Statistica Sinica Preprint No: SS-2025-0294	
Title	Estimation of Conditional Extremiles in Reproducing Kernel Hilbert Spaces with Application to Large Commercial Banks Data
Manuscript ID	SS-2025-0294
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0294
Complete List of Authors	Fang Chen and Caixing Wang
Corresponding Authors	Caixing Wang
E-mails	wangcaixing96@gmail.com
Notice: Accepted author version.	

Statistica Sinica

ESTIMATION OF CONDITIONAL EXTREMILES IN REPRODUCING KERNEL HILBERT SPACES WITH APPLICATION TO LARGE COMMERCIAL BANKS DATA

Fang Chen[1†], Caixing Wang[2†*]

[1]*Nanjing Forestry University and* [2]*Southeast University*

Abstract: As analogs of quantiles, extremiles are coherent spectral risk measures with explicit formulations and intuitive interpretations. Their inherent sensitivity to the magnitude of extreme outcomes makes them particularly suitable for heavy-tailed data. However, existing extremile estimation methods rarely exploit rich auxiliary covariate information, which limits their ability to capture conditional extreme patterns and to extrapolate reliably at very high risk levels. This paper proposes a new nonparametric framework for estimating conditional extremiles in the presence of multiple covariates. By combining reproducing kernel Hilbert spaces (RKHS) with a quantile regression process approximation, our method flexibly models the conditional extremile structure while enabling reliable extrapolation for heavy-tailed distributions. We establish the non-asymptotic error bound for the estimation error, rigorously justifying its theoretical validity.

* Corresponding author. Email: caixingwang@seu.edu.cn

† The two authors contributed equally to this work.

Simulation studies show that our approach outperforms existing competitors in both efficiency and extrapolation accuracy in heavy-tailed settings. An empirical application to large commercial banks further illustrates its practical value for extreme risk measurement.

Key words and phrases: Asymmetric least squares, extreme value theory, heavy tails, quantile regression.

1. Introduction

Rare events often incur catastrophic consequences, such as massive portfolio losses (Schaumburg, 2012; Odening and Hinrichs, 2003), extreme oil price fluctuations (Marimoutou et al., 2009), and severe meteorological disasters (Dupuis et al., 2015; Friederichs and Hense, 2007). For example, the 2022 U.S. bond market crash resulted in extreme portfolio losses that exceeded quantile-based risk forecasts by 26% (Adrian and Fleming, 2022), while in 2019 the inland flooding caused economic losses globally of USD82 billion due to mis-estimated extreme damage magnitudes (Aon, 2020). These failures, amplified by the 2008 global financial crisis, underscore a critical truth: for heavy-tailed distributions governing extreme events, it is essential to characterize tail quantities rather than central tendencies such as the mean.

Numerous risk metrics have been developed in the literature. However, existing tail risk measures generally fail to simultaneously achieve sensitivity to the magnitude of extreme values, compliance with coherence axioms, and practical interpretability, leaving a pressing methodological gap. Quantiles have long dominated tail risk analysis due to their intuitive threshold interpretation and computational tractability (Embrechts et al., 1997; Linsmeier and Pearson, 2000). However, quantile-based risk measures suffer from two fundamental drawbacks. First, they only exploit information on whether observations fall below or above a given threshold, ignoring the actual magnitudes of large losses (e.g., a \$10 million loss and a \$100 million loss are treated identically if both exceed $q_{0.01}$). Second, quantile-based measures are in general not coherent because they fail subadditivity (Artzner et al., 1999). To address these limitations, an alternative risk measure called expectiles was introduced by Newey and Powell (1987), a least-squares analog of quantiles that integrates both tail probabilities and extreme value magnitudes. In fact, expectiles generate the unique law-invariant coherent risk measure that is also elicitable (Girard et al., 2022). Nevertheless, expectiles lack explicit closed-form expressions, requiring numerical approximation even for simple distributions.

Daouia et al. (2019) introduced extremiles, a tail risk measure that unifies the interpretability of quantiles, the coherence of expectiles, and the sensitivity to extreme magnitudes. Formally, extremiles extend the quantile minimization framework by replacing the absolute deviation with the squared deviation, leveraging a tail-weighted loss function to prioritize extreme values. For a random variable y and a given quantile level $\tau \in (0, 1)$, the quantile q_τ can uniquely be defined as the generalized inverse $q_\tau = F^{-1}(\tau) = \inf\{y : F(y) \geq \tau\}$ of the cumulative distribution function F. Without loss of generality, we assume F is continuous. In fact, we can verify that the quantile q_τ is the median of the random variable Z_τ with the cumulative distribution function $F_{Z_\tau} = K_\tau(F)$, where

$$K_\tau(t) = \begin{cases} 1 - (1-t)^{s(\tau)} & \text{if } 0 < \tau \leq 1/2; \\ t^{r(\tau)} & \text{if } 1/2 \leq \tau < 1, \end{cases}$$

and $r(\tau) = s(1-\tau) = \log(1/2)/\log(\tau)$. Then, the τ-th quantile can be obtained by solving

$$q_\tau = \text{argmin}_\theta \, \mathbb{E}\left\{J_\tau(F(y)) \cdot [|y-\theta| - |y|]\right\},$$

where the weight-generating function is $J_\tau(\cdot) = K'_\tau(\cdot)$. The τ-th extremile ξ_τ modifies the objective function by using the squared deviation, yielding that

$$\xi_\tau = \text{argmin}_\theta \, \mathbb{E}\left\{J_\tau(F(y)) \cdot \left[|y-\theta|^2 - |y|^2\right]\right\},$$

and equivalently, it serves as the mean of the transformed tail variable Z_τ with $F_{Z_\tau} = K_\tau(F)$.

Asymmetric least squares (ALS) regression (Yao and Tong, 1996) provides a natural way to investigate some higher or lower regions of the sample space. Extremiles are defined on the asymmetric function and own some attractive merits compared with traditional risk measures. Compared with the quantiles, extremiles are comonotonically additive (Acerbi and Szekely, 2014) and depend on both the tail realizations and their probabilities, thus they suggest a better capability of fitting both locations and spread in data points. Compared with the expectiles, extremiles admit several equivalent explicit representations and more transparent interpretations. These estimators are available in closed form and are computationally convenient, so that inference on extremiles is typically simpler than for both expectiles and extreme quantiles. Moreover, extremile estimators strike a favorable balance between the robustness of ordinary quantiles and the high sensitivity of extreme quantiles. For moderate probability levels, they are more tail-sensitive than the corresponding quantiles, while for very high or low levels, they tend to be more stable.

In financial and actuarial applications, responses y (e.g., loss severity) are rarely observed in isolation, they are typically accompanied by p-dimensional covariates $\mathbf{x}$ (e.g., asset portfolio structure, market volatility, geopolitical risk indices) that drive conditional tail behavior. A substantial body of work has therefore focused on the estimation of extreme conditional risk measures, including conditional expectiles and quantiles. In the unconditional heavy-tailed setting, Daouia et al. (2018) and Daouia et al. (2020) study extreme expectile estimators. Wang et al. (2012) considered extreme quantile estimation in a linear regression framework by combining intermediate quantile estimators with extreme value theory, while Wang and Li (2013) relaxed the linearity assumption via a power transformation. For time series data, Yi et al. (2014) and Li and Wang (2019) developed extreme quantile estimators based on extreme value theory. In a covariate-dependent context, Gardes and Girard (2010) proposed an estimator of extreme conditional quantiles using neighborhoods in the covariate space, and Wang and Tsai (2009) studied extreme conditional quantiles under a covariate-dependent extreme value index. More recently, Xu et al. (2022b) introduced a semiparametric approach based on a tail single-index model. A local smoothing method to estimate the conditional extremiles of the response y is also investigated in Daouia et al. (2022), where the number of covariates is required to remain small. They estimated the conditional extremiles by using observations in a small neighborhood of the covariate, and thus the finite sample behavior of the estimate heavily depended on the richness of data in their neighborhood.

This paper proposes a nonparametric framework for estimating the conditional extremiles with multivariate covariates, leveraging reproducing kernel Hilbert spaces (RKHS) and quantile regression processes. In contrast to approaches that impose linearity or strong low-dimensional restrictions, our method is designed to adapt to complex covariate structures. In particular, when using a universal kernel such as the Gaussian kernel, the induced RKHS is rich enough that any continuous function can be approximated arbitrarily well in the sup norm by functions in the corresponding RKHS (Steinwart, 2005), which makes the proposed framework highly flexible for modeling conditional extremiles. Our work is closely related to the RKHS literature on conditional quantiles and expectiles. Li et al. (2007) studied nonparametric estimation of conditional quantile functions in an RKHS, while Zhang et al. (2016) extended kernel quantile regression to incorporate sparsity constraints. Yang et al. (2018) proposed an RKHS-based expectile regression method and developed an efficient algorithm for computing the entire solution path. Building on these ideas, we construct an RKHS estimator specifically tailored to extremiles. The resulting estimator is particularly promising for extreme value analysis, as it can naturally accommodate nonlinearity, non-additivity, and complex interaction effects in the covariates.

The remainder of the paper is organized as follows. Section 2 briefly outlines the problem formulation and the foundational concepts of reproducing kernels. The proposed methodology for estimating intermediate and extreme conditional extremiles is detailed in Section 3 and 4, respectively. And the corresponding theoretical results are established in Section 5. Section 6 compares extremiles with traditional risk measures and discusses their connections to the distortion function K_τ. Section 7 validates the method via simulation studies and a real-world case analysis. Concluding remarks are given in Section 8, and all technical proofs are deferred to the Supplementary Material.

2. Preliminaries

2.1 Conditional Extremiles

Let $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^p$ and $y \in \mathbb{R}$ be random variables following a Borel probability measure $\rho_{\mathbf{x},y}$ on $\mathcal{X} \times \mathbb{R}$. Suppose that $\mathcal{Z}^n = \{(\mathbf{x}_i, y_i) \in \mathcal{X} \times \mathbb{R}, 1 \leq i \leq n\}$ is a collection of n independent copies of $(\mathbf{x}, y)$. We consider the following model

$$y_i = f_0(\mathbf{x}_i) + \varepsilon_i, \quad i = 1, 2 \cdots, n,$$

where the error ε_i satisfies $\mathbb{E}(\varepsilon_i | \mathbf{x}_i) = 0$ and $\mathbb{E}\left(\varepsilon_i^2 | \mathbf{x}_i\right) = \sigma_i^2 < \infty$.

