Statistica Sinica Preprint No: SS-2026-0148	
Title	Function-on-Function Geometric Quantile Regression with a Distributed Extension
Manuscript ID	SS-2026-0148
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202026.0148
Complete List of Authors	Bingchen Li, Alan T.K. Wan, Miaomiao Yu and Yong Zhou
Corresponding Authors	Miaomiao Yu
E-mails	mmyu@fem.ecnu.edu.cn
Notice: Accepted author version.	

Statistica Sinica: Newly accepted Paper

Statistica Sinica

Function-on-Function Geometric Quantile Regression with a Distributed Extension

Bingchen Li[1], Alan T. K. Wan[2], Miaomiao Yu[1], Yong Zhou[1]

[1]*School of Statistics and Academy of Statistics and Interdisciplinary Sciences*

East China Normal University

[2]*Department of Decision Analytics and Operations, City University of Hong Kong*

Abstract: This paper proposes a novel Function-on-Function (FoF) quantile regression model for characterizing the influence of functional covariates across different parts of the conditional distribution of functional responses. We design a loss function based on geometric quantiles in infinite-dimensional spaces. This ensures that our method provides well-defined conditional quantiles for functional data, while retaining desirable statistical properties similar to those in the univariate setting. Additionally, the convergence rates of estimators are explicitly derived for both given and estimated basis functions. To address the challenges posed by large-scale functional data that are often distributed across multiple computing nodes, we further extend our framework to a distributed setting. We propose a communication-efficient smooth FoF geometric quantile regression algorithm that not only achieves computational and communication efficiency but also guarantees convergence rates comparable to those of global estimators.

Statistica Sinica: Newly accepted Paper

Key words and phrases: Communication efficiency, Distributed computation, Functional data, Function-on-function regression, Geometric quantile.

1. Introduction

Functional data, representing observations as functions, have become increasingly prevalent across various disciplines due to advancements in modern data technology. This trend is evident in fields such as biostatistics (Li et al., 2022), industrial statistics (Horváth et al., 2021), and econometrics (Li et al., 2024), among others. Consequently, functional data analysis (FDA) has become an important research focus. Within FDA, regression analysis is foundational, and much of the existing work has focused on function-on-scalar (FoS) and scalar-on-function (SoF) regression.

However, it is common in practice for both the response and the predictor to be functional data, making FoF regression a vital area of investigation; see, for instance, Benatia et al. (2017). Despite the significant practical value of FoF regression, existing research remains relatively limited. The main challenge stems from the inherently infinite-dimensional nature of both the response and predictor, leading to two major issues: first, it renders the model's computational complexity substantially higher than in finite-dimensional cases; second, the per-sample storage requirement

Statistica Sinica: Newly accepted Paper

far exceeds that of conventional data. Consequently, large-scale functional datasets, which are especially prevalent in the era of big data, often exceed the memory and computational capacity of a single machine and are therefore distributed across multiple nodes. Accordingly, the primary goal of this paper is to develop a novel FoF regression methodology and extend it to a distributed algorithm specifically designed for the demands of large-scale functional data analysis.

Among FoF regression methods, most existing work has focused on modeling conditional means. However, these approaches have notable limitations in practical applications. Firstly, mean regression methods are particularly sensitive to outliers and typically rely on the assumption of homoscedasticity. Consequently, when the data exhibit heteroscedasticity or contain prominent outliers, FoF mean regression may lead to substantial prediction errors (Beyaztas et al., 2025). Secondly, modeling only the conditional mean fails to completely illustrate the conditional distribution of the response. In many applications, the primary interest lies in understanding how predictors influence extreme regions (e.g., tails) of the response distribution rather than merely its center (Liu et al., 2020).

In recent years, functional quantile regression has been studied in both the SoF and FoS settings (Kato, 2012; Ma et al., 2019; Hu et al., 2024; Liu

et al., 2020; Zhang et al., 2022), and Dong et al. (2024) further considered a model that treats the quantile function as functional data. Nevertheless, none of the aforementioned models are applicable to the FoF regression setting, nor can they effectively address the practical challenge of large-scale functional data distributed across multiple computational nodes. Therefore, this paper aims to propose a novel FoF quantile regression method and to further extend this method to a distributed computing framework.

Several studies have examined FoF quantile regression. Beyaztas et al. (2022) proposed a partial quantile regression framework, which was further developed by Mutis et al. (2025) through two additional algorithms. In the linear setting, Beyaztas et al. (2025) developed a penalized linear FoF quantile regression method that discretizes the functional data and defines a loss function based on the check loss used in univariate quantile regression. To accommodate more complex nonlinear relationships and interactions, Zhu et al. (2023) introduced a semiparametric FoF quantile regression approach.

Although the aforementioned methods differ in model specification and estimation strategies, they share one common feature — reliance on a pointwise modeling strategy, i.e., treating the functional response $Y(t)$ as a scalar at each time point t and subsequently applying univariate quantile regression techniques. This strategy, however, neglects the inherent geometric

Statistica Sinica: Newly accepted Paper

structure of functional data by failing to consider the response as a random element in a Hilbert space. The estimators thus obtained merely characterize the trajectory of the one-dimensional conditional quantiles of $Y(t)$ over time, rather than capturing the overall conditional quantile structure of an infinite-dimensional random element. More critically, several fundamental concepts in univariate quantile inference, such as distribution functions and quantile intervals, lack natural extensions within this framework.

Geometric quantiles in infinite-dimensional Hilbert spaces (Chakraborty and Chaudhuri, 2014) provide a systematic way to address these challenges. Unlike pointwise methods, this approach captures the overall relative position of functional data within the infinite-dimensional space. In addition, geometric quantiles can define quantile regions, revealing the shape and structure of the data cloud in the infinite-dimensional space from a distributional perspective (Konen, 2025). Hence, this approach effectively reflects the holistic distributional structure of functional data.

Leveraging this framework, Chowdhury and Chaudhuri (2019) developed a nonparametric model for FoF quantile regression. However, the model relies on Nadaraya-Watson (NW) kernel estimation, making it susceptible to the curse of dimensionality, especially when both the predictor and response variables are infinite-dimensional functional data. Moreover,

Statistica Sinica: Newly accepted Paper

the method is computationally inefficient, as it requires iterative calculations for each predictor.

1.1 The contributions

In this paper, a novel FoF quantile regression framework is proposed. Furthermore, we employ a "divide-and-conquer" strategy by extending our proposed FoF quantile regression method to a distributed framework. The main contributions of this paper are summarized as follows.

Our first contribution is the development of a novel function-on-function quantile regression framework. Specifically, we propose a function-on-function geometric quantile regression framework that treats the functional response as a random element in a Hilbert space, rather than relying on a pointwise strategy. We develop an efficient sieve estimation procedure for the regression operator by explicitly computing the Fréchet derivatives. We establish the consistency and convergence rates of our estimator. For known basis functions, the rate decomposes into a statistical error and an approximation error from operator projection. Furthermore, when the basis functions are estimated, an additional term emerges in the convergence rate, and its magnitude decreases with the smoothness of the true basis functions.

Our second contribution is the development of a distributed FoF geo-

Statistica Sinica: Newly accepted Paper

metric quantile regression framework. In constructing this framework, we encounter two main challenges. The first concerns the estimation of basis functions when they are unknown and the data are stored across multiple computational nodes. In such distributed settings, estimating basis functions via traditional Functional Principal Component Analysis (FPCA) is difficult because of the substantial communication cost. To address this issue, we develop a distributed FPCA algorithm that achieves accurate basis estimation with much lower communication overhead.

The second challenge is to extend the proposed function-on-function geometric quantile regression framework to a distributed setting while preserving communication efficiency and achieving a convergence rate comparable to that of the global estimator. A widely used and effective approach constructs a surrogate empirical loss function, as in Jordan et al. (2019). However, the second-order Fréchet derivative of the loss function for the functional geometric quantile regression does not satisfy the regularity conditions in Jordan et al. (2019). Consequently, directly applying that method does not attain the same statistical efficiency as the global estimator. To address this limitation, we propose a smoothed surrogate loss that stabilizes the higher-order Fréchet derivatives and develop a communication-efficient distributed algorithm based on it. Under mild assumptions, we rigorously

Statistica Sinica: Newly accepted Paper

show that the proposed distributed algorithm attains a convergence rate comparable to that of the global estimator.

1.2 Organization of the article

The remainder of this paper is organized as follows. Section 2 introduces the FoF quantile regression model and presents the corresponding estimation framework. Section 3 provides a detailed analysis of the theoretical properties of the FoF quantile regression estimator. In Section 4, we extend the FoF quantile regression model to the distributed setting, introducing the communication-efficient smoothed distributed FoF geometric quantile regression algorithm and the distributed FPCA algorithm, along with their theoretical guarantees. Sections 5 and 6 present numerical simulations and a real data application, respectively, followed by a conclusion in Section 7.

