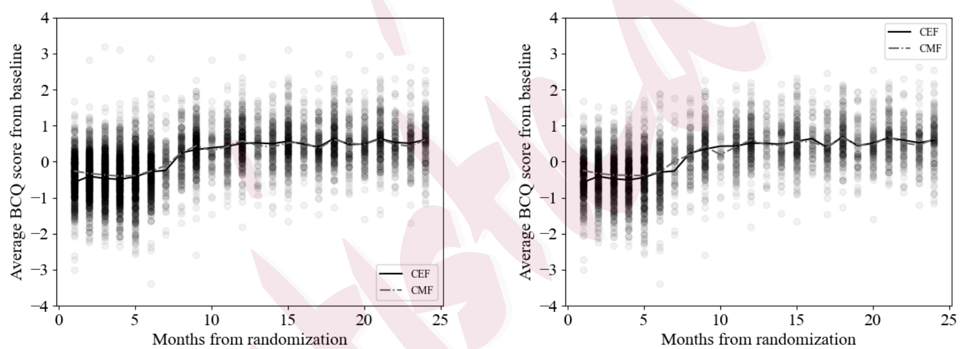


Figure 1: Average BCQ score change from baseline to Year 2 for all subjects (left). Average BCQ score change from baseline to Year 2 for subjects will complete followup till Year 2 (right). The blue solid line represents the mean change for the CEF group and the red dash line represents the CMF group.

For a specific subscale of QoL, such as the emotional subscale of BCQ, before the analysis, the score of the subscale is first calculated by the mean of the answers from patients to the questions defining the subscale. In this analysis, we use the mean score of all 30 questions as the response. The scores of a QoL subscale are restricted by the minimum and maximum scores of the questions defining the subscale.

For the BCQ emotional subscale score, 713 individuals (99.6%) have score at their first visit, but only 711 (99.3%), 708 (98.9%), 703 (98.2%), 699 (97.6%), 690 (96.4%), 668 (93.3%), 634 (88.5%), 607 (84.8%), 558 (77.9%), 500 (69.8%), 392 (54.7%), 214 (29.9%), and 5 (0.7%) individuals have BCQ emotional subscale score at the following measurements, respectively.

To address the issue of monotonic missingness in our dataset, we employed two distinct analytical approaches: 1) an analysis utilizing the unimputed raw data; 2) an analysis restricted to patients with complete two-year follow-up data. It is important to note that our proposed method does not necessitate identical sampling points across data points. However, as we will demonstrate, disregarding the missing data mechanism in this context may significantly bias the estimators at later months, predominantly influenced by patients who remain alive.

We analyzed each dataset using two models: 1) The first model con-

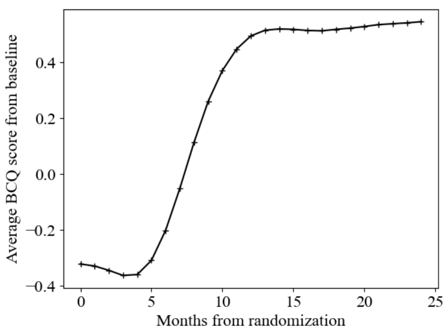
Statistica Sinica: Preprint
doi:10.5705/ss.202022.0206

siders treatment assignment as the sole covariate. 2) The second model incorporates additional prognostic biomarkers, including age, surgical type (1 for total mastectomy, 0 for partial mastectomy), and the presence of positive auxiliary nodes (1 for more than 10, 0 for 10 or fewer). Figures 2 and 3 depict the estimated intercepts at $\tau = 0.5$, and the estimated coefficient trajectories at $\tau = 0.1, 0.5, 0.9$ for both the treatment-only model and the model with four predictors, respectively, using raw data. Figures 4 and 5 provide analogous information but are based solely on patients with a complete two-year follow-up.

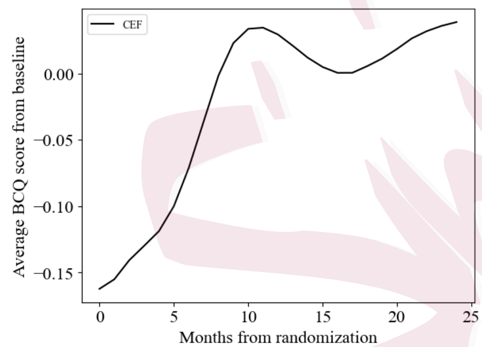
Analysis of the raw data reveals that the average BCQ score treatment effect of CEF (versus CMF) at the mean ($\tau = 0.5$) level is initially negative, which could be attributed to higher acute toxicity in the CEF group. However, this effect stabilizes and turns positive over time. At $\tau = 0.1$, the estimated coefficient trajectories are similar; at $\tau = 0.9$, there is a predominantly increasing pattern, with the magnitude of the positive treatment effect being more pronounced in the long term. In the model incorporating four predictors, the trend for CEF mirrors that observed in the treatment-only model, consistent with expectations from a randomized study. Notably, the impact of age appears to decrease over time. Figures 3 and 4, which include only patients with a full two-year follow-up, exhibit