2.1 Conditional Extremiles

The conditional extremiles of the response y is defined as the median of the random variable $Z_\tau^{\mathbf{x}}$ whose distribution function is given by $F_{Z_\tau^{\mathbf{x}}} = K_\tau\left(F_{\mathbf{x}}(y)\right)$, where $F_{\mathbf{x}}(\cdot)$ refers to the conditional cumulative distribution function of the response y given the covariate $\mathbf{x}$. Throughout this paper, we assume that $F_{\mathbf{x}}(\cdot)$ is continuous, and then we define the τ-th conditional extremiles of the response y given the covariate $\mathbf{x}$ analogously as that of Daouia et al. (2022) with the weight-generating function $J_\tau(\cdot) = K_\tau'(\cdot)$, that is

$$\xi_\tau(\mathbf{x}) = \underset{f}{\operatorname{argmin}}\, \mathbb{E}\left\{J_\tau\left(F_{\mathbf{x}}(y)\right)\left[|y - f(\mathbf{x})|^2 - |y|^2\right] \,\middle|\, \mathbf{x}\right\}, \tag{2.1}$$

Thus, it is straightforward to obtain that

$$\xi_\tau(\mathbf{x}) = \frac{\mathbb{E}\left[yJ_\tau(F_{\mathbf{x}}(y)) \middle| \mathbf{x}\right]}{\mathbb{E}\left[J_\tau(F_{\mathbf{x}}(y)) \middle| \mathbf{x}\right]}.$$

Note that $\mathbb{E}\left[J_\tau(F_{\mathbf{x}}(y)) \middle| \mathbf{x}\right] = 1$ by the continuity of $F_{\mathbf{x}}$ for all $\tau \in (0,1)$. Hence, we further obtain that

$$\xi_\tau(\mathbf{x}) = \mathbb{E}\left[yJ_\tau\left(F_{\mathbf{x}}(y)\right) | \, \mathbf{x}\right] = \int_0^1 J_\tau(t) q_t(\mathbf{x}) dt = \int_0^1 q_t(\mathbf{x}) dK_\tau(t), \tag{2.2}$$

where $q_\tau(\mathbf{x}) = F_{\mathbf{x}}^{-1}(\tau) = \inf\{y \in \mathbb{R} | F_{\mathbf{x}}(y) \geq \tau\}$ denotes the τ-th conditional quantile.

Since the quantile $q_\tau(\mathbf{x})$ coincides with the median of the random variable $Z_\tau^{\mathbf{x}}$, an argument analogous to Proposition 2 in Daouia et al. (2019) shows that, whenever $\mathbb{E}\,|Z_\tau^{\mathbf{x}}| < \infty$,

$$\xi_\tau(\mathbf{x}) = \mathbb{E}\left(Z_\tau^{\mathbf{x}}\right).$$

The integrability condition $\mathbb{E}\,|Z_\tau^{\mathbf{x}}| < \infty$ is in turn implied by $\mathbb{E}[|y||\,\mathbf{x}] < \infty$. Hence, extremiles of any order exist only when the response y admits a finite first conditional moment. Throughout this paper, we assume the existence of the first conditional moment of the response, i.e., $\mathbb{E}(|y|) < \infty$. This assumption is satisfied by many heavy-tailed models used in practice, such as Pareto-type distributions with tail index larger than one, which are commonly employed in financial risk measurement.

Since the objective function in (2.1) requires the estimation of certain unknown functions, appropriate penalty terms must be incorporated to avoid overfitting. In general, the regularized extremile regression problem can be written as

$$\min_{f\in\mathcal{F}} \mathbb{E}\left\{J_\tau\left(F_{\mathbf{x}}(y)\right)\left[|y-f(\mathbf{x})|^2\right]\right\} + \lambda T(f),$$

where $\mathcal{F}$ is the functional space of our interests, $T(\cdot)$ is the penalty functional, and λ is a regularization parameter controlling the strength of regularization.

2.2 Reproducing kernels

We focus on estimating conditional extremiles in a reproducing kernel Hilbert space (RKHS) framework and therefore only recall the basic notions needed for our development. For comprehensive treatments, we refer the reader to Wahba (1990), Shawe-Taylor and Cristianini (2004) and Berlinet and Thomas-Agnan (2004). Let $K : \mathcal{X}\times\mathcal{X}\to\mathbb{R}$ be a symmetric and positive definite Mercer kernel. The associated RKHS $\mathcal{H}_K$ is defined as the completion of the linear span of the functions $\{K_{\mathbf{x}}(\cdot) = K(\mathbf{x},\cdot), \mathbf{x}\in\mathcal{X}\}$ with the inner product given by $\langle K_{\mathbf{x}}, K_{\mathbf{u}}\rangle_K = K(\mathbf{x},\mathbf{u})$ for any $\mathbf{x},\mathbf{u}\in\mathcal{X}$. In particular, the RKHS $\mathcal{H}_K$ is uniquely determined by the kernel K.

Given a probability measure $\rho_{\mathbf{x}}$ on $\mathcal{X}$, the space $\mathcal{H}_K$ is continuously embedded in $L^2_{\rho_{\mathbf{x}}}$, where $L^2_{\rho_{\mathbf{x}}} = \{f : \int f^2(\mathbf{x})d\rho_{\mathbf{x}} < \infty\}$. For each $\mathbf{x} \in \mathcal{X}$, the function $\mathbf{z} \mapsto K(\mathbf{x}, \mathbf{z})$ belongs to $\mathcal{H}_K$, and the inner product is chosen so that $K(\mathbf{x}, \cdot)$ acts as the representer of point evaluation:

$$\langle f, K(\mathbf{x}, \cdot)\rangle_K = f(\mathbf{x}), \quad \text{for } f \in \mathcal{H}_K.$$

We denote by $\|g\|_K = \sqrt{\langle g, g\rangle_K}$ and $\|g\|_2 = \left(\int_{\mathcal{X}} g(\mathbf{x})^2 d\rho_{\mathbf{x}}\right)^{1/2}$ as the norm in $\mathcal{H}_K$ and $L^2_{\rho_{\mathbf{x}}}$ respectively. It is worth pointing out that the RKHS induced by some universal kernel, such as the Gaussian kernel, is a fairly large functional space in the sense that any continuous function can be arbitrarily well approximated by an intermediate function in its induced RKHS under the infinity norm (Steinwart, 2005).

Furthermore, we assume that the kernel function $K(\cdot, \cdot)$ is upper bounded by some constant, i.e., $\sup_{\mathbf{x}\in\mathcal{X}} K(\mathbf{x}, \mathbf{x}) \leq \kappa^2$, which is satisfied by many commonly used kernels, such as the Gaussian kernel and polynomial kernels defined on a bounded set. A similar assumption was previously used in Steinwart and Scovel (2007) and Blanchard et al. (2008). Reproducing kernel Hilbert spaces (RKHS) have been extensively used in traditional quantile and expectile regression (Takeuchi et al., 2006; Li et al., 2007; Yang et al., 2018). By allowing for nonlinear, nonadditive, and highly interactive effects among covariates, RKHS-based methods provide a flexible nonparametric framework for approximating complex underlying functions in applications.

3. Estimation of Intermediate Conditional Extremiles

3.1 Problem Formulation

We are interested in estimating extreme conditional extremiles at levels $\tau = \tau'_n$ with $\tau'_n \to 1$ at an extremely high rate. As a first step, we study the estimation problem at an intermediate level

$\tau_n \to 1$ satisfying the usual condition $n(1-\tau_n) \to \infty$, and then extrapolate to more extreme levels.

Assume that, for a fixed τ, the true conditional extremile function ξ_τ belongs to an RKHS $\mathcal{H}_K$ induced by a given kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. The conditional extremiles can subsequently be derived by solving the following optimization problem, which integrates the asymmetric weighted least-squares loss with a penalty term based on the squared Hilbert norm,

$$\min_{f \in \mathcal{H}_K} \mathbb{E}\left\{ J_\tau \left(F_{\mathbf{x}}(y)\right) \left[|y - f(\mathbf{x})|^2\right] \right\} + \lambda \left\|f\right\|_K^2 , \tag{3.3}$$

where $\lambda > 0$ is a regularization parameter and $\|f\|_K = \langle f, f\rangle_K^{1/2}$ is the RKHS-norm of f.

Given a random sample $\mathcal{Z}^n = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$, it is natural to estimate ξ_{τ_n} by the empirical analogue of (3.3),

$$\hat{\xi}_{\tau_n}(\mathbf{x}) = \underset{f \in \mathcal{H}_K}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n J_{\tau_n}\left(\widehat{F}_{\mathbf{x}_i}(y_i)\right)(y_i - f(\mathbf{x}_i))^2 + \lambda_{1n} \|f\|_K^2, \tag{3.4}$$

where $\widehat{F}_{\mathbf{x}}(\cdot)$ is an estimator of $F_{\mathbf{x}}(\cdot)$, and $\lambda_{1n} > 0$ is a regularization parameter controlling the complexity of the model. In this work, $\widehat{F}_{\mathbf{x}}(\cdot)$ is obtained via a quantile regression process approximation, see Section 3.2 for details. Compared with the Nadaraya–Watson kernel estimator (Horváth and Yandell, 1988) used in Daouia et al. (2022), which is essentially restricted to settings with a small number of covariates, our method remains effective when the covariate dimension p is moderate or large. As illustrated in the simulation study in Section 7, the resulting estimator exhibits satisfactory finite-sample performance.

