Mixed-Tail Quantile Functions for Extreme Value Analysis

Edoardo Redivo
University of Bologna
Alessio Farcomeni
University of Rome Tor Vergata
Cinzia Viroli
cinzia.viroli@unibo.it

University of Bologna

Research Article

Keywords: Order Statistics, Quantile Functions, M-estimation, Tail estimation

Posted Date: August 29th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7328610/v1

Additional Declarations: No competing interests reported.

Mixed-Tail Quantile Functions for Extreme Value Analysis

Edoardo Redivo[1†], Alessio Farcomeni[2†], Cinzia Viroli[3*†]

[2]University of Tor Vergata, Italy.
[1,3]University of Bologna, Italy.

*Corresponding author(s). E-mail(s): cinzia.viroli@unibo.it;
Contributing authors: edoardo.redivo@unibo.it;
alessio.farcomeni@uniroma2.it;
†These authors contributed equally to this work.

Abstract

This paper introduces a parametric mixed-tail quantile model which flexibly combines Gumbel, Fréchet, and Weibull tail behaviors through weighted quantile functions. Unlike classical extreme value models, our approach simultaneously captures multiple tail types, enhancing finite-sample adaptability and modeling accuracy. Parameters are estimated using a computationally efficient least squares approach based on the expected order statistics. We establish asymptotic properties, and address inference challenges via bootstrap methods. Theoretical results, including tail dominance and adaptive bias–variance decomposition, guide practical quantile estimation. Simulations and an application on global ice extents demonstrate improved performance in modeling extreme quantiles.

Keywords: Order Statistics, Quantile Functions, M-estimation, Tail estimation

1 Introduction

Extreme–value theory studies the stochastic behaviour of the *tails* of probability distributions and underpins applications ranging from climate-change assessment to production–frontier analysis and financial risk management (Coles, 2001; Chernozhukov et al., 2017).

A general framework for estimating conditional quantiles is *quantile regression* (Koenker and Bassett, 1978; Koenker, 2005). Given a response Y, covariates X and

a probability level $u \in (0,1)$, quantile regression models the conditional quantile $Q_{Y|X}(u) = x^\top \boldsymbol{\beta}(u)$ by minimising a check-loss function over $\boldsymbol{\beta}(u)$. Classical references include Koenker and D'Orey (1987) and Yu and Moyeed (2001), while time-series extensions are given by Cai and Stander (2008). For extremal quantiles, Chernozhukov (2005) showed that allowing covariates to interact multiplicatively with $Q(u)$ yields consistency and non-Gaussian limits under suitable regularity conditions.

A well-known drawback of fitting each u separately is the *crossing-quantile problem* (He, 1997; Bondell et al., 2010): estimated curves may violate monotonicity in u. Remedies include constrained regression, rearrangement, and penalisation (Dette and Volgushev, 2008; Chernozhukov et al., 2009, 2010; Neocleous and Portnoy, 2008; Liu and Wu, 2011; Andriyana et al., 2016). A fully parametric alternative is to impose a quantile–coefficient model $Q_{Y|X}(u;\boldsymbol{\theta})$ whose shape guarantees monotonicity; see Frumento and Bottai (2016) and the constrained refinement in Sottile and Frumento (2021).

Another route is to model the *entire* conditional quantile function parametrically. Comprehensive reviews are in Gilchrist (2000), and new families can be generated by algebraic operations on simpler quantile functions (Perepolkin et al., 2023). The power–Pareto model of Cai (2010) illustrates this idea for heavy-tailed data and has since been extended to censoring and dependence (Cai and Reeve, 2013; Cai, 2013, 2016). Formally,

$$Q_{Y|X}(u) = X^\top \boldsymbol{\beta} + Q(u, \boldsymbol{\theta}), \tag{1}$$

where $Q(u, \boldsymbol{\theta})$ captures departures from linearity, especially in the tails.

We propose a *mixed-tail* quantile function that blends the three classical domains of attraction—Gumbel, Fréchet and reversed Weibull—within a single parametric form. Like the Generalised Extreme-Value (GEV) distribution (Jenkinson, 1955, 1969), our model nests the three canonical shapes, but, crucially, it allows them to *coexist in finite samples* through data-driven weights. This extra flexibility is valuable when tail behaviour changes across sub-populations or regimes.