2. Functional geometric quantile regression

In this section, we first introduce the notation and the concepts of geometric quantiles in Hilbert spaces, and then formulate the FoF geometric quantile regression model together with its estimation procedure.

Statistica Sinica: Newly accepted Paper

2.1 Notation

We begin by introducing the notation used throughout this paper. For a separable Hilbert space $\mathcal{H}$, its inner product is denoted by $\langle \cdot, \cdot \rangle$, and the induced norm is given by $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$. Let X and Y be random elements taking values in separable Hilbert spaces $\mathcal{H}_X$ and $\mathcal{H}_Y$, respectively, and let $\{(X_i, Y_i)\}_{i=1}^n$ be an independent and identically distributed (i.i.d.) dataset. The operator norm of $\mathcal{O}$ is defined as $\|\mathcal{O}\|_{\mathcal{X}} = \sup_{\substack{x \in \mathcal{H}_X \\ \|x\| \leq 1}} \|\mathcal{O}(x)\|$, where $\mathcal{O}(x)$ represents the image of x under $\mathcal{O}$. In particular, we specifically denote $\| \cdot \|_2$ as the Euclidean norm in a finite-dimensional Euclidean space. For any measurable functional $\mathfrak{f}$, we define $P\mathfrak{f} = \mathbb{E}[\mathfrak{f}(X)]$ and $\mathbb{P}_n\mathfrak{f} = \frac{1}{n}\sum_{i=1}^n \mathfrak{f}(X_i)$. We also use $\mathbb{P}$ to denote the probability measure throughout this paper to avoid ambiguity. $N(\epsilon, \mathcal{F}, \| \cdot \|)$ denotes the ϵ-covering number of the function class $\mathcal{F}$ with respect to the norm $\| \cdot \|$, for $\epsilon > 0$.

2.2 Geometric quantile

This subsection introduces geometric quantiles and the corresponding conditional geometric quantiles. Geometric quantiles are defined based on the "depth" of a data cloud; see Subsection S1.1 of the Supplementary Material Li et al. (2026) for a more detailed illustration.

Following Chakraborty and Chaudhuri (2014), the spatial distribution

Statistica Sinica: Newly accepted Paper

at a point $y \in \mathcal{H}_Y$ is defined as

$$S_Y(y) = \mathbb{E}\left(\frac{y-Y}{\|y-Y\|}\mathbb{I}(\|y-Y\| \neq 0)\right), \tag{2.1}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. This represents the average direction vector of y within the data cloud. Correspondingly, the spatial depth at the point y is defined as

$$SD_Y(y) = 1 - \|S_Y(y)\|, \tag{2.2}$$

ensuring that a larger value indicates greater depth.

Building upon the spatial distribution, let $r \in [0,1)$ and $u \in \mathcal{H}_Y$ with $\|u\| = 1$. A geometric quantile of order r in direction u for Y is defined as

$$Q = \arg\min_{\tilde{Q}\in\mathcal{H}_Y} \mathbb{E}\left(\|\tilde{Q} - Y\| + r\langle u, Y - \tilde{Q}\rangle\right). \tag{2.3}$$

For notational convenience, we let $\tau = ru$ and define the geometric check function

$$\rho_\tau(z) = \|z\| + \langle \tau, z\rangle, \qquad z \in \mathcal{H}_Y.$$

Accordingly, for a fixed $x \in \mathcal{H}_X$, the conditional τ-geometric quantile of Y given $X = x$ is defined by

$$Q_{Y|X=x}(\tau) = \arg\min_{\tilde{Q}\in\mathcal{H}_Y} \mathbb{E}\Big[\rho_\tau(Y - \tilde{Q}) \mid X = x\Big]. \tag{2.4}$$

In practice, we set $\tau = ru$ to specify the target conditional geometric quantile $Q_{Y|X=x}(\tau)$, where r is the quantile order and $u \in \mathcal{H}_Y$ is a direc-

Statistica Sinica: Newly accepted Paper

tion of interest. Larger values of r correspond to more extreme conditional geometric quantiles in that direction, while $r = 0$ gives the conditional geometric median regardless of u. Further discussion is provided in Supplementary Subsection S1.2.

2.3 Function-on-Function Geometric Quantile Regression

Building on the definition of geometric quantiles introduced above, we formulate the FoF geometric quantile regression model and characterize the associated conditional geometric quantile operator.

For a fixed $\tau \in \mathcal{H}_Y$ satisfying $\|\tau\| < 1$, let $\mathcal{O}_\tau^* : \mathcal{H}_X \to \mathcal{H}_Y$ be the conditional τ-geometric quantile operator defined through $Q_{Y|X=x}(\tau) = \mathcal{O}_\tau^*(x)$. For ease of notation, we omit the subscript τ throughout this paper and write $\mathcal{O}^*$ whenever no ambiguity arises.

To model $\mathcal{O}^*$, we assume it belongs to a structured operator class $\mathcal{G}$. Each element of $\mathcal{G}$ is denoted by $\mathcal{O}_{\boldsymbol{\theta}}$, which is induced by a coefficient sequence $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top, \ldots)^\top \in \Theta$, where $\boldsymbol{\theta}_i = (\theta_{i1}, \theta_{i2}, \ldots)^\top$, and Θ denotes the admissible parameter space. Specifically, for each $\boldsymbol{\theta} \in \Theta$, the corresponding operator $\mathcal{O}_{\boldsymbol{\theta}} \in \mathcal{G}$ is defined by

$$\mathcal{O}_{\boldsymbol{\theta}}(x) = \sum_{i=1}^{\infty} f_i \left(\sum_{j=1}^{\infty} \theta_{ij} \langle x, \phi_j^{(X)} \rangle \right) \phi_i^{(Y)}, \qquad x \in \mathcal{H}_X. \tag{2.5}$$

Here $\{\phi_i^{(Y)}\}_{i=1}^{\infty}$ and $\{\phi_j^{(X)}\}_{j=1}^{\infty}$ are bases of $\mathcal{H}_Y$ and $\mathcal{H}_X$, respectively, and

Statistica Sinica: Newly accepted Paper

$\{f_i\}_{i=1}^{\infty}$ is a collection of pre-specified smooth real-valued link functions. This class contains the classical linear FoF model as a special case with $f_i(a) = a + b_i$, where b_i is a constant. For notational convenience, let $f_i(x;\boldsymbol{\theta}_i) := f_i\left(\sum_{j=1}^{\infty}\theta_{ij}\langle x, \phi_j^{(X)}\rangle\right)$.

Under (2.4) and the modeling assumption that $\mathcal{O}^* \in \mathcal{G}$, the conditional τ-geometric quantile operator is characterized as the minimizer of the population risk over $\mathcal{G}$:

$$\mathcal{O}^* = \arg\min_{\tilde{\mathcal{O}}\in\mathcal{G}} \mathbb{E}\Big[\rho_\tau\Big(Y - \tilde{\mathcal{O}}(X)\Big)\Big]. \tag{2.6}$$

Since each operator in $\mathcal{G}$ is represented by a parameter sequence in Θ, there exists $\boldsymbol{\theta}^* \in \Theta$ such that $\mathcal{O}^* = \mathcal{O}_{\boldsymbol{\theta}^*}$. Accordingly, (2.6) induces the following equivalent characterization of $\boldsymbol{\theta}^*$:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}\in\Theta} \mathbb{E}\left[\rho_\tau\left(Y - \sum_{i=1}^{\infty} f_i(X;\boldsymbol{\theta}_i)\phi_i^{(Y)}\right)\right]. \tag{2.7}$$

2.4 Estimation

In this subsection, we describe how to estimate the conditional τ-geometric quantile operator $\mathcal{O}^*$, or equivalently the associated parameter sequence $\boldsymbol{\theta}^*$, from the observed data $\{(X_i, Y_i)\}_{i=1}^n$ via the empirical loss function $\mathbb{P}_n(\rho_\tau)$.