Note that by the representer theorem (Wahba, 1990), the function f in the RKHS can be expressed as a linear combination of the kernel functions $\{K(\mathbf{x}_i, \cdot), i = 1, \ldots, n\}$, that is,

$$f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i K(\mathbf{x}_i, \mathbf{x}) = \boldsymbol{\alpha}\mathbf{K}_n(\mathbf{x}), \tag{3.5}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^\top$ and $\mathbf{K}_n(\mathbf{x}) = (K(\mathbf{x}_1, \mathbf{x}), \ldots, K(\mathbf{x}_n, \mathbf{x}))^\top$. By plugging (3.5) into (3.4), we further obtain an equivalent optimization problem expressed as

$$\widehat{\boldsymbol{\alpha}} = \min_{\alpha} \frac{1}{n} \sum_{i=1}^{n} J_{\tau_n}\left(\widehat{F}_{\mathbf{x}_i}(y_i)\right)\left(y_i - \left(\sum_{j=1}^{n} \alpha_j K(\mathbf{x}_i, \mathbf{x}_j)\right)\right)^2 + \lambda_{1n}\boldsymbol{\alpha}^\top \mathbf{K}\,\boldsymbol{\alpha}, \tag{3.6}$$

where $\mathbf{K}$ is the gram matrix with the elements $[K(\mathbf{x}_i, \mathbf{x}_j)]_{i,j=1}^{n}$. Let

$$\mathbf{W}_{\tau_n} = \mathrm{diag}\big(J_{\tau_n}(\widehat{F}_{\mathbf{x}_1}(y_1)), \ldots, J_{\tau_n}(\widehat{F}_{\mathbf{x}_n}(y_n))\big), \quad \mathbf{y} = (y_1, \ldots, y_n)^\top.$$

A straightforward calculation shows that the solution of (3.6) can be written as

$$\widehat{\boldsymbol{\alpha}} = \left(\mathbf{K}\,\mathbf{W}_{\tau_n}\,\mathbf{K} + n\lambda_{1n}\,\mathbf{K}\right)^{+}\mathbf{K}\,\mathbf{W}_{\tau_n}\,\mathbf{y},$$

where $(\cdot)^+$ denotes the Moore–Penrose generalized inverse.

The proposed method requires $O(n^2)$ memory for storing the matrices $\mathbf{K}$ and $\mathbf{W}_{\tau_n}$, and a time complexity of $O(n^3)$ for solving the optimization problem in (3.6) via matrix inversion operations. This is standard for kernel methods and is adequate for the moderate sample sizes considered in our simulations and empirical application. However, when n is very large, such a cost becomes prohibitively high. To address scalability, we can use popular low–rank kernel approximations such as Nyström approximations (Rudi et al., 2015) or random Fourier features

(Rahimi and Recht, 2007), reducing the storage cost to $O(mn)$ and the computation cost to $O(nm^2)$ with m being the dimension of random center and $m \ll n$. Additionally, when data are stored in different systems, we can adopt distributed approaches within our estimation framework (Lin et al., 2017). Since our objective is a weighted quadratic loss in RKHS, these scalable strategies can be used without altering the form of the extremile estimator and provide a natural route to handle very large n.

Remark 1. In practice, kernel choice is typically guided by reasonable prior beliefs about the smoothness and structure of the target extremile function. For standard real-valued regression or classification, the most standard choice is a Gaussian or Matérn kernel, which are universal on compact sets. Gaussian corresponds to a very smooth prior, while Matérn provides a tunable smoothness scale via its parameter, moving from rough to smooth functions. When a linear structure is plausible, a linear or dot kernel can reduce variance compared with very smooth RBF kernels. When the data have explicit structure (periodicity in time series, spatial correlation), one may use kernels that respect those invariances, such as periodic kernels, spatial Matérn kernel, and so on. For example, in the simulation study presented in Section 7, the underlying extremile function $\xi_\tau(x)$ exhibits high smoothness, rendering the Gaussian kernel both reasonable and effective. In addition, since we care about the heavy-tailed noise setting, kernel choice is less critical than loss choice and robust procedures. However, one may prefer kernels with bounded RKHS norm for simple functions and Lipschitz control, which are typically provided by Gaussian or Matérn kernels.

3.2 Quantile Regression Process Approximation

This section describes how we approximate the conditional distribution function $F_{\mathbf{x}}(\cdot)$ by estimating a finite collection of conditional quantile functions and interpolating them over the quantile levels τ.

We first construct a kernel quantile regression process on a grid of levels $0 < \tau_1 < \cdots < \tau_{s_n} < 1$. Specifically, we solve

$$\widehat{q}_{\tau_1}, \cdots, \widehat{q}_{\tau_{s_n}} = \underset{q_{\tau_1}, \cdots, q_{\tau_{s_n}} \in \mathcal{H}_K}{\operatorname{argmin}} \frac{1}{n s_n} \sum_{j=1}^{s_n} \sum_{i=1}^{n} \rho_{\tau_j} \left(y_i - q_{\tau_j}(\mathbf{x}_i) \right) + \frac{\lambda_{2n}}{s_n} \sum_{j=1}^{s_n} \|q_{\tau_j}\|_K^2, \tag{3.7}$$

where $\rho_\tau(u) = u\{\tau - \mathbf{1}(u \leq 0)\}$ is the check loss function and $\tau_1, \ldots, \tau_{s_n}$ are quantile levels chosen equispaced on $(0, 1)$. According to the representer theorem, each $q_{\tau_j} \in \mathcal{H}_K$ can be written as a finite kernel expansion in terms of the training inputs, so (3.7) reduces to a finite–dimensional optimization problem in $\mathbb{R}^n$, similar to those in (3.4)–(3.6) in Section 3. Although the primal problem has no closed-form solution, it is equivalent to a constrained quadratic program in its dual formulation, which can be solved efficiently using standard kernel quantile regression algorithms, see Takeuchi et al. (2006); Chen et al. (2021) for computational details. We can also consider the random Fourier features and distributed approach to solve the scalability issue in kernel quantile regression (Wang et al., 2024; Wang and Feng, 2024).

To approximate the entire conditional quantile process, we view the collection $\{\widehat{q}_{\tau_j} : j = 1, \ldots, s_n\}$ as evaluations of a curve in τ and interpolate between them by natural linear splines. Let the grid of quantile levels be

$$\Omega = \left\{ \frac{1}{s_n + 1} = \tau_1 < \cdots < \tau_{s_n} = \frac{s_n}{s_n + 1} \right\}.$$

For each fixed covariate vector $\mathbf{x}$, define the estimated conditional quantile curve $\widehat{q}(\tau | \mathbf{x})$ as the natural linear spline in τ with common knots Ω satisfying

$$\widehat{q}(\tau_j|\,\mathbf{x}) = \widehat{q}_{\tau_j}(\mathbf{x}), \qquad j = 1, \ldots, s_n.$$

Thus, $\tau \mapsto \widehat{q}(\tau|\,\mathbf{x})$ is continuous and piecewise linear on $(0,1)$ and, in practice, nearly monotone in τ. Since the collection of conditional quantile functions $\{q_\tau(\mathbf{x}) : \tau \in (0,1)\}$ characterizes the full conditional distribution of Y given $\mathbf{x}$, we can recover an estimate of $F_{\mathbf{x}}(y)$ by inverting the estimated quantile curve. In particular, for any $(\mathbf{x}, y)$ we define

$$\widehat{F}_{\mathbf{x}}(y) = \inf\big\{\tau \in (0,1) : \widehat{q}(\tau|\,\mathbf{x}) \geq y\big\},$$

which can be evaluated numerically by linear interpolation between adjacent knots in Ω. This provides a convenient estimator of the conditional distribution function $F_{\mathbf{x}}(y)$ based on the joint kernel quantile regression process in (3.7).

4. Estimation of Extreme Conditional Extremiles

4.1 Basic Assumptions and Properties

In this section, we first introduce the basic model assumptions and some key properties of extremiles. Let DA$(\cdot)$ denote the maximum domain of attraction of an extreme-value distribution, that is, the class of distribution functions whose suitably normalized maxima converge in distribution to a given extreme–value law. The Fréchet distribution corresponds to heavy–tailed cases and is typically used to model extremely large values when the upper tail decays slowly. It plays a central role in financial and actuarial applications (Embrechts et al., 1997; Resnick, 2007). In what follows, we focus on distributions whose maxima are attracted to the Fréchet family.

More precisely, following Proposition 3 in Daouia et al. (2019), for any fixed $\mathbf{x}$, if $F_{Y|\mathbf{x}} \in \mathrm{DA}\big(\Phi_{\gamma(\mathbf{x})}\big)$, where $\Phi_{\gamma(\mathbf{x})}(y) = \exp\{-y^{-1/\gamma(\mathbf{x})}\}$ denotes the Fréchet distribution and $\gamma(\mathbf{x}) < 1$, then

$$\frac{\xi_\tau(\mathbf{x})}{q_\tau(\mathbf{x})} \sim \Gamma\big(1-\gamma(\mathbf{x})\big)\,\{\log 2\}^{\gamma(\mathbf{x})} \qquad \text{as } \tau \to 1, \tag{4.8}$$

where $\Gamma(\cdot)$ is the Gamma function. According to De Haan and Ferreira (2006), the model assumption $F_{Y|x} \in DA\left(\Phi_{\gamma(x)}\right)$ is equivalent to the first-order regular variation condition, that is

$$\lim_{t\to\infty} \frac{q_{1-(tu)^{-1}}(\mathbf{x})}{q_{1-t^{-1}}(\mathbf{x})} = u^{\gamma(x)} \quad \text{for all } \; u > 0. \tag{4.9}$$

To derive the convergence rates of extreme conditional extremiles, we further need the second-order condition indexed by $(\gamma(\mathbf{x}), \varrho(\mathbf{x}), A(\cdot|\,\mathbf{x}))$, that is, there exists $\gamma(\mathbf{x}) \in (0,1)$, $\varrho(\mathbf{x}) \le 0$ and auxiliary function $A(\cdot|\,\mathbf{x})$ satisfying $\lim_{t\to\infty} A(t|\,\mathbf{x}) = 0$ such that

$$\lim_{t\to\infty} \frac{1}{A(t|\,\mathbf{x})} \left\{ \frac{q_{1-(tu)^{-1}}(\mathbf{x})}{q_{1-t^{-1}}(\mathbf{x})} - u^{\gamma(\mathbf{x})} \right\} = u^{\gamma(\mathbf{x})} \frac{u^{\varrho(\mathbf{x})} - 1}{\varrho(\mathbf{x})} \quad \text{for all } \; u > 0. \tag{4.10}$$

Moreover, we assume that $A(\cdot|\,\mathbf{x})$ is a regularly varying function with index $\varrho(\mathbf{x})$, that is, $\lim_{t\to\infty} A(tz|\,\mathbf{x})/A(t|\,\mathbf{x}) = z^{\varrho(\mathbf{x})}$ (De Haan and Ferreira, 2006). We can substitute $log(z)$ in place of $\frac{z^{\varrho(\mathbf{x})}-1}{\varrho(\mathbf{x})}$ when $\varrho(\mathbf{x}) = 0$. This second-order condition controls the convergence rate of $\frac{q_{1-(tu)^{-1}}(\mathbf{x})}{q_{1-t^{-1}}(\mathbf{x})}$. Many commonly used continuous distributions fit the second-order condition. For example, it holds for the normal distribution with $\gamma = \varrho = 0$. If $\gamma > 0$ and $\varrho < 0$, then (4.10) is also equivalent to

$$q_{1-t^{-1}}(\mathbf{x}) = ct^{\gamma(\mathbf{x})}\left[1 + \frac{A(t|\,\mathbf{x})}{\varrho(\mathbf{x})}\{1+o(1)\}\right], \quad \text{as} \quad t \to \infty.$$

We refer the reader to De Haan and Ferreira (2006) for more details of extreme value theory, including numerous interpretations and examples.