Estimation is performed jointly for all quantile levels via a closed-form least-squares criterion matching the model to the *expected* order statistics. Because the expectations of the three tail components admit analytic Beta–integral expressions, the objective can be evaluated without numerical integration and optimised by an alternating scheme: (i) a weighted least-squares update for the location/scale block, and (ii) a gradient-based step for the soft-max weights and the two shape parameters. The algorithm scales linearly in n and is typically much faster than fitting many separate check-loss regressions.

Section 5 shows that, under standard M-estimation regularity conditions (van der Vaart, 1998), the estimator is consistent and asymptotically normal; a simple bootstrap accounts for boundary cases in the mixing weights. Section 6 proves that, as $u \to 1$, exactly one of the three components dominates, giving a quantile-scale counterpart of the Fisher–Tippett–Gnedenko classification. For Fréchet-type tails, a second-order expansion yields a bias expression which, together with the usual $1/(n\epsilon)$ variance term, leads to an adaptive bias–variance decomposition and an explicit rule for selecting the extrapolation level in practice.

The paper is organised as follows. Section 2 reviews the three classical tail types. Section 3 defines the mixed-tail quantile function and its key properties. Section 4 details least-squares estimation via expected order statistics. Section 5 states the large-sample theory (consistency, asymptotic normality, bootstrap for boundary cases). Section 6 presents the dominance result, second-order expansions and the bias–variance decomposition. Section 7 reports simulation and ice-extent results, and Section 8 concludes. Proofs are given in the Appendix.

2 Preliminaries: Extreme Value Distributions

We briefly recall the classical domains of attraction from the Fisher–Tippett–Gnedenko Theorem (Resnick, 1987), which underpin most extreme value models based on the block maxima approach (David and Gumbel, 1960; Dombry, 2013). If properly normalized, the maximum of i.i.d. random variables converges to one of the three standard types of limiting distributions: Gumbel, Fréchet, or reversed Weibull.

Each type is associated with a canonical quantile function:

- Type I (Gumbel) — exponential-type tails:

$$Q^{(1)}(u;\mu,\sigma) = \mu - \sigma \log\{-\log u\}, \quad u \in (0,1). \tag{2}$$

- Type II (Fréchet) — heavy-tailed, polynomial decay:

$$Q^{(2)}(u;\mu,\sigma,\xi) = \mu + \sigma\{-\log u\}^{-1/\xi}, \quad \xi > 0. \tag{3}$$

- Type III (Weibull) — finite upper endpoint:

$$Q^{(3)}(u;\mu,\sigma,\xi) = \mu - \sigma\{-\log u\}^{1/\xi}, \quad \xi > 0. \tag{4}$$

A popular parametric family encompassing all three behaviours is the Generalized Extreme Value (GEV) distribution (Jenkinson, 1955, 1969), whose quantile function is:

$$Q(u;\mu,\sigma,\xi) = \begin{cases} \mu - \sigma \log\{-\log u\}, & \text{if } \xi = 0, \\ \mu + \dfrac{\sigma}{\xi}\left[\{-\log u\}^{-\xi} - 1\right], & \text{if } \xi \neq 0. \end{cases}$$

As is well known, $\xi = 0$ recovers the Gumbel form, $\xi > 0$ corresponds to the Fréchet case, and $\xi < 0$ leads to the reversed Weibull. The GEV is widely used in practice due to its ability to adapt to different tail behaviours with a single shape parameter.

However, in finite samples or heterogeneous populations, tail behaviour may not conform to a single limiting type. Motivated by this need, our model combines the quantile functions in (2)–(4) into a finite mixture with covariate-adjusted weights, enabling simultaneous representation of multiple tail regimes and improved modelling flexibility.

3 A Mixed Quantile Function

,

Suppose we have a random sample $y_1, \ldots, y_n$ drawn from a distribution function F_Y, with corresponding quantile function $Q_Y(u) = F_Y^{-1}(u)$ for $u \in (0, 1)$. Our goal is to construct a parametric quantile function that can flexibly accommodate multiple types of tail decay—namely Gumbel, Fréchet, and Weibull-like behaviour—within a single model.