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0206

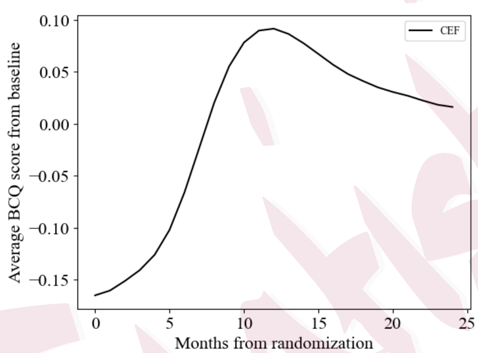
patterns similar to Figures 1 and 2. This similarity likely arises because these patients contribute more data, thus heavily influencing the estimation process.



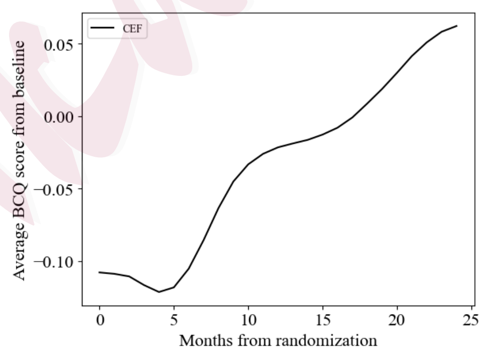
(a) Intercept with $\tau = 0.5$



(b) Coefficient with $\tau = 0.5$

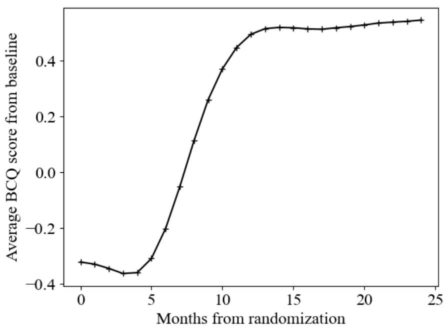


(c) Coefficient with $\tau = 0.1$

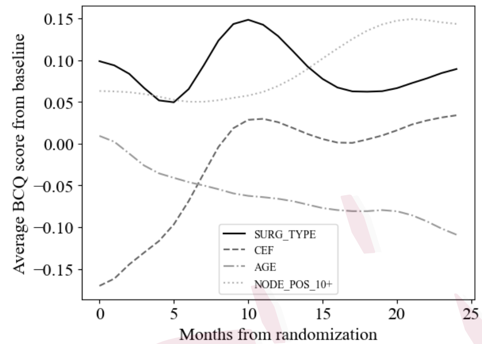


(d) Coefficient with $\tau = 0.9$

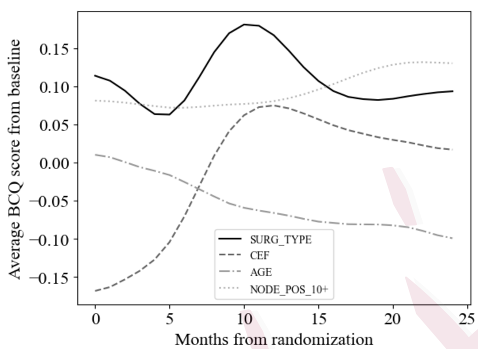
Figure 2: Estimated regression coefficient at different τs for the treatment-only model using raw data