Since the parameter sequence remains infinite-dimensional, we approximate the structured class by a sieve space. Specifically, we set $f_i(\cdot) \equiv 0$

Statistica Sinica: Newly accepted Paper

for $i > L_Y(n)$ and $\theta_{ij} \equiv 0$ whenever $i > L_Y(n)$ or $j > L_X(n)$, where $L_X(n)$ and $L_Y(n)$ are positive integers satisfying $L_X(n) \to \infty$ and $L_Y(n) \to \infty$ as $n \to \infty$. Under this sieve approximation, elements of the structured operator class are represented as

$$\mathcal{O}_{{}^{(n)}\boldsymbol{\theta}}(X) = \sum_{i=1}^{L_Y(n)} f_i(X; {}^{(n)}\boldsymbol{\theta}_i)\phi_i^{(Y)},$$

where ${}^{(n)}\boldsymbol{\theta} = ({}^{(n)}\boldsymbol{\theta}_1^T, {}^{(n)}\boldsymbol{\theta}_2^T, \cdots, {}^{(n)}\boldsymbol{\theta}_{L_Y(n)}^T, 0, 0, \cdots)^T$ and ${}^{(n)}\boldsymbol{\theta}_i = (\theta_{i1}, \cdots, \theta_{iL_X(n)}, 0, 0, \cdots)^T$. Accordingly, estimation reduces to optimization over the finite-dimensional sieve parameter space Θ_n:

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta} \in \Theta_n} \mathbb{P}_n \left[\rho_\tau \left(Y - \sum_{i=1}^{L_Y(n)} f_i(X; \boldsymbol{\theta}_i)\phi_i^{(Y)} \right) \right], \tag{2.8}$$

where Θ_n is defined such that for all $\boldsymbol{\theta} \in \Theta_n$, $(\boldsymbol{\theta})_{ij} = 0$ if $i > L_Y(n)$ or $j > L_X(n)$.

Based on the first- and second-order Fréchet derivatives given in Equations (S2.5) and (S2.6), we develop an optimization algorithm for computing the sieve FoF Geometric Quantile Regression (FGQR) estimator. The complete procedure is presented in Algorithm 1, and additional details are provided in Subsection S2.1 of the Supplementary Material. Furthermore, given the important role of the truncation dimensions $L_X(n)$ and $L_Y(n)$ in the estimation procedure, we select them jointly by cross-validation in this paper.

Statistica Sinica: Newly accepted Paper

The foregoing derivation is presented for the case where the basis functions are known. However, in practice, the true basis functions are typically unknown. To address this issue, we employ the FPCA approach to obtain estimates of the underlying basis functions (He et al., 2010; Kato, 2012). These estimated basis functions are then used in place of the true ones in the subsequent analysis. Specifically, we define $\hat{f}_i(X;\boldsymbol{\theta}_i) = f_i\left(\sum_{j=1}^{\infty}\theta_{ij}\langle X,\hat{\phi}_j^{(X)}\rangle\right)$, where $\hat{\phi}_j^{(X)}$ denotes the FPCA-based estimate of the j-th true eigenfunction $\phi_j^{(X)}$. Similarly, $\hat{\phi}_i^{(Y)}$ represents the FPCA estimate of $\phi_i^{(Y)}$. Following the formulation in (2.8), we estimate the parameter under the estimated basis functions by solving

$$\hat{\hat{\boldsymbol{\theta}}}_n = \arg\min_{\boldsymbol{\theta}\in\Theta_n}\mathbb{P}_n\left[\rho_\tau\left(Y-\sum_{i=1}^{L_Y(n)}\hat{f}_i(X;\boldsymbol{\theta}_i)\hat{\phi}_i^{(Y)}\right)\right]. \tag{2.9}$$

Remark 1. The structured operator class $\mathcal{G}$ in (2.5) is constructed using the pre-specified link functions $\{f_i\}_{i=1}^{\infty}$. Thus, $\mathcal{O}^*\in\mathcal{G}$ ensures correct specification of the structured operator class. If the pre-specified link functions are misspecified so that $\mathcal{O}^*\notin\mathcal{G}$, estimation within $\mathcal{G}$ introduces an additional error due to link-function misspecification. Specifically, minimizing the population risk over $\mathcal{G}$ yields the operator $\mathcal{O}^\dagger\in\mathcal{G}$, rather than the true operator $\mathcal{O}^*$. Consequently, the sieve estimator considered in this subsection targets $\mathcal{O}^\dagger$ instead of $\mathcal{O}^*$, thereby introducing an additional bias of $\|\mathcal{O}^\dagger-\mathcal{O}^*\|_{\mathcal{X}}$. One feasible approach to mitigating this bias is to model the

link functions nonparametrically using B-splines and jointly optimize over spline coefficients and the corresponding parameters. Detailed discussions and a simulation study of the consequences of link function misspecification and the feasibility of nonparametrically estimating the link functions are provided in Section S5 of the Supplementary Material.

3. Asymptotic properties of estimates

This section establishes convergence rates for the sieve estimator with known basis functions in Subsection 3.1 and estimated basis functions in Subsection 3.2.

3.1 Sieve estimator with known basis functions

Let $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^{*T}, \boldsymbol{\theta}_2^{*T}, \cdots)^T$ denote the true parameter vector that satisfies (2.7), where $\boldsymbol{\theta}_i^* = (\theta_{i,1}^*, \theta_{i,2}^*, \cdots)^T$. Let ${}^{(n)}\boldsymbol{\theta}^* = ({}^{(n)}\boldsymbol{\theta}_1^{*T}, {}^{(n)}\boldsymbol{\theta}_2^{*T}, \cdots, {}^{(n)}\boldsymbol{\theta}_{L_Y(n)}^{*T}, 0, 0, \cdots)^T$ be the truncated true parameter, and ${}^{(n)}\boldsymbol{\theta}_i^* = (\theta_{i,1}^*, \cdots, \theta_{i,L_X(n)}^*, 0, 0, \cdots)^T$. To establish the asymptotic theory, we impose the following assumptions:

A_1 For every $\delta > 0$, $\inf_{\boldsymbol{\theta} \in \Theta: \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_2 \geq \delta} \mathbb{E}[\rho_\tau(Y - \mathcal{O}_{\boldsymbol{\theta}}(X))] > \mathbb{E}[\rho_\tau(Y - \mathcal{O}_{\boldsymbol{\theta}^*}(X))]$;

A_2 There exist constants $c_0 > 0$ and $n_0 \geq 1$ such that, for every $n \geq n_0$,

$$\inf_{\substack{\|\boldsymbol{\theta}\|_2 = 1 \\ \boldsymbol{\theta} \in \Theta_n}} \left\langle \boldsymbol{\theta}, \mathbb{E}D^2 \left[\rho_\tau \left(Y - \sum_{i=1}^{\infty} f_i(X; \boldsymbol{\theta}_i^*) \phi_i^{(Y)} \right) \right] \boldsymbol{\theta} \right\rangle \geq c_0.$$

Statistica Sinica: Newly accepted Paper

A_3 There exists a constant $\epsilon > 0$ such that if $\|\mathcal{O}(X) - \sum_{i=1}^{\infty} f_i(X; \boldsymbol{\theta}_i^*)\phi_i^{(Y)}\| < \epsilon$, then $\mathbb{E}\left(\frac{1}{\|Y - \mathcal{O}(X)\|}\right) < \infty$;

A_4 There exist no $y_1, y_2 \in \mathcal{H}_Y$ such that

$$\mathbb{P}\left(\{y \in \mathcal{H}_Y : y = y_1 + ty_2, t \in \mathbb{R}\} \,|X\right) = 1; \tag{3.10}$$

A_5 The function sequence $\{f_i(\cdot)\}_{i=1}^{\infty}$ consists of twice differentiable functions. Moreover, there exist constants $C_1, C_2, C_3 > 0$, independent of i, such that f_i, f_i', and f_i'' are Lipschitz continuous with Lipschitz constants C_1, C_2, and C_3, respectively;

A_6 There exists a constant C_4 such that, for any $\boldsymbol{\theta}$ satisfying $\|\boldsymbol{\theta}\|_2 < C_4$, we have $\left\|\sum_{i=1}^{\infty} f_i(X; \boldsymbol{\theta}_i)\phi_i^{(Y)}\right\| < \infty$;

A_7 The random element $X \in \mathcal{H}_X$ satisfies $\mathbb{E}\|X\|^2 < \infty$.