3.1 Combining Tail Types via Convex Sums

A noteworthy property of quantile functions is that, under certain conditions, they can be linearly combined to form new valid quantile functions (Karvanen, 2006). Specifically, if $Q_1(u)$ and $Q_2(u)$ are valid quantile functions, then, for any $\lambda \in [0, 1]$, the function

$$\lambda Q_1(u) \; + \; (1 - \lambda)\, Q_2(u)$$

also defines a valid quantile function, provided monotonicity in u is preserved. This motivates us to build a *mixed* model that blends the three canonical tail forms (Type I, Type II, Type III) in a single framework. In contrast to the GEV distribution, which restricts the tail shape to a single regime (Gumbel, Fréchet, or Weibull), our mixture allows multiple tail types to coexist in finite samples. We will later show that, as $n \to \infty$, one component asymptotically dominates in the extreme tails, consistently with the Fisher–Tippett–Gnedenko Theorem. This added flexibility is particularly valuable when data originate from heterogeneous sources, each potentially exhibiting distinct tail characteristics.

3.2 Approximating the Canonical Tails Near $u \to 1$

Directly adding the exact quantile functions of Type I, Type II, or Type III from Section 2 can yield expressions too complex for practical estimation. In particular, they do not admit closed-form expressions for expected order statistics, which are essential in our estimation strategy. To gain simpler but asymptotically equivalent forms for extreme values (u close to 1), we approximate each canonical tail by its Taylor expansion around $u \to 1$ and truncate after the leading terms.

Type I (Gumbel).
Consider the quantile form

$$Q^{(1)}(u; \mu, \sigma) \; = \; \mu - \sigma \, \log\{-\log(u)\}.$$

Focusing on the simplified case $\mu = 0$, $\sigma = 1$, let $t = 1-u$ and expand $\log\{-\log(1-t)\}$ for small $t > 0$. Since $-\log(1 - t) = t + t^2/2 + O(t^3)$, we have

$$\log\{-\log(1 - t)\} = \log t + \frac{t}{2} + O(t^2),$$

and therefore

$$Q^{(1)}(u) \; \approx \; -\log(1 - u) \; - \; \frac{1}{2}\,(1 - u) \; = \; \tilde{Q}^{(1)}(u). \tag{5}$$

Here $-\log(1-u)$ is the exponential quantile capturing the leading behaviour as $u \to 1$, and the correction term improves finite-sample accuracy.

Type II (Fréchet/Pareto).
For the Fréchet family with lower endpoint μ_l and $\sigma = 1$:

$$Q^{(2)}(u;\, \mu_l, \xi) \;=\; \mu_l + \{-\log(u)\}^{-1/\xi}, \quad \xi > 0.$$

With $t = 1 - u$ small, $-\log(1-t) = t + t^2/2 + O(t^3)$, so

$$\{-\log(1-t)\}^{-1/\xi} = t^{-1/\xi}\left(1 - \frac{t}{2\xi} + O(t^2)\right),$$

yielding

$$Q^{(2)}(u;\, \mu_l, \xi) \;\approx\; \mu_l + (1-u)^{-\frac{1}{\xi}}\left(1 - \frac{1-u}{2\,\xi}\right) \;=\; \tilde{Q}^{(2)}(u;\, \mu_l, \xi). \tag{6}$$

The term $(1-u)^{-1/\xi}$ resembles the Pareto quantile, with $1/\xi$ the *extreme value index* controlling the tail heaviness.

Type III (Reversed Weibull).
For the Weibull-type tail with upper endpoint μ_u and $\sigma = 1$:

$$Q^{(3)}(u;\, \mu_u, \xi) \;=\; \mu_u \;-\; \{-\log(u)\}^{1/\xi}, \quad \xi > 0.$$

Expanding as before,

$$\{-\log(1-t)\}^{1/\xi} = t^{1/\xi}\left(1 + \frac{t}{2\xi} + O(t^2)\right),$$

we obtain

$$Q^{(3)}(u;\, \mu_u, \xi) \;\approx\; \mu_u \;-\; (1-u)^{1/\xi}\left(1 + \frac{1-u}{2\,\xi}\right) \;=\; \tilde{Q}^{(3)}(u;\, \mu_u, \xi). \tag{7}$$

This corresponds to a power-type tail approaching the endpoint at a rate determined by ξ.