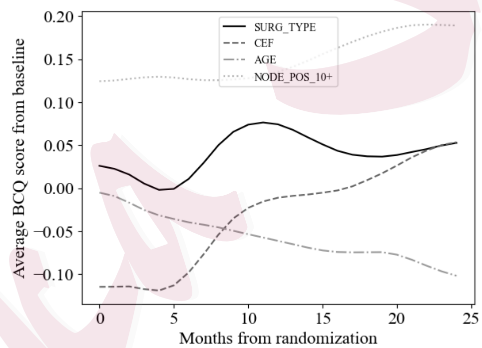
(a) Intercept with $\tau = 0.5$



(b) Coefficient with $\tau = 0.5$

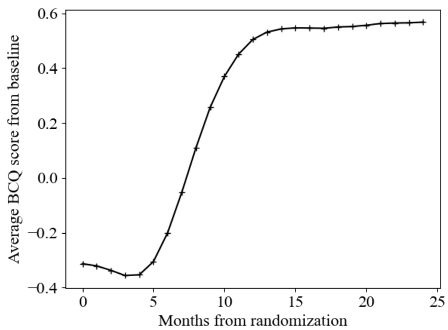


(c) Coefficient with $\tau = 0.1$

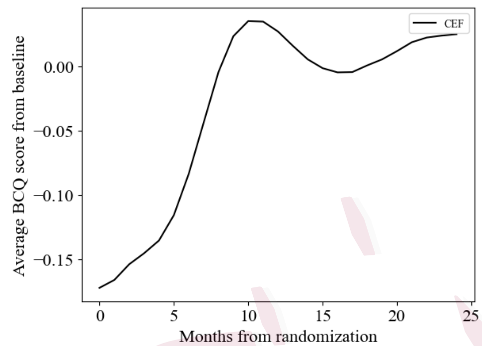


(d) Coefficient with $\tau = 0.9$

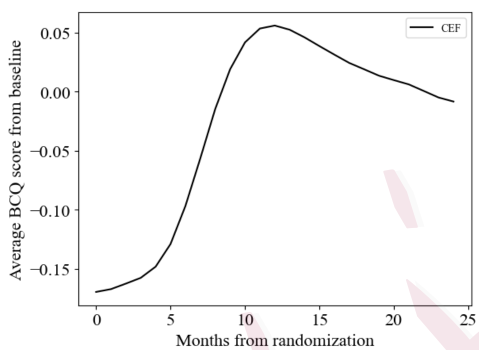
Figure 3: Estimated regression coefficient at different τs with 4 predictors using raw data



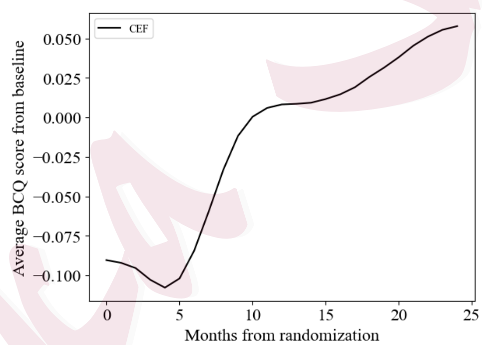
(a) Intercept with $\tau = 0.5$



(b) Coefficient with $\tau = 0.5$



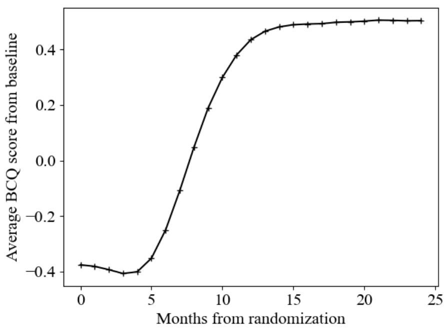
(c) Coefficient with $\tau = 0.1$



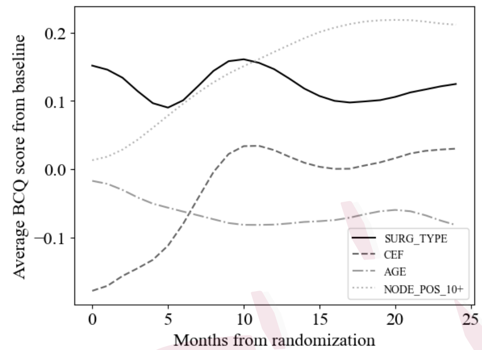
(d) Coefficient with $\tau = 0.9$

Figure 4: Estimated regression coefficient at different τs for the treatment-only model using alive patients only

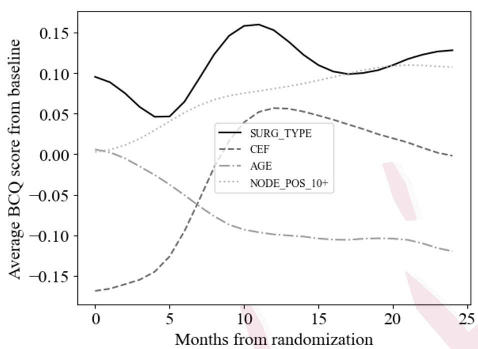
Statistica Sinica: Preprint
doi:10.5705/ss.202022.0206