Assumption A_1 ensures that the parameter $\boldsymbol{\theta}^*$ defined in (2.7) is uniquely identifiable. Assumption A_2 requires the Fréchet second derivative of the population loss with respect to $\boldsymbol{\theta}$ to be positive definite at $\boldsymbol{\theta}^*$. More specifically, the curvature lower bound is required to be uniform over the relevant sieve spaces. Both Assumptions A_1 and A_2 are standard in M-estimation theory (Chakraborty and Chaudhuri, 2014; Chowdhury and Chaudhuri, 2019). By the continuity of $D^2\left[\rho_\tau\left(Y - \sum_{i=1}^{\infty} f_i(X; \boldsymbol{\theta}_i)\phi_i^{(Y)}\right)\right]$ with respect

Statistica Sinica: Newly accepted Paper

to $\boldsymbol{\theta}$ and Assumption A_2, there exists a neighborhood of $\boldsymbol{\theta}^*$ on which the corresponding population second derivative remains positive definite. Assumption A_3 holds when $(Y - \mathcal{O}(X))$ is a continuous Gaussian Markov process on an interval under the L^2 norm (Chowdhury and Chaudhuri, 2019). Assumption A_4 guarantees existence and uniqueness of the geometric quantile in Equation (2.4), as is commonly assumed in the related literature (Cardot et al., 2013; Chowdhury and Chaudhuri, 2019). Assumption A_5 ensures that the estimation error of the operator is bounded by the parameter estimation error. This assumption is relatively mild and is satisfied by many smooth Lipschitz link functions. Assumption A_6 further ensures that the operator induced by $\boldsymbol{\theta}$ remains bounded for $\boldsymbol{\theta}$ in a bounded neighborhood. Assumption A_7 also holds for many commonly used functional data structures, such as Gaussian processes with square-integrable covariance functions. For notational convenience, let $L(n) = \max\{L_X(n), L_Y(n)\}$. We then state the following theorem.

Theorem 1. *Suppose that Assumptions A_1–A_7 hold. Then*

$$\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\|_2 = O_p\left(\sqrt{\frac{L(n)^2}{n} + M({}^{(n)}\boldsymbol{\theta}^*) - M(\boldsymbol{\theta}^*)}\right),$$

where

$$M({}^{(n)}\boldsymbol{\theta}^*) := P\left[\rho_\tau\left(Y - \sum_{i=1}^{L_Y(n)} f_i(X; {}^{(n)}\boldsymbol{\theta}_i^*)\phi_i^{(Y)}\right)\right],$$

Statistica Sinica: Newly accepted Paper

$$M(\boldsymbol{\theta}^*) := P\left[\rho_\tau\left(Y - \sum_{i=1}^{\infty} f_i(X;\boldsymbol{\theta}_i^*)\phi_i^{(Y)}\right)\right].$$

We observe that the convergence rate consists of two components. The first part corresponds to the estimation error, given by $\frac{L(n)}{\sqrt{n}}$, where $L(n)^2$ provides an upper bound for the dimension of the truncated parameter space. The second part reflects the approximation error, which arises from projecting the infinite-dimensional operator space onto a finite-dimensional subspace. Due to the continuity of the loss function and the fact that ${}^{(n)}\boldsymbol{\theta}^* \to \boldsymbol{\theta}^*$ as $n \to \infty$, this approximation error term vanishes asymptotically.

3.2 Sieve estimator with unknown basis functions

In this subsection, the basis functions are unknown and must be estimated from the data. Since FPCA eigenfunctions are identifiable only up to sign, throughout the theoretical analysis we choose $\hat{\phi}_j^{(X)}$ and $\hat{\phi}_i^{(Y)}$ such that $\langle\hat{\phi}_j^{(X)}, \phi_j^{(X)}\rangle_{\mathcal{H}_X} \geq 0$ and $\langle\hat{\phi}_i^{(Y)}, \phi_i^{(Y)}\rangle_{\mathcal{H}_Y} \geq 0$, respectively. According to Equation (2.9), the resulting parameter estimate is denoted by $\hat{\hat{\boldsymbol{\theta}}}_n$. We now introduce the assumptions required under this setting.

A_8 The estimated basis functions converge in expectation, satisfying

$$\sum_{i=1}^{L_Y(n)} \mathbb{E}\left\|\hat{\phi}_i^{(Y)} - \phi_i^{(Y)}\right\| + \sum_{j=1}^{L_X(n)} \mathbb{E}\left\|\hat{\phi}_j^{(X)} - \phi_j^{(X)}\right\| \lesssim \alpha_n,$$

Statistica Sinica: Newly accepted Paper

where $\alpha_n = o(1)$.

A_9 For each $i = 1, \ldots, L_Y(n)$ and each $j = 1, \ldots, L_X(n)$, there exist function classes $\mathcal{F}_i^{(Y)}$ and $\mathcal{F}_j^{(X)}$ such that $\hat{\phi}_i^{(Y)} \in \mathcal{F}_i^{(Y)}$ and $\hat{\phi}_j^{(X)} \in \mathcal{F}_j^{(X)}$, and the covering numbers of these classes satisfy

$$\log N\left(\epsilon, \mathcal{F}_i^{(Y)}, L_2(P)\right) \lesssim \epsilon^{-v},$$
$$\log N\left(\epsilon, \mathcal{F}_j^{(X)}, L_2(P)\right) \lesssim \epsilon^{-v},$$

for some constant $v < 2$.

Assumption A_8 states that the sequence of estimated basis functions converges in expectation, where α_n is a quantity related to the specific basis function estimation method. When FPCA is employed, the explicit form of α_n and its derivation can be found in Bosq (2000); Yao et al. (2005), where the latter also imposes similar assumptions. Such entropy conditions in Assumption A_9 are commonly satisfied by many well-studied function spaces. For example, they hold for VC-type classes, where the constant v is related to the VC dimension, and for Sobolev spaces, where v depends on both the dimension of the domain and the smoothness of the functions. In particular, they can be satisfied by ensuring that the estimated basis functions possess sufficient smoothness. For notational convenience, we let $\mathcal{V} := \max\left\{1, \frac{1+v}{2}\right\}$, and now state the following theorem.

Statistica Sinica: Newly accepted Paper

Theorem 2. *Under Assumptions A_1-A_9, and provided that $\frac{L(n)^{\mathcal{V}}}{\sqrt{n}} = o(1)$, we have*

$$\|\hat{\hat{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}^*\|_2 = O_p\left(\sqrt{\frac{L(n)^{2\mathcal{V}}}{n} + \alpha_n^2 + M({}^{(n)}\boldsymbol{\theta}^*) - M(\boldsymbol{\theta}^*)}\right). \tag{3.11}$$

Compared to Theorem 1, the approximation error remains the same, i.e., $M({}^{(n)}\boldsymbol{\theta}^*) - M(\boldsymbol{\theta}^*)$. However, the use of estimated basis functions introduces additional estimation error. When the basis functions employed exhibit sufficient smoothness (i.e., $0 < v \leq 1$), the additional estimation error arises primarily from the estimation of these basis functions, and thus equals α_n. In contrast, when the basis functions are relatively rough (i.e., $1 < v < 2$), apart from the basis estimation error α_n, the use of such insufficiently smooth basis functions introduces an extra error in parameter estimation, yielding a total additional error of $\alpha_n + \frac{L(n)^{\mathcal{V}}}{\sqrt{n}}$.

Remark 2. The convergence rates established above for the parameter estimators also yield corresponding rates for the induced operator estimators. Under the Lipschitz condition on the link functions, the operator estimation error is controlled by the parameter estimation error and, when the basis functions are estimated, the additional basis estimation error. Formal statements and proofs are deferred to Corollaries 1 and 2 of the Supplementary Material.

4. Distributed statistical inference of FoF geometric quantile regression

In this section, we study the estimation of the FoF geometric quantile regression model via a distributed algorithm designed for large sample sizes N under limited memory constraints. Suppose the dataset $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^N$ is randomly divided into K subsets $\mathcal{D}_1, \ldots, \mathcal{D}_K$, each of size $n_k = n = N/K$, and stored on machines $\mathcal{M}_1, \ldots, \mathcal{M}_K$, respectively. In this context, $\mathbb{P}_{n_i}\mathfrak{f} := \frac{1}{n_i}\sum_{j=1}^{n_i}\mathfrak{f}(X_j)$ for $X_j \in \mathcal{D}_i$ denotes the empirical mean over the local dataset $\mathcal{D}_i$, and $\mathbb{P}_N\mathfrak{f}$ denotes the empirical mean over the full dataset. Throughout this section, we focus on the setting with unknown basis functions.