Within the domain of attraction $\mathrm{DA}\left(\Phi_{\gamma(\mathbf{x})}\right)$, this asymptotic relation shows that extremiles $\xi_\tau(\mathbf{x})$ are more sensitive to the heaviness of the right tail than the corresponding quantiles $q_\tau(\mathbf{x})$ as $\tau \to 1$. Here the extreme value index $\gamma(\mathbf{x})$ governs the tail thickness of $F_{Y|\mathbf{x}}$, with larger positive values indicating heavier tails. Therefore, it is crucial to develop estimation methods that can reliably quantify $\xi_\tau(\mathbf{x})$ for τ close to 1, in order to obtain accurate assessments of tail risk.

4.2 Estimation Procedure

In this section, we extrapolate the intermediate extremile estimator to an extreme level τ_n', which approaches one at a rate in the sense that $n(1-\tau_n') \to c$, for some constant $c > 0$.

We first consider the estimation of extreme conditional quantiles. To extrapolate beyond the range of observed data in the extreme tail, a classical and efficient tool is the Weissman estimator (Weissman, 1978), which exploits the asymptotic behavior of extremes. Define a sequence of intermediate levels $\tau_{n-k} < \tau_{n-k+1} < \ldots < \tau_{n-1} \in (0,1)$ with $\tau_{n-j} = (n-j)/(n+1)$ for $j = 1, \ldots, k$ and $\tau_{n-j} \to 1$ as $n \to \infty$. For the extreme level τ_n', we define the Weissman-type estimator of the conditional quantile $q_{\tau_n'}(\mathbf{x})$ by

$$\widehat{q}_{\tau_n'}(\mathbf{x}) = \left(\frac{1-\tau_{n-k}}{1-\tau_n'}\right)^{\widehat{\gamma}(\mathbf{x})} \widehat{q}_{\tau_{n-k}}(\mathbf{x}), \tag{4.11}$$

where $\widehat{q}_{\tau_n'}(\mathbf{x})$ and $\widehat{\gamma}(\mathbf{x})$ are the estimators of $q_{\tau_n'}(\mathbf{x})$ and $\gamma(\mathbf{x})$, respectively. In this construction, the intermediate level is typically taken as $\tau_{n-k} = (n-k)/(n+1)$, where $k = k(n)$ is an integer sequence satisfying $k \to \infty$ and $k/n \to 0$ as $n \to \infty$.

Under the Assumption $F_{Y|x} \in DA\left(\Phi_{\gamma(x)}\right)$ or equivalently the first-order regular variation condition (4.9), we can derive from (4.8) that

$$\frac{\xi_{\tau_n'}(\mathbf{x})}{q_{\tau_n'}(\mathbf{x})} \sim G(\gamma(\mathbf{x})) \sim \frac{\xi_{\tau_{n-k}}(\mathbf{x})}{q_{\tau_{n-k}}(\mathbf{x})} \quad \text{as } n \to \infty,$$

and hence

$$\frac{\xi_{\tau_n'}(\mathbf{x})}{\xi_{\tau_{n-k}}(\mathbf{x})} \sim \frac{q_{\tau_n'}(\mathbf{x})}{q_{\tau_{n-k}}(\mathbf{x})} \quad \text{as } n \to \infty. \tag{4.12}$$

Combining (4.11) and (4.12) motivates the following purely extremile-based estimator:

$$\widehat{\xi}_{\tau_n'}(\mathbf{x}) = \left(\frac{1-\tau_{n-k}}{1-\tau_n'}\right)^{\widehat{\gamma}(\mathbf{x})} \widehat{q}_{\tau_{n-k}}(\mathbf{x}) G(\hat{\gamma}(\mathbf{x})) = \left(\frac{1-\tau_{n-k}}{1-\tau_n'}\right)^{\widehat{\gamma}(\mathbf{x})} \widehat{\xi}_{\tau_{n-k}}(\mathbf{x}), \tag{4.13}$$

where $\widehat{\xi}_{\tau_{n-k}}(\mathbf{x})$ is estimated by the method of Section 3. Thus, the final extremile-based estimator at the extreme level can be viewed as the extrapolation of the intermediate one. Furthermore, for the tail index $\gamma(x)$, we consider Hill's estimator (Hill, 1975), which is a pivotal tool for tail index estimation in extreme value theory, that is

$$\widehat{\gamma}(\mathbf{x}) = \frac{1}{k-1}\sum_{j=1}^{k}\log\frac{\widehat{q}_{\tau_{n-j}}(\mathbf{x})}{\widehat{q}_{\tau_{n-k}}(\mathbf{x})} = \frac{1}{k-1}\sum_{j=1}^{k}\log\frac{\widehat{\xi}_{\tau_{n-j}}(\mathbf{x})}{\widehat{\xi}_{\tau_{n-k}}(\mathbf{x})}. \tag{4.14}$$

Hill estimator achieves optimal convergence rates under mild regularity conditions (Hill, 1975).

Remark 2. In the estimation procedure, k is the number of intermediate level extremiles used for the tail extrapolation. Actually, the choice of k depends on the sample size n, that is, $k = k(n)$ satisfying $k \to \infty$ and $k/n \to 0$ as $n \to \infty$. In the extreme value literature (Hill, 1975; Wang et al., 2012), the selection of k is very important. The value k can be viewed as the effective sample size for tail extrapolation. A smaller k leads to estimators with larger variance, while a larger k results in more bias. In practice, a common approach to select k is to plot the estimation error versus k, then choose a suitable k corresponding to the smallest error. In our simulation, we employ this approach for selecting k, see Figure 1 in Section 7.

5. Main Results

In this section, we present the theoretical results for the proposed estimator. We begin by introducing an integral operator that plays a central role in our analysis. Let $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a Mercer kernel. It induces a linear operator L_K on $L^2_{\rho_{\mathbf{x}}}$ defined by

$$(L_K f)(\mathbf{x}) = \int_{\mathcal{X}} K(\mathbf{x}, \mathbf{u})\, f(\mathbf{u})\, d\rho_{\mathbf{x}}(\mathbf{u}), \qquad f \in L^2_{\rho_{\mathbf{x}}},$$

where $L^2_{\rho_{\mathbf{x}}} = \{f : \int f^2(\mathbf{x}) d\rho_{\mathbf{x}} < \infty\}$ and $\rho_{\mathbf{x}}$ denotes the marginal distribution of $\rho_{\mathbf{x},y}$. The following technical assumptions are needed to establish the consistency and convergence rates for the proposed estimator.

Assumption 1. There exists some constant $\kappa > 0$ such that $\sup_{\mathbf{x}\in\mathcal{X}} \|K_{\mathbf{x}}\|_K \leq \kappa$.

Assumption 2. For $r \in [1/2, 1]$, we assume $\xi_{\tau_n} = L_K^r h_{\tau_n}$ for some $h_{\tau_n} \in L^2_{\rho_{\mathbf{x}}}$, where L_K^r denotes the r-th power of L_K.

Assumption 3. The effective sample size $k = k(n) \to \infty$ is an integer sequence, and as $n \longrightarrow \infty$, $k/n \to 0$, $\sqrt{k}A\left(n/k\right) \to \phi \in \mathbb{R}$, where $A(\cdot)$ is a regularly varying function.

Assumption 1 implies that we can consider those bounded kernels, a conventional choice extensively discussed in RKHS studies. Popular employed kernels include Gaussian kernel and Sobolev kernel (Smale and Zhou, 2007; Yang et al., 2016). Assumption 2 imposes a regularity condition on ξ_{τ_n} and controls the RKHS approximation error of the estimator. It is a classical source condition that has been extensively used in approximation theory (Smale and Zhou, 2007; Lin et al., 2017; He et al., 2021). The case $r = 1/2$ corresponds to the least favorable situation and is essentially equivalent to making no additional smoothness assumption beyond belonging to the RKHS. Assumption 3 specifies the increasing rate of k with respect to the sample size n. $k/n \to 0$ is a condition on the effective sample size in extreme value theory (EVT). The regularly varying function $A(\cdot)$ is introduced in Section 4.1, which is also frequently used in EVT. More specific choices of k can be referred to Theorem 2 and our simulation analysis.

Under the aforementioned assumptions and conditions (4.9) and (4.10) introduced in Section 4.1, we are now ready to give the convergence rates of the proposed estimators. In Proposition 1, we first establish the consistency of the quantile regression process estimator.

Proposition 1. *Suppose Assumption 1 is satisfied, and*

$$\frac{s_n^{1/2}\left(\log n\right)^{1/2}\left(1+\kappa\lambda_{2n}^{-\frac{1}{2}}\right)^{1/2}}{n^{1/4}} \to 0, \quad \textit{as } n, s_n \to \infty,$$

then we obtain that

$$\sup_{\tau\in\left[\frac{1}{s_n+1},\frac{s_n}{s_n+1}\right]}\|\widehat{q}_\tau - q^*_\tau\|_K = o_p(1),$$

where $q^*_\tau(\cdot)$ *denotes the true* τ*-th quantile function.*

This proposition gives the uniform consistency of the estimated quantile process on the set sequence approaching the whole quantile interval $(0,1)$, which implies that the estimate $\hat{F}_{\mathbf{x}}(\cdot)$ is also consistent.