3.3 Constructing the Mixed Quantile Function

Having approximated each tail form by $\tilde{Q}^{(j)}(u)$, $j = 1, 2, 3$, we now combine them with nonnegative weights w_1, w_2, w_3 that sum to 1. Let $\mu \in \mathbb{R}$ be a shift parameter and $\sigma > 0$ a scale parameter. We define the *mixed quantile function* as

$$Q^{(m)}\big(u;\, \boldsymbol{\theta}\big) = \mu + \sigma\left[w_1\, \tilde{Q}^{(1)}(u) + w_2\, \tilde{Q}^{(2)}\big(u;\, \mu_l, \xi_1\big) + w_3\, \tilde{Q}^{(3)}\big(u;\, \mu_u, \xi_2\big)\right], \tag{8}$$

where $w_1 + w_2 + w_3 = 1$ and each w_j controls the relative contribution of the corresponding tail type. In a finite sample, multiple components may be active simultaneously, capturing richer tail dynamics than a single GEV-like family. The parameters μ_l and μ_u appear separately from μ because they represent the lower and upper endpoints for the Fréchet and Weibull components, whereas μ is an intermediate shift for the Gumbel (Type I) part.

Remark 1 The inverse of $Q^{(m)}$ is not available in closed form. While (8) resembles a "mixture" of quantile functions, it is *not* the quantile function of a mixture distribution. Rather, it is a generator that produces a weighted average of component quantiles at the same u. This means an extreme quantile under $Q^{(m)}$ is a weighted average of extremes from components with different decay rates. For this reason, we refer to the specification as *mixed* rather than *mixture*.

Identifiability issues.

In simulation experiments, the model adapts well to the true tail type. However, when ξ_1 increases, the Fréchet tail $Q^{(2)}$ becomes lighter and approaches exponential decay, making it less distinguishable from Gumbel; similarly, for small ξ_2, the Weibull tail $Q^{(3)}$ can mimic Gumbel behaviour. To reduce this confounding, we constrain $1 < \xi_1 < a$ and $0 < \xi_2 < b$ with $a = 3$ and $b = 2$ in applications.

Example.

We generated $n = 200$ samples from: (i) a Gumbel distribution (location 0, scale 1), (ii) a Fréchet distribution (location 0, scale 1, shape $\xi = 2$), (iii) a reversed Weibull (location 0, scale 1, shape $\xi = 2$), and (iv) an equal-weight combination of the three quantiles. Table 1 reports the average estimated weights over 100 replications. The model correctly identifies the dominant tail component in most cases, but shows bias when components have similar tail heaviness.

Table 1: Average estimated weights $\hat{w}_j$ from 100 simulated samples.

True w	$\hat{w}_1$	$\hat{w}_2$	$\hat{w}_3$
$(1, 0, 0)$	0.68	0.04	0.28
$(0, 1, 0)$	0.24	0.72	0.04
$(0, 0, 1)$	0.26	0.00	0.74
$(1/3, 1/3, 1/3)$	0.46	0.31	0.23

3.4 Mixed-Tail Quantile Regression

Cai (2016); Cai and Reeve (2013) proposed a conditional quantile model of the form

$$Q_{Y|X}(u) = X^\top \boldsymbol{\beta} + Q(u, \theta),$$

where $Q(u, \theta)$ is a parametric quantile function capturing both body and tail behaviour. This enables fully parametric modelling of the entire conditional distribution, in contrast to fitting each quantile separately. Notably, Cai and Reeve (2013) specified $Q(u, \theta)$ as a product of a power quantile and a Pareto tail, suited for heavy-tailed outcomes.