4.1 Distributed FPCA Algorithm

In this subsection, we propose a distributed FPCA (DFPCA) algorithm to estimate the eigenfunctions when simultaneous access to the global dataset is unavailable. Let L be the number of retained components. On each machine $k \in [K]$, we compute the first L principal component functions via standard FPCA, denoted as $\{\hat{\phi}_i^{(k)}\}_{i=1}^L$. Each machine then sends its estimates of the first L eigenfunctions to the central machine. For $m =$

Statistica Sinica: Newly accepted Paper

$1, \ldots, L$, define the averaged covariance operator

$$\Sigma^{(m)} = \frac{1}{K} \sum_{k=1}^{K} \hat{\phi}_m^{(k)} \otimes \hat{\phi}_m^{(k)}.$$

For the first component, the central machine computes $\Sigma^{(1)}$ and takes its leading eigenfunction as the estimate $\tilde{\phi}_1$. For each subsequent component $m = 2, \ldots, L$, the central machine computes $\Sigma^{(m)}$ and extracts its leading eigenfunction $V^{(m)}$. To ensure orthogonality with previously estimated eigenfunctions and unit norm, we orthogonalize and normalize $V^{(m)}$ via

$$\tilde{\phi}_m = \frac{\left(I - \sum_{i=1}^{m-1} \tilde{\phi}_i \otimes \tilde{\phi}_i\right) V^{(m)}}{\left\| \left(I - \sum_{i=1}^{m-1} \tilde{\phi}_i \otimes \tilde{\phi}_i\right) V^{(m)} \right\|}.$$

The resulting function $\tilde{\phi}_m$ is then used as the estimate of the m-th eigenfunction. Applying this procedure to all L components yields the estimated eigenfunctions $\{\tilde{\phi}_i\}_{i=1}^{L}$, with the full algorithm presented in Algorithm 2 of the Supplementary Material.

We briefly discuss the differences between DFPCA and existing distributed principal component analysis (PCA) methods, with further details in Section S3 of the Supplementary Material. DFPCA is part of our distributed framework for FoF geometric quantile regression and sequentially estimates the individual eigenfunctions to replace the unknown basis functions in Equation (2.5). In contrast, Fan et al. (2019) and Shen et al. (2026) focus on recovering the leading eigenspace. Chen et al. (2022) also com-

Statistica Sinica: Newly accepted Paper

pute the leading eigenvectors sequentially, but use a multi-round distributed procedure in which the central and local machines repeatedly exchange information to update the current eigenvector estimate. However, DFPCA does not require iterative communication between the central and local machines. Once the local eigenfunction estimates have been transmitted, the central machine completes all subsequent estimation steps. Furthermore, Fan et al. (2019), Chen et al. (2022), and Shen et al. (2026) consider distributed PCA for multivariate data, whereas DFPCA is developed for functional data. For large-scale functional data, Yan and He (2026) use Poisson subsampling and decentralized communication to construct a covariance operator estimate, from which the eigenfunctions are subsequently obtained. DFPCA instead estimates eigenfunctions directly, without transmitting local covariance operators.

4.2 Smoothed Distributed Algorithm for FoF Geometric Quantile Regression

The parameter estimation procedure described in Equation (2.8) and its corresponding algorithm cannot be directly applied in distributed environments, as their reliance on the global dataset $\{(X_i, Y_i)\}_{i=1}^{N}$ would incur substantial communication overhead. To address these challenges, we develop

Statistica Sinica: Newly accepted Paper

a parameter estimation framework tailored for distributed settings.

We now propose a loss function suitable for FoF Geometric Quantile Regression in distributed environments. Specifically, we define the smoothed check function as

$$\rho_{n,\tau}(z) := \mathcal{K}_n(z) + \langle \tau, z \rangle, \tag{4.12}$$

where the smoothing kernel function $\mathcal{K}_n(z)$ is defined by

$$\mathcal{K}_n(z) = \|z\| \cdot \mathbb{I}\{\|z\| > h_n\} + \left(\frac{1}{h_n}\|z\|^2 - \frac{1}{3h_n^2}\|z\|^3 + \frac{1}{3}h_n \right) \cdot \mathbb{I}\{\|z\| \leq h_n\}, \tag{4.13}$$

and h_n is a smoothing parameter satisfying $h_n \to 0$ as $n \to \infty$. The proposed smooth kernel is specifically constructed to ensure that both the first- and second-order Fréchet derivatives of the loss function satisfy Lipschitz continuity.

Since we consider the case where the basis functions are unknown, we first estimate the basis functions using the DFPCA algorithm introduced in Subsection 4.1. Based on these estimated basis functions, we define the following smoothed loss function:

$$\tilde{m}_{n,\boldsymbol{\theta}} := \rho_{n,\tau}\left(Y - \sum_{i=1}^{L_Y(n)} \tilde{f}_i(X; \boldsymbol{\theta}_i)\tilde{\phi}_i^{(Y)} \right), \tag{4.14}$$

where $\tilde{f}_i(X; \boldsymbol{\theta}_i) = f_i(\sum_{j=1}^{\infty} \theta_{i,j}\langle X, \tilde{\phi}_j^{(X)} \rangle)$. The final form of the smoothed

Statistica Sinica: Newly accepted Paper

surrogate loss function becomes:

$$\mathbb{P}_{n_1}\tilde{m}_{n,\boldsymbol{\theta}} - \left\langle \boldsymbol{\theta}, \mathbb{P}_{n_1} D\tilde{m}_{n,\bar{\boldsymbol{\theta}}} - \mathbb{P}_N D\tilde{m}_{n,\bar{\boldsymbol{\theta}}} \right\rangle, \tag{4.15}$$

where $\bar{\boldsymbol{\theta}}$ serves as the surrogate expansion point. Accordingly, the smoothed distributed estimator is defined as

$$\tilde{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\Theta_n} \mathbb{P}_{n_1}\tilde{m}_{n,\boldsymbol{\theta}} - \left\langle \boldsymbol{\theta}, \mathbb{P}_{n_1} D\tilde{m}_{n,\bar{\boldsymbol{\theta}}} - \mathbb{P}_N D\tilde{m}_{n,\bar{\boldsymbol{\theta}}} \right\rangle. \tag{4.16}$$

Based on this formulation, we develop a communication-efficient Smoothed Distributed FoF Geometric Quantile Regression (SDFGQR) algorithm, applying Newton's method to iteratively solve the optimization problem in Equation (4.16).

We first estimate the basis functions by applying Algorithm 2 to the predictor and response samples, respectively, and then obtain an initial estimator on $\mathcal{M}_1$ using Algorithm 1. At each iteration, the current estimate is broadcast to all machines, the local gradients are computed and aggregated, and the central machine updates the parameter by minimizing the smoothed surrogate objective in (4.16). The full procedure is summarized in Algorithm 3 of the Supplementary Material.

Statistica Sinica: Newly accepted Paper

4.3 Asymptotic properties

We begin by analyzing the convergence rate of the proposed distributed FPCA estimator. Let $\{\tilde{\phi}_i\}_{i=1}^L$ denote the collection of estimated eigenfunctions obtained via DFPCA algorithm proposed in Subsection 4.1. For comparison, let $\{\hat{\phi}_i\}_{i=1}^L$ represent the eigenfunctions estimated from a single machine using standard FPCA. Correspondingly, let $\{\phi_i\}_{i=1}^L$ be the set of true eigenfunctions associated with the covariance operator Σ. Define $\lambda_1 \geq \lambda_2 \geq \cdots$ as the ordered eigenvalues of Σ. For each $i \in [L]$, similarly define $\lambda_{i,1}^* \geq \lambda_{i,2}^* \geq \cdots$ as the ordered eigenvalues of the operator $\mathbb{E}[\hat{\phi}_i \otimes \hat{\phi}_i]$. The following assumptions are introduced to develop the asymptotic theory.

B_1 The random element $Z \in \mathcal{H}_Z$ has a covariance operator Σ and satisfies the moment condition $\mathbb{E}\|Z\|^4 < \infty$.

B_2 Define the function class $\mathcal{F} := \{Z \in \mathcal{H}_Z : \|Z\| \leq 1\}$. The covering number of $\mathcal{F}$ with respect to the $L_2(P)$ norm satisfies

$$\log N\left(\epsilon, \mathcal{F}, L_2(P)\right) \lesssim \epsilon^{-v},$$

for some constant $v < 2$.

Assumption B_1 imposes a stronger moment condition compared to Assumption A_7. Nevertheless, this assumption remains mild. In particular, for

Statistica Sinica: Newly accepted Paper

a centered Gaussian process Z with covariance operator Σ, Assumption B_1 is fulfilled whenever the sequence of eigenvalues $\{\lambda_i\}_{i=1}^{\infty}$ of Σ satisfies both $\sum_{i=1}^{\infty}\lambda_i < \infty$ and $\sum_{i=1}^{\infty}\lambda_i^2 < \infty$. This guarantees that $\|Z\|^2 = \sum_{i=1}^{\infty}\lambda_i\xi_i^2$ has a finite second moment, where ξ_i are uncorrelated standard normal variables. Moreover, Assumption B_1 is naturally satisfied if the random variable $\|Z\|$ is sub-Gaussian. Assumption B_2 is conceptually analogous to Assumption A_9 and is satisfied when $\mathcal{F}$ is a VC-type class or the unit ball of a Sobolev space.