In this paper, we focus on the right heavy tail of variables. We begin by providing the estimation result at the intermediate level following $n\to\infty,\ n(1-\tau_n)\to\infty$.

Theorem 1. *Suppose Assumptions 1–2 and the conditions in Proposition 1 are satisfied, then there exists some positive constant* C *such that, for any* $\delta_n\in(0,1)$*, with probability at least* $1-\delta_n$*, there holds*

$$\|\widehat{\xi}_{\tau_n}-\xi_{\tau_n}\|_K \le \log\frac{4}{\delta_n}\left(\frac{6\kappa r_{\tau_n} M_n^{\xi}}{\lambda_{1n} n^{1/2}} + \frac{C\kappa r_{\tau_n}^2 M_n^{\xi}}{\lambda_{1n}\alpha_n}(1+\kappa\lambda_{1n}^{-1/2}) + \lambda_{1n}^{2r-1}\left\|L_K^{-r}\xi_{\tau_n}\right\|_2\right),$$

where $M_n^{\xi}=\max\left\{\kappa\|f_0\|_K+\sqrt{\frac{2\sum_{i=1}^n\sigma_i^2}{\delta_n}}, \|\xi_{\tau_n}\|_K\right\}$ *and* $\alpha_n>0$*.*

Here α_n is determined by the regularization parameter λ_{2n} and the number of quantile levels s_n, and it represents the convergence rate of the estimated conditional distribution function $\widehat{F}_{\mathbf{x}}(\cdot)$, i.e., $\sup_{y\in\mathbb{R}}|\widehat{F}_{\mathbf{x}}(y)-F_{\mathbf{x}}(y)| = O_p(\alpha_n)$. The explicit form of α_n is given in the proof provided in the Supplementary Material. The error bound in Theorem 1 consists of two main components: $\log\frac{4}{\delta_n}\left(\frac{6\kappa r_{\tau_n}M_n^{\xi}}{\lambda_{1n}n^{1/2}}+\frac{C\kappa r_{\tau_n}^2 M_n^{\xi}}{\lambda_{1n}\alpha_n}(1+\kappa\lambda_{1n}^{-1/2})\right)$ and $\log\frac{4}{\delta_n}\lambda_{1n}^{2r-1}\left\|L_K^{-r}\xi_{\tau_n}\right\|_2$. The first term can be viewed as the empirical error, as quantified in Lemma 2 of the Supplementary Material, while the second term corresponds to the approximation bias induced by the RKHS through the source condition on ξ_{τ_n}. This non-asymptotic control at the intermediate level is a key ingredient for deriving error bound at the extreme level. The next theorem provides the convergence rate of the estimated conditional extremiles at the extreme level τ_n'.

Theorem 2. *Suppose Assumptions 1-3, conditions in Proposition 1 and conditions* (4.9) *and* (4.10) *are satisfied, denote* $p_n = 1-\tau_n'$, $\tau_n' \to 1$, $np_n = o(k)$ *and* $\log(np_n) = o(\sqrt{k})$. *Then there exists some positive constant* C *such that, for any* $\delta_n \in (0,1)$, *with probability at least* $1-\delta_n$, *there holds*

$$\frac{\sqrt{k}}{\log(k/np_n)}\left(\left\|\frac{\widehat{\xi}_{\tau_n'}}{\xi_{\tau_n'}}\right\|_K - 1\right) = \frac{\phi}{1-\varrho} + \frac{\sqrt{k}B_{n,k}}{C},$$

where $B_{n,k} = \log\frac{4}{\delta_n}\left(\frac{6\kappa r_{\tau_{n-k}} M^{\xi}_{n,k}}{\lambda_{1n} n^{1/2}} + \frac{C\kappa r^2_{\tau_{n-k}} M^{\xi}_{n,k}}{\lambda_{1n}\alpha_n}(1+\kappa\lambda_{1n}^{-1/2}) + \lambda_{1n}^{2r-1}\left\|L_K^{-r}\xi_{\tau_{n-k}}\right\|_2\right)$ *with* $M^{\xi}_{n,k} = \max\left\{\kappa\left\|f_0\right\|_K + \sqrt{\frac{2\sum_{i=1}^n \sigma_i^2}{\delta_n}}, \|\xi_{\tau_{n-k}}\|_K\right\}$.

Theorem 2 provides an upper bound on the convergence rates at the extreme levels τ_n'. The conditions imposed on τ_n' are analogous to those used in Wang et al. (2012) and Wang and Li (2013). As discussed in Remark 2, the number of intermediate extremile levels k plays a key role in balancing bias and variance. Note that the upper bound involves two terms, $\frac{\log\{k/(np_n)\}}{\sqrt{k}}\frac{\phi}{1-\varrho}$ and $\frac{\log\{k/(np_n)\}B_{n,k}}{C}$. The first term corresponds to the variance of the estimator which shrinks as k increases, and the second term represents the bias of the estimator which grows proportionally with $\log k$. The influence of k on the finite-sample performance of the estimator is further investigated in the simulation study in Section 7.

Remark 3. Our theoretical analysis is developed under a fixed covariate dimension p, in line with the general theory of kernel methods (Smale and Zhou, 2007; He et al., 2021), where convergence rates are not explicitly in terms of p. In particular, our results cover settings in which the conditional extremiles $\xi_\tau(\cdot)$ lie in a relatively low-complexity subspace of $\mathcal{H}_K$, even if the ambient dimension is moderate or large. A full $p \gg n$ theory is beyond the present scope, and we leave this as a future direction.

6. Comparisons and Discussion

6.1 Advantages of Extremiles Over Traditional Risk Measures

As emphasized in Daouia et al. (2019), extremiles address critical limitations of Value at Risk (VaR) and expectiles by integrating tail sensitivity, coherence, statistical estimation efficiency, and intuitive interpretability, which are tailored for extreme risk analysis. VaR is defined as the τ-th quantile q_τ of a non-negative loss distribution with τ close to 1 or $-q_\tau$ of a real-valued profit-loss distribution with τ close to 0. VaR only depends on the frequency of losses exceeding a threshold, not on their values. Thus, it is insensitive to the magnitude of those excess losses. Moreover, VaR is in general not coherent because it fails subadditivity. It can lead to inconsistent capital requirements for portfolios. By contrast, the extremile at level τ is defined as a tail-weighted least-squares functional, which depends on both the probability of tail events and the magnitude of tail losses. Extremile-based risk measure is coherent, satisfying subadditivity, monotonicity, and law invariance. More importantly, they also meet the co-monotonic additivity property for co-monotonic losses, as introduced in Section 4.1 in Daouia et al. (2019).

Expectiles are also based on asymmetric least squares and thus share with extremiles the dependence on both tail probabilities and magnitudes. Expectiles are coherent in most respects but fail comonotonic additivity (Acerbi and Szekely, 2014), which is a critical property for aggregating correlated risks. In addition, they typically do not admit simple closed-form expressions and can be harder to interpret in terms of extreme events. However, extremiles enjoy several explicit representations, i.e., $\xi_\tau = \int_0^1 J_\tau(t) q_t dt$. At very high levels, extremiles have a direct asymptotic link to extreme quantiles and the tail index, which we used in our theoretical development. This makes extremiles particularly suitable for modeling extreme conditional losses, while still retaining a least-squares structure that is suitable for RKHS methods.

6.2 Connection between Extremiles and $K_\tau(t)$

In fact, the distribution transformation $K_\tau(t) : [0,1] \to [0,1]$ is a piecewise distortion function, given by:

$$K_\tau(t) = \begin{cases} 1-(1-t)^{s(\tau)} & \text{if } 0 < \tau \leq 1/2; \\ t^{r(\tau)} & \text{if } 1/2 \leq \tau < 1, \end{cases}$$

where $r(\tau) = s(1-\tau) = \log(1/2)/\log(\tau)$. The key insight is that $K_\tau(t)$ acts as a distributional distortion that connects quantiles and extremiles of a functional of the random variable Z_τ. Specifically, the τ-th quantile q_τ of Y is equivalent to the median of Z_τ, where Z_τ has CDF $F_{Z_\tau} = K_\tau(F)$. Here we provide a brief proof. First, it is easy to verify that $K_\tau(\tau) = 1/2$ whenever $\tau \in (0, 1/2]$ or $\tau \in [1/2, 1)$, according to the definition of $K_\tau(\cdot)$. Since $K_\tau(\cdot)$ is increasing and $F(\cdot)$ are non-decreasing, when $y < q_\tau$, we have $F_{Z_\tau}(y) = K_\tau(F(y)) < K_\tau(\tau) = \frac{1}{2}$; when $y \geq q_\tau$, we have $F_{Z_\tau}(y) = K_\tau(F(y)) \geq K_\tau(\tau) = \frac{1}{2}$, where we use the definition of τ-th quantile q_τ. Thus, we can conclude that $q_\tau = \text{median}(Z_\tau)$, or

$$q_\tau \in \text{argmin}_\theta \mathbb{E}_{Z_\tau}|Z_\tau - \theta|.$$

Note that $J_\tau(\cdot) = K'_\tau(\cdot) \geq 0$, we have

$$\begin{aligned} \mathbb{E}_{Z_\tau}|Z_\tau - \theta| &= \int |y-\theta| dF_{Z_\tau}(y) = \int |y-\theta| dK_\tau(F(y)) \\ &= \int J_\tau(F(y))|y-\theta| dF(y) = \mathbb{E}_y[J_\tau(F(y))|y-\theta|]. \end{aligned}$$

Thus, we can define the τ-th quantile q_τ as the minimizer of the following:

$$q_\tau = \text{argmin}_\theta \, \mathbb{E}\left\{J_\tau(F(y)) \cdot [|y-\theta| - |y|]\right\}.$$

The τ-th extremile of the response y is the parallel counterpart of the τ-th quantile, which is given by replacing the absolute deviation with the squared deviation

$$\xi_\tau = \text{argmin}_\theta \, \mathbb{E}\left\{J_\tau(F(y)) \cdot \left[|y-\theta|^2 - |y|^2\right]\right\}.$$

And we can similarly prove that the τ-th extremile ξ_τ of Y is equivalent to the mean of Z_τ.