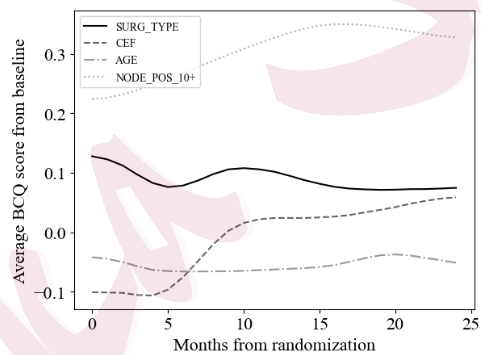
(a) Intercept with $\tau = 0.5$



(b) Coefficient with $\tau = 0.5$



(c) Coefficient with $\tau = 0.1$



(d) Coefficient with $\tau = 0.9$

Figure 5: Estimated regression coefficient at different τs with 4 predictors using alive patients only

5. Conclusion

In this study, a function-on-scalar expectile regression was investigated, focusing on a varying-coefficient model with a functional response. We introduced a penalized expectile RKHS estimator and established its optimal rate of convergence under both random and fixed design settings, assuming relatively mild conditions. The estimator's efficiency and robustness were thoroughly assessed through extensive simulation studies, encompassing diverse design scenarios and data generation processes. Overall, our findings indicate that larger sample sizes, higher sampling frequencies, and an even design typically enhance estimator performance, while the presence of correlated noise negatively impact the estimation.

The development of our methodology was motivated by the need to analyze longitudinal health-related quality of life data from a breast cancer clinical trial. This trial, characterized by a fixed and uneven design with more frequent observations in early stages (a common approach in clinical trials), served as a practical test case. Applying our method, we found results consistent with the trial's main analysis, suggesting that while the experimental drug may slightly reduce quality of life during treatment due to toxicities, it could offer comparable long-term quality of life outcomes compared to the control group. Unlike standard methods that compare mean BCQ scores at a specific time point, our estimator allowed for an assessment of the treatment's impact throughout the sampling period at various τ levels. This approach proves particularly valuable in accounting for the high variability typically seen in the effects of cancer treatments.

Supplementary Material

This supplemental material contains the technical proofs for Theorems 1–4.

Acknowledgments

Yi Liu is supported by the York University startup fund. Bei Jiang and Linglong Kong were partially supported by grants from the Canada CIFAR AI Chairs program, the Alberta Machine Intelligence Institute (AMII), and Natural Sciences and Engineering Council of Canada (NSERC), and Linglong Kong was also partially supported by grants from the Canada Research Chair program from NSERC.

References

Aigner, D. J., T. Amemiya, and D. J. Poirier (1976). On the estimation of production frontiers: maximum likelihood estimation of the parameters of a discontinuous density function. International Economic Review, 377–396.

Aronszajn, N. (1950). Theory of reproducing kernels. Transactions of the American mathematical society 68(3), 337–404.

Berlinet, A. and C. Thomas-Agnan (2011). Reproducing kernel Hilbert spaces in probability and statistics. Springer Science & Business Media.

Breiman, L. and J. H. Friedman (1985). Estimating optimal transformations for multiple regression and correlation. Journal of the American statistical Association 80(391), 580–598.

Cai, J. and C. Weng (2016). Optimal reinsurance with expectile. Scandinavian Actuarial Journal 2016(7), 624–645.

Cai, T. T., P. Hall, et al. (2006). Prediction in functional linear regression. The Annals of Statistics 34(5), 2159–2179.

Cardot, H., C. Crambes, and P. Sarda (2005). Quantile regression when the covariates are functions. Nonparametric Statistics 17(7), 841–856.

Chen, K. and H.-G. Müller (2012). Conditional quantile analysis when covariates are functions, with application to growth data. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 74(1), 67–89.

Chen, Y., J. Goldsmith, and R. T. Ogden (2016). Variable selection in function-on-scalar regression. Stat 5(1), 88–101.

Cleveland, W. S. and E. Grosse (1991). Computational methods for local regression. Statistics and computing 1(1), 47–62.

Ferraty, F. and P. Vieu (2006). Nonparametric functional data analysis: theory and practice. Springer Science & Business Media.

Goldsmith, J. and T. Kitago (2016). Assessing systematic effects of stroke on motorcontrol by using hierarchical function-on-scalar regression. Journal of the Royal Statistical Society. Series C, Applied statistics 65(2), 215.

Goldsmith, J., V. Zipunnikov, and J. Schrack (2015). Generalized multilevel function-on-scalar regression and principal component analysis. Biometrics 71(2), 344–353.