We now state the following theorem.

Theorem 3. *Suppose Assumptions B_1 and B_2 hold. Assume further that $\lambda_i > \lambda_{i+1}$ and $\lambda_{i,1}^* > \lambda_{i,2}^*$ for every $i \in [L]$, and define $\Delta_i := \min\{\lambda_{i-1} - \lambda_i, \lambda_i - \lambda_{i+1}\}$ for $2 \le i \le L$. Then, the estimation error of the i-th eigenfunction satisfies the following bounds. In particular,*

$$\mathbb{E}\|\tilde{\phi}_1 - \phi_1\| \lesssim \frac{1}{(\lambda_{1,1}^* - \lambda_{1,2}^*)(\lambda_1 - \lambda_2)}\frac{1}{\sqrt{N}} + \frac{1}{(\lambda_1 - \lambda_2)^2}\frac{1}{n}, \tag{4.17}$$

and for all $2 \le i \le L$,

$$\mathbb{E}\|\tilde{\phi}_i - \phi_i\| \lesssim \frac{1}{(\lambda_{i,1}^* - \lambda_{i,2}^*)\Delta_i}\frac{1}{\sqrt{N}} + \frac{1}{\Delta_i^2}\frac{1}{n} + \sum_{j=1}^{i-1}\mathbb{E}\|\tilde{\phi}_j - \phi_j\|. \tag{4.18}$$

Theorem 3 provides upper bounds on the first-moment estimation errors of the eigenfunctions obtained from the proposed distributed FPCA algorithm. For the leading eigenfunction $\tilde{\phi}_1$, the error bound consists of

Statistica Sinica: Newly accepted Paper

two distinct components. The first component, proportional to $\frac{1}{\sqrt{N}}$, captures the statistical estimation error and aligns with the convergence rate achieved by global FPCA. However, the factor $\frac{1}{(\lambda^*_{1,1}-\lambda^*_{1,2})(\lambda_1-\lambda_2)}$ in the first component contains an additional term, $\frac{1}{(\lambda^*_{1,1}-\lambda^*_{1,2})}$, which is absent in the global setting. This discrepancy arises because the distributed setting does not have access to the full data for directly estimating the global covariance operator $\mathbb{E}(X\otimes X)$, and must instead rely on the surrogate $\mathbb{E}(\hat{\phi}_1\otimes\hat{\phi}_1)$. The second component of the error arises from the approximation error between $\mathbb{E}(\hat{\phi}_1\otimes\hat{\phi}_1)$ and $\phi_1\otimes\phi_1$, which can be interpreted as a bias term. This bias becomes negligible when the number of local samples per machine is sufficiently large, specifically when $\sqrt{\frac{K}{n}} < \frac{(\lambda_1-\lambda_2)}{(\lambda^*_{1,1}-\lambda^*_{1,2})}$. For higher-order eigenfunctions $\tilde{\phi}_i$ with $2\le i\le L$, the error bound exhibits a recursive dependence on the estimation errors of the preceding eigenfunctions due to the orthogonalization step.

We next present the convergence rate of the communication-efficient smoothed sieve distributed algorithm. This result requires several additional assumptions.

A_{10} Let $p>2$ and assume that $\mathbb{E}\|X\|^{3p}$ and $\mathbb{E}\left(\frac{1}{\|Y-\sum_{i=1}^{\infty} f_i(X;\boldsymbol{\theta}^*_i)\phi_i^{(Y)}\|}\right)^p$ are finite.

Statistica Sinica: Newly accepted Paper

A_{11} The estimated basis functions converge in expectation, satisfying

$$\sum_{i=1}^{L_Y(n)} \mathbb{E}\left\|\tilde{\phi}_i^{(Y)} - \phi_i^{(Y)}\right\| + \sum_{j=1}^{L_X(n)} \mathbb{E}\left\|\tilde{\phi}_j^{(X)} - \phi_j^{(X)}\right\| \lesssim \kappa_n,$$

where $\kappa_n = o(1)$.

Although Assumption A_{10} imposes stronger conditions on the distributions of X and Y, it remains relatively mild. In particular, if $\|X\|$ is sub-Gaussian, then all its moments exist, ensuring the finiteness of $\mathbb{E}\|X\|^{3p}$. The second part of the assumption ensures that the residual norm $\|\mathcal{E}\|$ does not concentrate too heavily near zero. Notably, this condition is satisfied by any continuous Gaussian Markov process defined on an interval and equipped with the L_2-norm (Chowdhury and Chaudhuri, 2019).

We define $\boldsymbol{\theta}_n^*$ as the true parameter under the smoothed population loss function:

$$\boldsymbol{\theta}_n^* = \arg\min_{\boldsymbol{\theta}\in\Theta} P\left[\rho_{n,\tau}\left(Y - \sum_{i=1}^{\infty} f_i(X;\boldsymbol{\theta}_i)\phi_i^{(Y)}\right)\right], \tag{4.19}$$

and let $M_n(\boldsymbol{\theta}) = P\left[\rho_{n,\tau}\left(Y - \sum_{i=1}^{\infty} f_i(X;\boldsymbol{\theta}_i)\phi_i^{(Y)}\right)\right]$. Here, $\boldsymbol{\theta}_n^* = (\boldsymbol{\theta}_{n,1}^{*T}, \boldsymbol{\theta}_{n,2}^{*T}, \ldots)^T$ with each component $\boldsymbol{\theta}_{n,i}^* = (\theta_{n,i,1}^*, \theta_{n,i,2}^*, \ldots)^T$. We also denote the truncated version by ${}^{(n)}\boldsymbol{\theta}_n^* = ({}^{(n)}\boldsymbol{\theta}_{n,1}^{*T}, \ldots, {}^{(n)}\boldsymbol{\theta}_{n,L_Y(n)}^{*T}, 0, 0, \ldots)^T$, where ${}^{(n)}\boldsymbol{\theta}_{n,i}^* = (\theta_{n,i,1}^*, \ldots, \theta_{n,i,L_X(n)}^*, 0, 0, \ldots)^T$. Let $\eta_0 = (\{\phi_j^{(X)}\}_{j=1}^{\infty}, \{\phi_i^{(Y)}\}_{i=1}^{\infty})$ denote the true basis functions, and write $m_{n,\boldsymbol{\theta},\eta_0} := \rho_{n,\tau}\left(Y - \mathcal{O}_{\boldsymbol{\theta}}(X)\right)$ for notational convenience.

Statistica Sinica: Newly accepted Paper

Define $\tilde{f}_i(X;\boldsymbol{\theta}_i) = f_i(\sum_{j=1}^{\infty}\theta_{i,j}\langle X,\tilde{\phi}_j^{(X)}\rangle)$ and $\tilde{\mathcal{O}}_n(X) := \sum_{i=1}^{L_Y(n)}\tilde{f}_i(X;\tilde{\boldsymbol{\theta}}_{n,i})\tilde{\phi}_i^{(Y)}$. The following theorem establishes the asymptotic behavior of the operator $\tilde{\mathcal{O}}_n$.

Theorem 4. *Suppose Assumptions A_1–A_{11} hold. In addition, let h_n satisfy*

$$h_n^2 > \max\left\{\frac{L(n)^{\mathcal{V}}\,\omega(1)}{\sqrt{N}},\ \sqrt{M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)}\,\omega(1),\ \frac{L(n)}{\sqrt{n}}\,\omega(1)\right\}$$

$$h_n^2 > \frac{8}{1-2C_4}\left\|PD_{\boldsymbol{\theta}}m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\eta_0} - PD_{\boldsymbol{\theta}}m_{n,\boldsymbol{\theta}_n^*,\eta_0}\right\|$$

$$h_n^3 > \kappa_n\,\omega(1)$$

Here, $\omega(1)$ denotes an arbitrary sequence diverging to infinity. Then we have

$$\|\tilde{\mathcal{O}}_n - \mathcal{O}^*\|_{\mathcal{X}} = O_p\left(\frac{L(n)}{\sqrt{n}h_n}\sqrt{\frac{L(n)^{2\mathcal{V}}}{n} + \kappa_n^2 + M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)}\right.$$
$$\left. + \sqrt{\frac{L(n)^{2\mathcal{V}}}{N} + \kappa_n^2 + M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)} + h_n^p\right). \tag{4.20}$$

The convergence rate presented in Theorem 4 consists of three components. First, for the leading term, the choice of h_n ensures that $\frac{L(n)}{h_n\sqrt{n}} = o(1)$. Therefore, as long as the number of machines is not excessively large, this term becomes asymptotically negligible compared to the second term. In particular, it suffices to require that the number of machines satisfies $K < \frac{h_n^2\sqrt{n}}{L(n)^2}$ to guarantee that the first term is asymptotically dominated.