The distortion function $K_\tau(t)$ used in this paper is not arbitrary. In this paper, we adopt the distortion function K_τ of Daouia et al. (2019), so that our target functional coincides with their quantile and extremile. In particular, with this specific $K_\tau(t)$, the quantile q_τ and the extremile ξ_τ are respectively the median and the mean of the same distribution $K_\tau(F)$. This yields explicit integral formulas expressing ξ_τ as a probability-weighted moment. These representations are central to the theory of extremiles and to their extreme-value asymptotics. However, the framework can be generalized to other distortion functions that generate valid weighting functions, and could be used to study other tail-weighted least-squares functionals beyond classical extremiles.

7. Numerical Experiments

7.1 Simulation Examples

This section investigates the finite-sample performance of the proposed method and compares it with ordinary extremile regression without extrapolation to extreme levels. We consider the following nonlinear regression model:

$$y_i = \sum_{j=1}^{p} \beta_i \sin(2\pi x_{ij}) + (1 + r\bar{x}_i)\varepsilon_i, \quad i = 1, 2, \ldots, n,$$

where $\beta_i \sim \text{Unif}(0, 0.1)$, $x_{ij} \sim \text{Unif}(0, 1)$ with $\bar{x}_i = \frac{1}{p}\sum_{j=1}^{p} x_{ij}$. Here, p is the covariate dimension, and $r \in \{0, 1\}$ controls the error structure: $r = 0$ corresponds to the homoscedastic case, whereas $r = 1$ generates heteroscedastic errors. The noise variables ε_i are i.i.d. with heavy-tailed distributions and are independent of the covariates $\{x_{ij}\}_{j=1}^{p}$. We consider the following error distributions:

1. Pareto distribution with extreme value index parameter $\gamma = 1/3$;

2. Fréchet distribution with extreme value index parameter $\gamma = 1/4$;

3. Student's t distribution with degree 5.

For a t distribution with ψ degrees of freedom, the extreme value index is $\gamma = 1/\psi$, so in Example 3 we have $\gamma = 1/5$.

According to (2.2), the τ-th conditional extremile of the response can be written as

$$\xi_\tau(\mathbf{x}_i) = \mathbf{f}(\mathbf{x}_i) + (1 + r\bar{x}_i)\int_0^1 J_\tau(t) q_\varepsilon(t) dt, \quad i = 1, \ldots, n,$$

where $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$, $\mathbf{f}(\mathbf{x}_i) = \sum_{j=1}^{p} \beta_i \sin(2\pi x_{ij})$ and $q_\varepsilon(t)$ is the t-th quantile of ε_i. To evaluate the performance of the proposed estimator, we use two summary measures. The first

Table 1: The mean absolute errors and root mean squared errors of those three estimators in the case with the error distribution Pareto(1/3), where $\tau_n' = 0.99, 0.995$ and 0.999.

	Sample size	Method	MAE			RMSE		
			0.99	0.995	0.999	0.99	0.995	0.999
$r = 0$	$n = 200$	KEE-QR	4.69	6.75	10.78	4.83	6.87	11.46
		OKE-QR	5.60	8.89	17.98	5.64	8.91	18.01
		KEE-NW	5.81	7.50	15.72	5.93	7.61	15.94
	$n = 500$	KEE-QR	3.13	3.22	10.11	3.66	3.82	10.67
		OKE-QR	4.45	5.53	18.05	4.69	5.61	18.07
		KEE-NW	3.59	4.21	11.49	4.60	5.71	12.3
	$n = 1000$	KEE-QR	2.65	2.88	6.31	2.75	2.96	6.54
		OKE-QR	3.32	4.14	12.65	3.51	4.21	12.21
		KEE-NW	3.01	3.78	10.21	4.38	3.92	11.59
$r = 1$	$n = 200$	KEE-QR	3.36	4.84	8.53	4.26	5.22	9.25
		OKE-QR	5.18	7.59	17.94	5.31	7.64	17.97
		KEE-NW	12.62	8.91	13.15	17.54	9.03	13.34
	$n = 500$	KEE-QR	3.07	3.19	6.99	3.31	3.70	8.35
		OKE-QR	4.86	6.88	18.02	4.92	6.96	18.05
		KEE-NW	6.1	7.41	10.94	8.38	7.94	12.35
	$n = 1000$	KEE-QR	2.51	2.75	2.45	2.87	3.15	2.90
		OKE-QR	3.25	4.76	8.36	3.49	4.84	8.37
		KEE-NW	4.76	5.73	10.59	6.09	7.95	15.96

is the mean absolute error (MAE) defined as

$$\mathrm{MAE} = n^{-1} \sum_{i=1}^{n} \left| \widehat{\xi}_{\tau_n}(\mathbf{x}_i) - \xi_{\tau_n}(\mathbf{x}_i) \right|,$$

and the second measurement is the square root of mean squared error (RMSE)

$$\mathrm{RMSE} = \left(n^{-1} \sum_{i=1}^{n} \left\{ \widehat{\xi}_{\tau_n}(\mathbf{x}_i) - \xi_{\tau_n}(\mathbf{x}_i) \right\}^2 \right)^{1/2},$$

where $\widehat{\xi}_{\tau_n}$ and $\xi_{\tau_n}(x_i)$ are the estimated and true conditional extremiles, respectively.

Table 2: The mean absolute errors and root mean squared errors of those three estimators in the case with the error distribution Fréchet(1/4), where $\tau_n' = 0.99, 0.995$ and 0.999.

	Sample size	Method	MAE			RMSE		
			0.99	0.995	0.999	0.99	0.995	0.999
$r = 0$	$n = 200$	KEE-QR	2.59	3.82	5.45	2.72	3.89	5.88
		OKE-QR	3.46	5.49	9.43	3.50	5.49	9.44
		KEE-NW	3.16	4.35	7.09	3.28	4.42	7.29
	$n = 500$	KEE-QR	2.28	3.42	4.61	2.44	3.61	4.98
		OKE-QR	3.16	5.02	9.41	3.21	5.04	9.43
		KEE-NW	3.25	4.30	5.12	3.36	4.44	5.37
	$n = 1000$	KEE-QR	2.01	2.94	3.82	2.45	3.01	3.95
		OKE-QR	2.65	3.71	7.21	2.73	3.76	7.23
		KEE-NW	2.32	3.54	4.92	2.48	3.71	5.01
$r = 1$	$n = 200$	KEE-QR	1.52	2.17	4.41	1.76	2.51	4.77
		OKE-QR	2.38	4.19	9.46	2.5	4.24	9.47
		KEE-NW	2.41	4.73	6.45	2.62	4.87	6.59
	$n = 500$	KEE-QR	1.39	1.48	2.71	1.67	1.75	3.22
		OKE-QR	2.08	3.77	9.45	2.29	3.81	9.46
		KEE-NW	2.42	2.72	3.84	2.77	3.08	4.39
	$n = 1000$	KEE-QR	1.12	1.24	2.15	1.15	1.26	2.17
		OKE-QR	1.85	3.26	7.43	1.91	3.37	7.59
		KEE-NW	1.93	2.33	3.12	1.95	2.53	3.20

We consider extremiles computed at the risk levels $\tau_n' = 0.99, 0.995$, and 0.999, with covariate dimension $p = 10$. We use the standard Gaussian kernel

$$K(\mathbf{u}, \mathbf{v}) = \exp\left(-\frac{\| \mathbf{u} - \mathbf{v} \|_2^2}{2}\right)$$

for our estimation, which is widely used in the literature of RKHS (Smale and Zhou, 2007; He et al., 2021). Under each model setting, the experiments are repeated $B = 100$ times for sample sizes $n = \{200, 500, 1000\}$. In accordance with the conditions of Theorem 2, the number effective sample size used for tail extrapolation is set as $k = [cn^{1/2}]$ with $c = 3$ for a stable default choice.

We also investigate alternative values of k obtained by varying the constant c, i.e., see Figure 1 for a sensitivity analysis. Moreover, we choose the regularization parameter λ_{1n} by minimizing the MAE on an independently generated dataset in each scenario. The searching grid is set in $\{10^{-7}, 10^{-6}, \ldots, 10^{-1}, 10^{0}, 10\}$. To compare with our proposed method, named kernel extremile estimator (KEE-QR), we implement two alternatives:

(i) the ordinary kernel method (OKE-QR), which is derived based on (3.4) without any extrapolations as the benchmark to show the necessity of extrapolation when estimating at extreme levels;

(ii) the conditional extremile estimator proposed by Daouia et al. (2022) denoted as KEE-NW.

Tables 1–3 report the MAE and RMSE of the different estimators of $\xi_{\tau_n'}$ across all experimental scenarios.

The results in Tables 1–3 show that all three methods improve in terms of MAE and RMSE as the sample size increases. Leveraging the quantile process approximation in an RKHS, and thus avoiding the purely local Nadaraya–Watson smoothing when p is moderate, KEE-QR systematically outperforms KEE-NW in our settings. Moreover, KEE-QR also dominates OKE-QR in all cases at extreme risk levels, especially at $\tau_n' = 0.999$, confirming that direct extremile regression without extrapolation becomes unstable in the tail region due to data sparsity. As n grows, the RMSE of KEE-QR at $\tau_n' = 0.999$ decreases markedly, indicating that the proposed procedure can successfully extrapolate beyond the range of observed data. As expected, both MAE and RMSE increase with the risk level τ_n' for all three error distributions. Regarding the heteroscedasticity parameter r, the performance of KEE-QR is robust in both the homoscedastic ($r = 0$) and heteroscedastic ($r = 1$) settings.

Table 3: The mean absolute errors and root mean squared errors of those three estimators in the case with the error distribution t(5), where $\tau_n' = 0.99, 0.995$ and 0.999.