Hall, P., J. L. Horowitz, et al. (2007). Methodology and convergence rates for functional linear regression. Annals of Statistics 35(1), 70–91.

Hastie, T., C. Loader, et al. (1993). Local regression: Automatic kernel carpentry. Statistical Science 8(2), 120–129.

Hastie, T. and R. Tibshirani (1993). Varying-coefficient models. Journal of the Royal Statistical Society: Series B (Methodological) 55(4), 757–779.

Kato, K. et al. (2012). Estimation in functional linear quantile regression. The Annals of Statistics 40(6), 3108–3136.

Koenker, R. and G. Bassett Jr (1978). Regression quantiles. Econometrica: journal of the Econometric Society, 33–50.

Koenker, R. and K. F. Hallock (2001). Quantile regression. Journal of economic perspectives 15(4), 143–156.

Levine, M. N., V. H. Bramwell, K. I. Pritchard, B. D. Norris, L. E. Shepherd, H. Abu-Zahra, B. Findlay, D. Warr, D. Bowman, J. Myles, et al. (1998). Randomized trial of intensive cyclophosphamide, epirubicin, and fluorouracil chemotherapy compared with cyclophosphamide, methotrexate, and fluorouracil in premenopausal women with node-positive breast cancer. national cancer institute of canada clinical trials group. Journal of Clinical Oncology 16(8), 2651–2658.

Levine, M. N., G. H. Guyatt, M. Gent, S. De Pauw, M. D. Goodyear, W. M. Hryniuk, A. Arnold, B. Findlay, J. R. Skillings, and V. H. Bramwell (1988). Quality of life in stage ii breast cancer: an instrument for clinical trials. Journal of Clinical Oncology 6(12), 1798–1810.

Li, Y., Y. Liu, and J. Zhu (2007). Quantile regression in reproducing kernel hilbert spaces. Journal of the American Statistical Association 102(477), 255–268.

Lv, Y., G. Qin, Z. Zhu, and D. Tu (2019). Quantile regression and empirical likelihood for the analysis of longitudinal data with monotone missing responses due to dropout, with applications to quality of life measurements from clinical trials. Statistics in Medicine 38(16), 2972–2991.

Nesterov, Y. (1983). A method for unconstrained convex minimization problem with the rate of convergence o (1/k2). In Dokl. Akad. Nauk. SSSR, Volume 269, pp. 543.

Newey, W. K. and J. L. Powell (1987). Asymmetric least squares estimation and testing. Econometrica: Journal of the Econometric Society, 819–847.

Ramsay, J. and B. Silverman (2005). Principal components analysis for functional data. Functional data analysis.

Ramsay, J. O. and B. W. Silverman (2007). Applied functional data analysis: methods and case studies. Springer.

Reiss, P. T., L. Huang, and M. Mennes (2010). Fast function-on-scalar regression with penalized basis expansions. The international journal of biostatistics 6(1).

Riesz, F. and S. Nagy (1955). B.(1990). functional analysis. Dover Publications, Inc., New York. First published in 3(6), 35.

Stigler, S. M. (1984). Studies in the history of probability and statistics xl boscovich, simpson and a 1760 manuscript note on fitting a linear relation. Biometrika 71(3), 615–620.

Wahba, G. (1990). Spline models for observational data. SIAM.

Wang, X. (2016). Optimal estimation for quantile regression with functional response.

Zhang, J. (2014). Analysis of variance for functional data. Monographs on statistics and applied probability 127, 127.

Zhang, Z., X. Wang, L. Kong, and H. Zhu (2021). High-dimensional spatial quantile function-on-scalar regression. Journal of the American Statistical Association, 1–37.

Zheng, X., G. Qin, and D. Tu (2017). A generalized partially linear mean-covariance regression model for longitudinal proportional data, with applications to the analysis of quality of life data from cancer clinical trials. Statistics in Medicine 36(12), 1884–1894.

Zhu, H., R. Li, and L. Kong (2012). Multivariate varying coefficient model for functional responses. The Annals of Statistics 40(5), 2634–2666.

doi:10.5705/ss.202022.0206

Department of Mathematical and Statistical Sciences, York University, 4700 Keele Street, Toronto, ON, Canada.

E-mail: yliu42@yorku.ca

Department of Public Health Sciences and Canadian Cancer Trials Group, Queen's University, Kingston, ON, Canada.

E-mail: wei.tu@queensu.ca

Department of Mathematical Sciences, University of Essex, Colchester, UK.

E-mail: ybaoa@essex.ac.uk

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB Canada.

E-mail: bei1@ualberta.ca

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB Canada.

E-mail: lkong@ualberta.ca