Statistica Sinica: Newly accepted Paper

Next, we examine the second and third terms. Suppose the same set of basis functions is used in both the distributed and oracle settings. In the oracle framework, we define the smoothed estimator of the parameter as

$$\hat{\tilde{\boldsymbol{\theta}}}_n = \arg\min_{\boldsymbol{\theta}\in\Theta_n} \mathbb{P}_N \left[\rho_{n,\tau} \left(Y - \sum_{i=1}^{L_Y(n)} \tilde{f}_i(X;\boldsymbol{\theta}_i)\tilde{\phi}_i^{(Y)} \right) \right]. \tag{4.21}$$

Based on $\hat{\tilde{\boldsymbol{\theta}}}_n$ and the estimated basis functions, we can similarly construct the oracle version of the Smoothed FoF Geometric Quantile Regression (SFGQR) operator estimator, denoted as $\hat{\tilde{\mathcal{O}}}(X) := \sum_{i=1}^{L_Y(n)} \tilde{f}_i(X;\hat{\tilde{\boldsymbol{\theta}}}_{n,i})\tilde{\phi}_i^{(Y)}$. Under this construction, the second and third terms jointly determine the convergence rate of $\hat{\tilde{\mathcal{O}}}$ to the true operator $\mathcal{O}^*$, where the second term quantifies the estimation error in estimating $\boldsymbol{\theta}_n^*$ and the third term reflects the smoothing bias. It follows that, given the same basis functions, the convergence rate of the proposed SDFGQR estimator is asymptotically equivalent to that of the SFGQR.

Moreover, when the parameter p in Assumption A_{10} is sufficiently large, the third term becomes negligible relative to the second. Under such conditions, the primary difference in convergence rate between SDFGQR and FGQR arises solely from the estimation error of the basis functions and the corresponding approximation error.

Statistica Sinica: Newly accepted Paper

5. Simulation

In this section, we conduct Monte Carlo simulations with 500 repetitions to evaluate the effectiveness of the proposed methodology.

We now describe the data generation process used in the simulations. Both the covariate $X(s)$ and the response $Y(t)$ belong to $\mathcal{L}_2([0,1])$, the Hilbert space of square-integrable functions on $[0,1]$. To generate $X(s)$, we specify its covariance function as $c_X(s_1,s_2) = \exp\{-(|s_1-s_2|/\sigma)^{1.2}\}$ and compute the first ten eigenpairs $\{(\lambda_i^{(X)}, \phi_i^{(X)})\}_{i=1}^{10}$. We then generate $X(s) = \sum_{i=1}^{10} \xi_i^{(X)} \phi_i^{(X)}(s)$, where the scores $\xi_i^{(X)} \sim N(0, \lambda_i^{(X)})$ are independent. Next, we generate the error process $\mathcal{E}(t)$ independently of X, using the same procedure with the covariance function $c_{\mathcal{E}}(t_1,t_2) = \exp\{-(|t_1-t_2|/\sigma)^{1.6}\}$. The response $Y(t)$ is then generated from the linear model

$$Y(t) = \beta_0(t) + \int_0^1 \beta(t,s)X(s)ds + \mathcal{E}(t). \tag{5.22}$$

For each Monte Carlo replication and each geometric quantile index τ, we jointly select the truncation dimensions for FGQR and SDFGQR by five-fold cross-validation, using ρ_τ to compute the average validation loss. For SDFGQR, we also select the smoothing parameter jointly with the truncation dimensions (see Subsection S4.1 of the Supplementary Material for implementation details).

Statistica Sinica: Newly accepted Paper

A study on the effect of sample size on FGQR is reported in Subsection S7.2 of the Supplementary Material. The results show that the estimation error decreases substantially as the sample size increases, while the effect of the geometric quantile index remains limited.

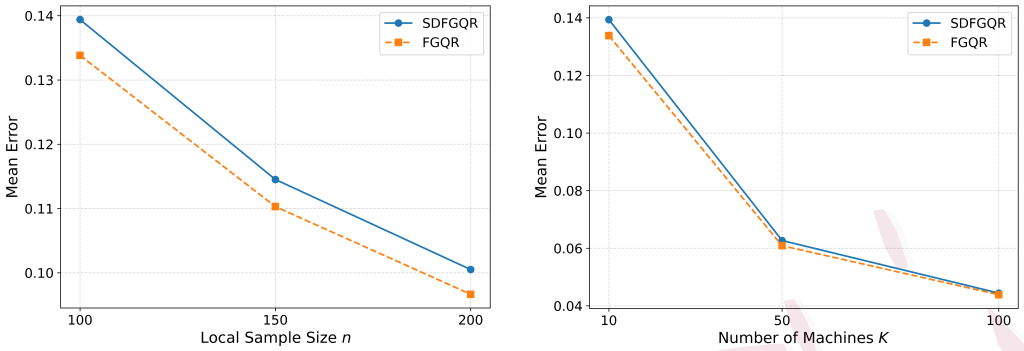
5.1 Estimation Accuracy in Distributed Settings

In this subsection, we verify the effectiveness of the DFPCA method proposed in Subsection 4.1 and the SDFGQR algorithm proposed in Subsection 4.2.

The proposed DFPCA method achieves estimation accuracy close to that of Oracle FPCA across a range of settings, with the estimation error decreasing as either the local sample size or the number of machines increases. This confirms the effective information aggregation and scalability of DFPCA in distributed environments. Detailed results are provided in Subsections S7.3 and S7.4 of the Supplementary Material.

We therefore focus here on the estimation accuracy of SDFGQR relative to FGQR. At $\tau = 0$, we fix $K = 10$ and vary the local sample size over $n = 100, 150, 200$, and then fix $n = 100$ and vary the number of machines over $K = 10, 50, 100$. Figure 1 shows that the mean error (ME; defined in Equation (S7.17) of the Supplementary Material) of SDFGQR decreases as

Statistica Sinica: Newly accepted Paper



(a) Impact of local sample size on ME (b) Impact of machine number on ME

Figure 1: Comparison of SDFGQR and FGQR under different settings: (a) varying local sample size with fixed $K = 10$; (b) varying machine number with fixed $n = 100$.

either the local sample size or the number of machines increases. Although FGQR consistently achieves slightly smaller ME, the absolute gap remains small and narrows as the local sample size grows, indicating that SDFGQR attains estimation accuracy close to that of the global estimator.

5.2 Comparisons

In this subsection, we compare our proposed method with several existing approaches. The benchmark methods used for comparison are Nonparametric Depth and Quantile Regression for Functional Data (NDQR; Chowdhury and Chaudhuri, 2019), Penalized Function-on-Function Linear Quantile Re-

Statistica Sinica: Newly accepted Paper

gression (PFLQR; Beyaztas et al., 2025), and Function-on-Function Partial Quantile Regression (FPQR; Mutis et al., 2025). A more detailed introduction to these benchmark methods is provided in Subsection S1.5 of the Supplementary Material.

In this subsection, we fix the total sample size at $N = 10{,}000$ and set the number of machines for SDFGQR to $K = 10$. We then compare the performance of the proposed and benchmark methods at several pointwise quantile levels q. However, FGQR, SDFGQR, and NDQR estimate conditional geometric quantiles in a Hilbert space, whereas PFLQR and FPQR estimate pointwise conditional quantile functions. Therefore, to enable a direct comparison, we calibrate a corresponding geometric quantile index τ_q so that the true conditional geometric quantile coincides with the true pointwise conditional quantile function. Details of this calibration are provided in Subsection S4.2 of the Supplementary Material. With this calibration, we evaluate the estimation accuracy of all five methods at $q \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$ using the mean relative error (MRE; see Equation (S4.9) in Subsection S4.2) on an independent test sample. The results are summarized in Table 1.

Table 1 shows that FGQR and SDFGQR yield the smallest MREs for estimating the common conditional quantile target across all five values

Table 1: Comparison of different methods

Pointwise quantile level q	FGQR	SDFGQR	NDQR	PFLQR	FPQR
0.10	**0.0208**	**0.0210**	0.6132	0.8092	0.7577
0.25	**0.0232**	**0.0232**	0.5949	0.7904	0.8954
0.50	**0.0248**	**0.0249**	0.6133	0.8277	0.9607
0.75	**0.0200**	**0.0202**	0.5515	0.8042	0.8096
0.90	**0.0180**	**0.0183**	0.5462	0.7667	0.6727

of q. These results demonstrate that the proposed methods provide substantially more accurate estimates of conditional geometric quantiles than the competing methods, not only at the median but also at non-central quantiles.