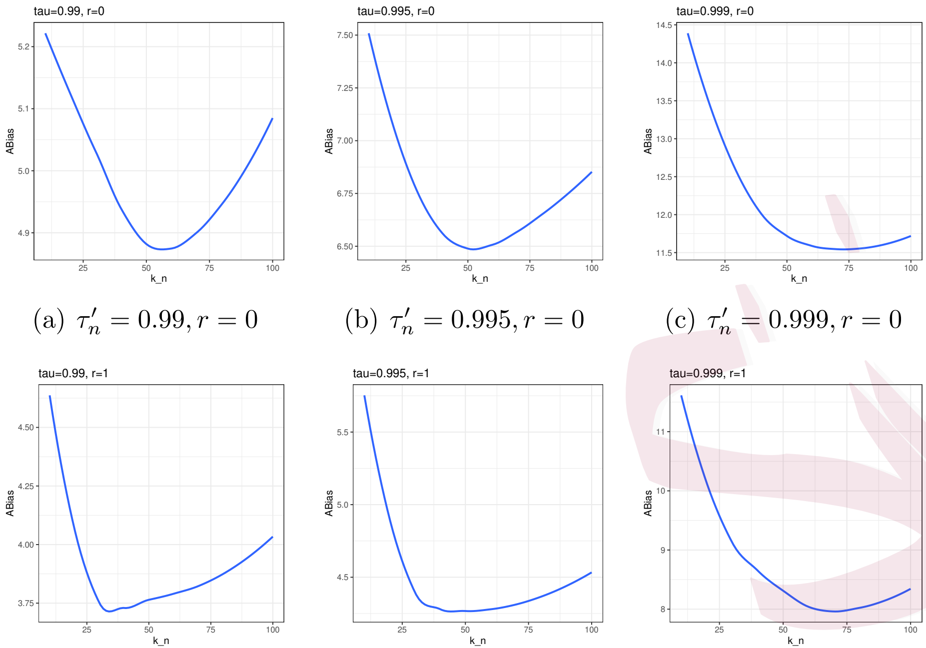
	Sample size	Method	MAE			RMSE		
			0.99	0.995	0.999	0.99	0.995	0.999
$r = 0$	$n = 200$	KEE-QR	2.81	3.25	3.95	2.92	3.36	5.11
		OKE-QR	5.02	6.43	9.56	5.03	6.44	9.68
		KEE-NW	4.13	5.81	7.12	4.31	5.89	7.33
	$n = 500$	KEE-QR	1.73	2.31	3.72	1.91	2.54	4.05
		OKE-QR	4.29	6.27	8.34	4.31	6.28	8.51
		KEE-NW	2.62	4.11	5.69	2.93	4.54	6.83
	$n = 1000$	KEE-QR	1.32	2.01	3.31	1.41	2.43	3.54
		OKE-QR	3.67	5.92	7.92	3.78	6.02	8.01
		KEE-NW	2.34	3.88	5.12	2.76	3.94	5.32
$r = 1$	$n = 200$	KEE-QR	1.54	2.63	4.29	1.95	3.62	4.68
		OKE-QR	3.72	6.15	9.65	3.75	6.16	9.66
		KEE-NW	4.25	5.06	8.23	4.41	5.28	8.48
	$n = 500$	KEE-QR	1.06	1.79	3.62	1.3	2.06	3.97
		OKE-QR	3.51	5.32	9.56	3.54	5.57	9.57
		KEE-NW	2.93	4.78	7.48	3.39	5.12	7.80
	$n = 1000$	KEE-QR	0.84	1.43	2.54	1.02	1.45	2.90
		OKE-QR	2.95	4.78	8.21	3.14	4.98	8.64
		KEE-NW	2.23	3.56	5.87	3.56	3.87	6.31

Figure 1 displays the MAE of KEE-QR as a function of the number of intermediate levels k for the Pareto(1/3) errors with $n = 200$. We observe that the MAE remains small for $k \in [40, 70]$, a range that includes our default choice $k = \lfloor 3n^{1/2} \rfloor$. The empirically optimal k increases with the risk level, suggesting that a larger number of intermediate levels should be used for tail extrapolation when estimating more extreme extremiles.

7.2 Real Data Application: Large Commercial Banks Data

This section focuses on the analysis of the large commercial banks dataset, which has been previously studied by Wang et al. (2014) and Xu et al. (2022a). The data contain weekly

7.2 Real Data Application: Large Commercial Banks Data



(a) $\tau_n' = 0.99, r = 0$ (b) $\tau_n' = 0.995, r = 0$ (c) $\tau_n' = 0.999, r = 0$

(d) $\tau_n' = 0.99, r = 1$ (e) $\tau_n' = 0.995, r = 1$ (f) $\tau_n' = 0.999, r = 1$

Figure 1: The mean absolute errors against the numbers of intermediate levels k (Pareto(1/3) errors and sample size $n = 200$).

observations for four major banks: Bank of America (BAC), Citigroup (Citi), JPMorgan Chase (JPM), and Wells Fargo (WFC). The sample sizes for BAC, Citi, JPM, and WFC are 1771, 1386, 2210, and 2210, respectively. The sample period runs from 1971 to 2013 and covers six recessions (1974–1975, 1980, 1981, 1990–1991, 2001, and 2007–2009) as well as several notable financial crises (1987, 1994, 1997, 1998, 2000, 2008, and 2011).

We take the negative weekly return as the response variable Y, and use the following covariates: weekly market return (x_1), three-month yield change (x_2), equity volatility (x_3), credit spread change (x_4), term spread change (x_5), short-term TED spread (x_6), and real estate excess return (x_7). The heavy-tailed nature of the loss distributions for these four banks have been documented in Xu et al. (2022a) using boxplots of weekly market losses, which justifies an extreme-value–based analysis. In our empirical study, we adopt the polynomial kernel

$$K(\mathbf{x},\mathbf{x}') = \big(\langle \mathbf{x},\mathbf{x}'\rangle + 1\big)^d$$

for the RKHS estimation with degree $d = 3$. Prior to the analysis, both the response and the covariates are standardized to make the results comparable across banks. The effective tail sample sizes k used for extrapolation are set to 70, 55, 40, and 45 for BAC, Citi, JPM, and WFC, respectively.

Our main objective is to examine how risk exposure varies across different risk management standards. To this end, we estimate the conditional extremiles at risk levels $\tau = 0.99, 0.995$, and 0.999. As illustrated in Figure 2, the estimated conditional extremiles increase monotonically with τ for each bank. Specifically, the larger the risk level, the higher extremile value obtained for a company. Among the four companies under investigation, Citi exhibits the largest conditional extremiles at all risk thresholds, indicating the highest tail risk. BAC and JPM display comparable risk estimates, with the similarity being particularly pronounced when the risk quantile τ is moderately low. In contrast, WFC tends to have lower extremile estimates, suggesting comparatively stronger risk control among the four banks considered. These findings align with the conclusions reported by Xu et al. (2022a) where they considered the conditional expectiles of these banks.

Furthermore, we estimate the weekly conditional extremiles for each of the four banks at the risk level $\tau = 0.995$. We then compare these estimates with (i) the corresponding conditional expectiles at the same risk level $\tau = 0.995$ and (ii) the actual weekly losses. For the conditional expectiles, we adopt a linear expectile regression model: intermediate conditional expectiles are first estimated and then extrapolated to the extreme level using Hill's estimator for the tail

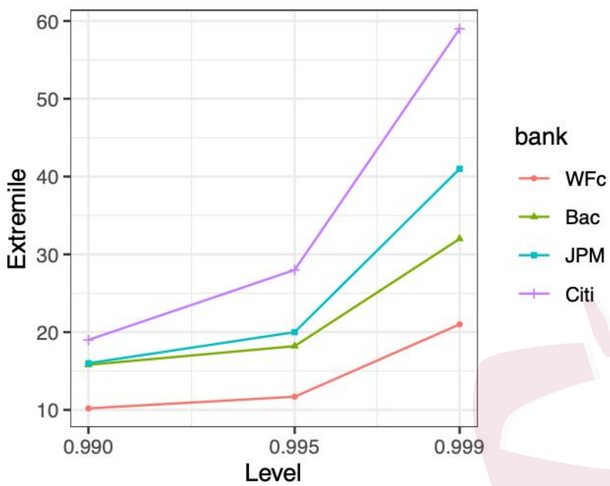


Figure 2: Estimated conditional extremiles across different risk levels.

index. The R package `expectreg` is employed to obtain the conditional expectile estimates. For the extremile regression in (3.4), the regularization parameter is set to 0.2.

Figure 3 reports the in-sample weekly conditional extremiles, conditional expectiles, and actual weekly losses for the four banks over the period 1990–2013, all computed at $\tau = 0.995$. Visually, the yellow curves (extremiles) typically envelope the green curves (actual losses) from above, indicating that extremiles provide a conservative risk benchmark capable of guarding against potential actual losses and thus serving as a useful instrument for risk management. In contrast, the expectile curves sometimes intersect or lie below the realized losses, whereas the extremile curves tend to yield systematically higher risk estimates, implying higher required capital or risk charges. The sample period includes six recessions (1974–1975, 1980, 1981, 1990–1991, 2001, and 2007–2009) and several major financial crises (1987, 1994, 1997, 1998, 2000, 2008, and 2011). Most of these stress episodes are clearly reflected in the in-sample extremile estimates for all four banks, as shown in Figure 3. Despite the differing profit and loss patterns across institutions, the extremile curves exhibit a pronounced time-varying association

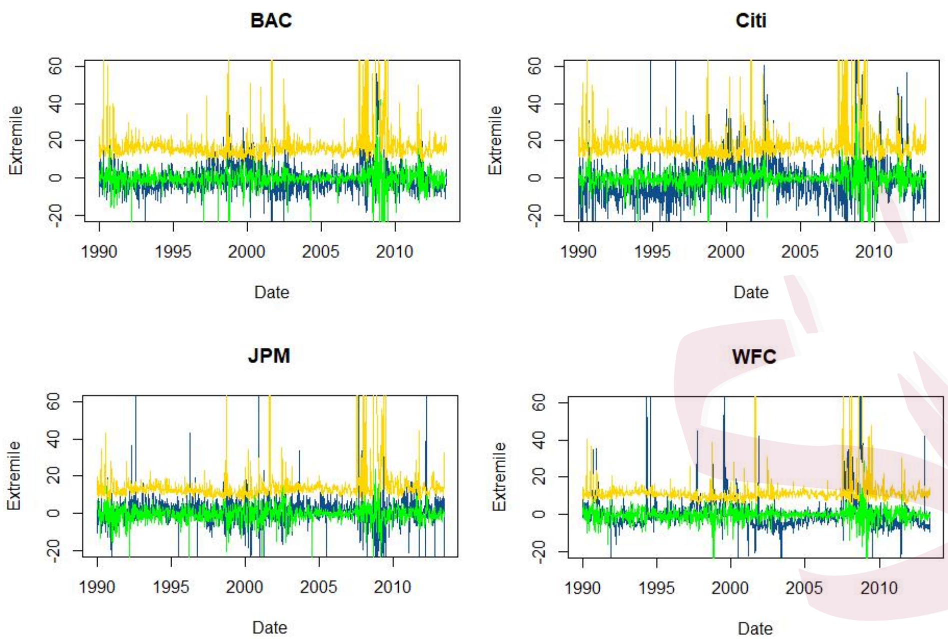


Figure 3: Time series of in-sample conditional extremiles (yellow), expectiles (blue), and actual weekly losses (green) for four large-scale financial institutions.

with weekly losses, responding strongly during turbulent periods.