6. Real Case Study

In this section, we demonstrate the superiority of the proposed method using data gathered from the oil pipeline transportation process.

Our dataset comprises measurements from 9 stations within an oil and gas pipeline network in China, recorded at various time points. Each station is equipped with sensors to record oil density, pressure, and temperature

measurements every six seconds. We treat density as the functional response and regard $X = (X_1, X_2)$ as the joint functional predictor taking values in the Hilbert direct sum $\mathcal{H}_X := \mathcal{H}_{X_1} \oplus \mathcal{H}_{X_2}$, where $X_1 \in \mathcal{H}_{X_1}$ and $X_2 \in \mathcal{H}_{X_2}$ denote the pressure and temperature trajectories, respectively. We obtain the datasets of nine stations, denoted as $\mathcal{D}_1, \cdots, \mathcal{D}_9$. They are combined to form the global dataset, represented as $\mathcal{D} := \bigcup_{k=1}^{9} \mathcal{D}_k$.

We perform median regression using FGQR, SDFGQR, PFLQR, NDQR, and FPQR. The geometric quantile index is set to $\tau = 0$ for FGQR, SDFGQR, and NDQR, while the pointwise quantile level is set to $q = 0.5$ for PFLQR and FPQR. Only SDFGQR uses distributed data, while the other methods are fitted to the global dataset. Table 2 reports the relative prediction errors on the test set. FGQR achieves the smallest error of 6.98×10^{-5}, followed by SDFGQR with an error of 8.40×10^{-5}. Further details of the

Table 2: Relative Prediction Error

	FGQR	SDFGQR	PFLQR	NDQR	FPQR
Relative Prediction Error	6.98×10^{-5}	8.40×10^{-5}	0.003198	0.000301	1.68×10^{-4}

real case study are provided in Section S8 of the Supplementary Material.

Statistica Sinica: Newly accepted Paper

7. Conclusion

This paper introduces a FoF geometric quantile regression with a structured sieve estimation procedure based on the concept of geometric quantiles in Hilbert spaces. Based on the proposed model, we investigate the consistency and convergence rates of the constructed FoF geometric quantile regression estimator in scenarios with both known and unknown basis functions.

Furthermore, to address the computational and storage demands of large-scale functional data, we extend the proposed FoF geometric quantile regression framework to a distributed computing environment. First, we develop a DFPCA algorithm that achieves estimation accuracy comparable to global FPCA while reducing communication costs. Second, we develop a communication-efficient distributed FoF geometric quantile regression algorithm using a smoothed surrogate loss, achieving a convergence rate comparable to that of the global estimator under the stated conditions.

Supplementary Materials

The Supplementary Material contains preliminaries, detailed algorithms, a discussion of the differences between DFPCA and existing distributed PCA methods, implementation details for the simulation studies, discussions of link function misspecification and nonparametric link modeling, operator

Statistica Sinica: Newly accepted Paper

convergence rates, additional simulation results, further details of the real case study, and complete proofs of the theoretical results.

Acknowledgements

We thank the anonymous referees, the Associate Editor, and the Editor for their constructive comments, which have helped improve the quality of this paper. This work is supported in part by the National Natural Science Foundation of China (72301108), the State Key Program of the National Natural Science Foundation of China (72531003 and 72331005), the National Key R&D Program of China (2021YFA1000100 and 2021YFA1000101), the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (JYB2025XDXM904), the Shanghai Pilot Program for Basic Research (TQ20240201), the Natural Science Foundation of Shanghai (23JS1400500), and the Shanghai Municipal Education Commission (2024AI01002).

References

Benatia, D., M. Carrasco, and J.-P. Florens (2017). Functional linear regression with functional response. *Journal of Econometrics 201*(2), 269–291.

Beyaztas, U., H. L. Shang, and A. Alin (2022). Function-on-function partial quantile regression.

Statistica Sinica: Newly accepted Paper

Journal of Agricultural, Biological and Environmental Statistics 27, 149–174.

Beyaztas, U., H. L. Shang, and S. Saricam (2025). Penalized function-on-function linear quantile regression. *Computational Statistics 40*, 1–29.

Bosq, D. (2000). *Linear processes in function spaces: theory and applications*, Volume 149. Springer Science & Business Media.

Cardot, H., P. Cénac, and P.-A. Zitt (2013). Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm. *Bernoulli 19*(1), 18–43.

Chakraborty, A. and P. Chaudhuri (2014). The spatial distribution in infinite dimensional spaces and related quantiles and depths. *Annals of Statistics 42*, 1203–1231.

Chen, X., J. D. Lee, H. Li, and Y. Yang (2022). Distributed estimation for principal component analysis: An enlarged eigenspace analysis. *Journal of the American Statistical Association 117*(540), 1775–1786.

Chowdhury, J. and P. Chaudhuri (2019). Nonparametric depth and quantile regression for functional data. *Bernoulli 25*, 395–423.

Dong, C., R. Chen, Z. Xiao, and W. Liu (2024). Functional quantile autoregression. *Journal of Econometrics 244*(2), 105765.

Fan, J., D. Wang, K. Wang, and Z. Zhu (2019). Distributed estimation of principal eigenspaces. *The Annals of Statistics 47*(6), 3009 – 3031.

Statistica Sinica: Newly accepted Paper

He, G., H. G. Müller, J. L. Wang, and W. Yang (2010). Functional linear regression via canonical analysis. *Bernoulli 16*, 705–729.

Horváth, L., P. Kokoszka, and S. Wang (2021). Monitoring for a change point in a sequence of distributions. *Annals of Statistics 49*, 2271–2291.

Hu, B., X. Hu, H. Liu, J. You, and J. Cao (2024). Simultaneous functional quantile regression. *Statistica Sinica 34*(2), 867–888.

Jordan, M. I., J. D. Lee, and Y. Yang (2019). Communication-efficient distributed statistical inference. *Journal of the American Statistical Association 114*(526), 668–681.

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

Konen, D. (2025). PDE characterization of geometric distribution functions and quantiles. *Bernoulli 31*, 2077–2104.

Li, B., A. T. K. Wan, M. Yu, and Y. Zhou (2026). *Supplement to "Function-on-Function Geometric Quantile Regression with a Distributed Extension"*.

Li, C., L. Xiao, and S. Luo (2022). Joint model for survival and multivariate sparse functional data with application to a study of Alzheimer's Disease. *Biometrics 78*, 435–447.

Li, Y., K. Chen, X. Zheng, and C. Y. Yau (2024). Functional threshold autoregressive model. *Statistica Sinica 34*(2), 817–836.

Liu, Y., M. Li, and J. S. Morris (2020). Function-on-scalar quantile regression with application

Statistica Sinica: Newly accepted Paper

to mass spectrometry proteomics data. *Annals of Applied Statistics 14*(2), 521.

Ma, H., T. Li, H. Zhu, and Z. Zhu (2019). Quantile regression for functional partially linear model in ultra-high dimensions. *Computational Statistics & Data Analysis 129*, 135–147.

Mutis, M., U. Beyaztas, F. Karaman, and H. L. Shang (2025). On function-on-function linear quantile regression. *Journal of Applied Statistics 52*, 814–840.

Shen, S., J. Lu, and X. Lin (2026). Dimension reduction for large-scale federated data: Statistical rate and asymptotic inference. *Journal of the American Statistical Association 121*(553), 585–597.

Yan, X. and S. He (2026). Poisson subsampling for large-scale functional data analysis. *Computational Statistics 41*, 101.

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

Zhang, Z., X. Wang, L. Kong, and H. Zhu (2022). High-dimensional spatial quantile function-on-scalar regression. *Journal of the American Statistical Association 117*(539), 1563–1578.

Zhu, H., Y. Zhang, Y. Li, and H. Lian (2023). Semiparametric function-on-function quantile regression model with dynamic single-index interactions. *Computational Statistics and Data Analysis 182*, 107727.

Bingchen Li, *Miaomiao Yu* and *Yong Zhou*

School of Statistics and Academy of Statistics and Interdisciplinary Sciences, East China Normal University.

Statistica Sinica: Newly accepted Paper

E-mail: bingchenli_math@163.com; mmyu@fem.ecnu.edu.cn; yzhou@fem.ecnu.edu.cn

Alan T. K. Wan

Department of Decision Analytics and Operations, City University of Hong Kong.

E-mail: Alan.Wan@cityu.edu.hk

ORCID:

Miaomiao Yu: 0000-0002-4536-2078

Statistica Sinica