Overall, these empirical results suggest that conditional extremiles offer a more conservative and informative view of tail risk than conventional measures such as expectiles, and therefore merit further investigation in financial risk management applications.

8. Conclusion

This paper focuses on the estimation of conditional extremiles and proposes a nonparametric estimation framework rooted in the quantile regression process within a reproducing kernel Hilbert space (RKHS). We derive an RKHS estimator for conditional extremiles, establish its asymptotic properties under suitable regularity and tail conditions, and illustrate its finite-sample behavior through simulation studies and a real data application.

Our work complements and extends the local smoothing approach of Daouia et al. (2022), which was the first comprehensive investigation of conditional extremile estimation. Their method relies on the minimization of a local linear check function and yields an explicit closed form for the estimator, which is computationally convenient in low dimensions. However, it is inherently exposed to the curse of dimensionality and can struggle in the presence of nonlinear and non-additive interactions among covariates that are common in financial and other applied contexts. In contrast, the RKHS framework can naturally accommodate nonlinearity, non-additivity, and complex interaction effects through the choice of the kernel, while global regularization provides a principled way to control model complexity. At the same time, our theoretical analysis is developed under a fixed covariate dimension and does not yet constitute a full $p \gg n$ theory. In high-dimensional applications, further gains can be expected by incorporating sparsity and variable selection into the RKHS extremile estimator, for example via additive or partially additive kernels, sparse feature representations, or gradient-based screening rules (He et al., 2021; Chen et al., 2021). Developing systematic high-dimensional theory and practical algorithms for such sparse RKHS extremile models is an important direction for future research.

Supplementary Materials

The supplementary materials contain some useful lemmas and the detailed proofs of the main results in this paper.

Acknowledgments

The authors thank the editor, the associate editor, and two anonymous referees for their constructive suggestions, which significantly improved this paper.

References

Acerbi, C. and Szekely, B. (2014). Back-testing expected shortfall. *Risk*, 27(11):76–81.

Adrian, T. and Fleming, M. (2022). The bond market selloff in historical perspective. Liberty street economics working paper, Federal Reserve Bank of New York.

Aon (2020). Economic losses from natural disasters top $232 billion in 2019 as the costliest decade on record comes to a close. Technical report, Aon Catastrophe Report.

Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

Berlinet, A. and Thomas-Agnan, C. (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. New York: Springer.

Blanchard, G., Bousquet, O., and Massart, P. (2008). Statistical performance of support vector machines. *Annals of Statistics*, 36(2):489–531.

Chen, F., He, X., and Wang, J. (2021). Learning sparse conditional distribution: An efficient kernel-based approach. *Electronic Journal of Statistics*, 15:1610–1635.

Daouia, A., Gijbels, I., and Stupfler, G. (2019). Extremiles: A new perspective on asymmetric least squares. *Journal of the American Statistical Association*, 114(527):1366–1381.

Daouia, A., Gijbels, I., and Stupfler, G. (2022). Extremile regression. *Journal of the American Statistical Association*, 117(539):1579–1586.

Daouia, A., Girard, S., and Stupfler, G. (2018). Estimation of tail risk based on extreme expectiles. *Journal of the Royal Statistical Society: Series B*, 80(2):263–292.

Daouia, A., Girard, S., and Stupfler, G. (2020). Tail expectile process and risk assessment. *Bernoulli*, 26(1):531–556.

De Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer.

Dupuis, D., Sun, Y., and Wang, H. (2015). Detecting change-points in temperature extremes. *Statistics and Its Interface*, 8:19–31.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. New York: Springer.

Friederichs, P. and Hense, A. (2007). Statistical downscaling of extreme precipitation events using censored quantile regression. *Monthly Weather Review*, 135(7):2388–2401.

Gardes, L. and Girard, S. (2010). Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return level. *Extremes*, 13:177–204.

Girard, S., Stupfler, G., and Usseglio-Carleve, A. (2022). Nonparametric extreme conditional expectile estimation. *Scandinavian Journal of Statistics*, 49(1):78–115.

He, X., Wang, J., and Lv, S. (2021). Efficient kernel-based variable selection with sparsistency. *Statistica Sinica*, 31(4):2123–2151.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3(5):1163–1174.

Horváth, L. and Yandell, B. S. (1988). Asymptotics of conditional empirical processes. *Journal of Multivariate Analysis*, 26(2):184–206.

Li, D. and Wang, H. J. (2019). Extreme quantile estimation for autoregressive models. *Journal of Business & Economic Statistics*, 37(4):661–670.

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

Lin, S.-B., Guo, X., and Zhou, D.-X. (2017). Distributed learning with regularized least squares. *Journal of Machine Learning Research*, 18(92):1–31.

Linsmeier, T. J. and Pearson, N. D. (2000). Value at risk. *Financial Analysts Journal*, 56:47–67.

Marimoutou, V., Raggad, B., and Trabelsi, A. (2009). Extreme value theory and value at risk: Application to oil market. *Energy Economics*, 31(4):519–530.

Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55(4):819–847.

Odening, M. and Hinrichs, J. (2003). Using extreme value theory to estimate value at risk. *Agricultural Finance Review*, 63:55–73.

Rahimi, A. and Recht, B. (2007). Random features for large-scale kernel machines. *Advances in Neural Information Processing Systems*, 20:1177–1184.

Resnick, S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer.

Rudi, A., Camoriano, R., and Rosasco, L. (2015). Less is more: Nyström computational regularization. *Advances in Neural Information Processing Systems*, 28:1657–1665.

Schaumburg, J. (2012). Predicting extreme value at risk: Nonparametric quantile regression with refinements from extreme value theory. *Computational Statistics and Data Analysis*, 56(12):4081–4096.

Shawe-Taylor, J. and Cristianini, N. (2004). *Kernel Methods for Pattern Analysis*. Cambridge University Press.

Smale, S. and Zhou, D. (2007). Learning theory estimates via integral operators and their approximations. *Constructive Approximation*, 26(2):153–172.

Steinwart, I. (2005). Consistency of support vector machines and other regularized kernel classifiers. *IEEE Transactions on Information Theory*, 51(1):128–142.

Steinwart, I. and Scovel, C. (2007). Fast rates for support vector machines using Gaussian kernels. *Annals of Statistics*, 35(2):575–607.

Takeuchi, I., Le, Q. V., Sears, T. D., and Smola, A. J. (2006). Nonparametric quantile estimation. *Journal of Machine Learning Research*, 7:1231–1264.

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

Wang, C. and Feng, X. (2024). Optimal kernel quantile learning with random features. In *International Conference on Machine Learning*, pages 50419–50452. PMLR.

Wang, C., Li, T., Zhang, X., Feng, X., and He, X. (2024). Communication-efficient nonparametric quantile regression via random features. *Journal of Computational and Graphical Statistics*, 33(4):1175–1184.

Wang, H. and Tsai, C. L. (2009). Tail index regression. *Journal of the American Statistical Association*, 104(485):1233–1240.

Wang, H. J. and Li, D. (2013). Estimation of extreme conditional quantiles through power transformation. *Journal of the American Statistical Association*, 108(503):1062–1074.

Wang, H. J., Li, D., and He, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association*, 107(498):1453–1464.

Wang, S., Shao, J., and Kim, J. K. (2014). An instrumental variable approach for identification and estimation with nonignorable nonresponse. *Statistica Sinica*, 24(3):1097–1116.

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73(364):812–815.

Xu, W., Hou, Y., and Li, D. (2022a). Prediction of extremal expectile based on regression models with heteroscedastic extremes. *Journal of Business & Economic Statistics*, 40:522–536.

Xu, W., Wang, H., and Li, D. (2022b). Extreme quantile estimation for single index model. *Statistica Sinica*, 32:893–914.

Yang, L., Lv, S., and Wang, J. (2016). Model-free variable selection in reproducing kernel Hilbert space. *Journal of Machine Learning Research*, 17(1):2885–2908.

Yang, Y., Zhang, T., and Zou, H. (2018). Flexible expectile regression in reproducing kernel Hilbert spaces. *Technometrics*, 60(1):26–35.

Yao, Q. and Tong, H. (1996). Asymmetric least squares regression estimation: A nonparametric approach. *Journal of Nonparametric Statistics*, 6(2-3):273–292.

Yi, Y., Feng, X., and Huang, Z. (2014). Estimation of extreme value-at-risk: an EVT approach for quantile GARCH model. *Economics Letters*, 124(3):378–381.

Zhang, C., Liu, Y., and Wu, Y. (2016). On quantile regression in reproducing kernel Hilbert spaces with the data sparsity constraint. *Journal of Machine Learning Research*, 17(1):1374–1418.

College of Science, Nanjing Forestry University

E-mail: fchen@njfu.edu.cn

School of Statistics and Data Science, Southeast University

E-mail: caixingwang@seu.edu.cn