We extend this idea by replacing $Q(u, \theta)$ with our mixed-tail quantile function, allowing light (Weibull-type), exponential (Gumbel-type), and heavy (Fréchet-type) tails to coexist:

$$Q^{(m)}_{Y|X}(u; X, \boldsymbol{\theta}) = \mu + X^\top \boldsymbol{\beta} + \sigma \sum_{j=1}^{3} w_j \, \tilde{Q}^{(j)}(u; \mu_l, \mu_u, \xi_1, \xi_2), \tag{9}$$

where the weights w_j are parameterised for identifiability via two logits relative to a baseline component (e.g., w_3):

$$w_j = \frac{\exp(\gamma_j)}{1 + \exp(\gamma_1) + \exp(\gamma_2)}, \quad j = 1, 2, \quad w_3 = 1 - w_1 - w_2.$$

This avoids the invariance to additive constants in $\boldsymbol{\gamma}$ that occurs with an unconstrained softmax on three parameters.

4 Model Estimation

In this section, we outline the least squares estimation procedure for the proposed mixed quantile model. This extends the approach in Redivo et al. (2023) to the mixed-tail specification. The main advantage is computational efficiency, especially for quantile functions admitting closed-form expectations of order statistics.

Endpoint constraints.

To improve identifiability and reduce overparameterisation, the endpoints μ_l and μ_u are not treated as free parameters. They could in principle be taken as the minimum and maximum of the observed range of y, but this choice is sensitive to outliers. To improve robustness, we instead set them to empirical quantiles of the data (e.g., the 0.05 and 0.95 empirical quantiles of y). This stabilises estimation in the presence of overlapping tail shapes.

4.1 Least Squares Estimation

Let $y_1, \ldots, y_n$ denote the observed responses, with associated covariate matrix $X \in \mathbb{R}^{n \times p}$. The conditional quantile at level u is modelled as

$$Q^{(m)}_{Y|X}(u; X, \boldsymbol{\theta}) = \mu + X^\top \boldsymbol{\beta} + \sigma \, \tilde{Q}^{(w)}(u; \boldsymbol{\vartheta}),$$

where $\tilde{Q}^{(w)}$ is the weighted combination of the three tail approximations introduced earlier:

$$\tilde{Q}^{(w)}(u;\boldsymbol{\vartheta}) = w_1\,\tilde{Q}^{(1)}(u) + w_2\,\tilde{Q}^{(2)}\big(u;\,\mu_l,\xi_1\big) + w_3\,\tilde{Q}^{(3)}\big(u;\,\mu_u,\xi_2\big),$$

with nonnegative weights summing to one.

To ensure identifiability of the weight parameters, we represent them through two logits relative to the third component:

$$w_j = \frac{\exp(\gamma_j)}{1+\exp(\gamma_1)+\exp(\gamma_2)}, \quad j=1,2, \qquad w_3 = 1-w_1-w_2.$$

This reparameterisation avoids the invariance to additive constants in $(\gamma_1,\gamma_2,\gamma_3)$ that would occur with an unconstrained softmax, and it improves numerical stability. The shape parameters (ξ_1,ξ_2) are further constrained to $1<\xi_1<a$ and $0<\xi_2<b$ in order to mitigate confounding with the Gumbel component when tail shapes become similar.

The estimation strategy relies on matching the fitted model to the empirical order statistics. Let $y_{(1)} \le \ldots \le y_{(n)}$ denote the ordered sample. Under the model, the expected value of the ith order statistic can be written as

$$E[Y_{(i)};\boldsymbol{\theta}] = \mu + X^\top\boldsymbol{\beta} + \sigma\Big(w_1E^{(1)}[Y_{(i)}] + w_2E^{(2)}[Y_{(i)};\xi_1] + w_3E^{(3)}[Y_{(i)};\xi_2]\Big),$$

where $E^{(j)}[Y_{(i)}]$ is the corresponding expectation under the jth tail component. These expectations can be expressed in closed form by integrating the component quantile functions against the Beta density $f_{i,n}(u) = \frac{u^{i-1}(1-u)^{n-i}}{B(i,n-i+1)}$:

$$\begin{aligned}
E^{(1)}[Y_{(i)}] &= \psi_0(n+1) - \psi_0(n-i+1) - \frac{n-i+1}{2(n+1)},\\
E^{(2)}[Y_{(i)};\xi_1] &= \mu_l + \frac{-1+\xi_1\left(2+\frac{i}{\xi_1 n+\xi_1-1}\right)}{2\xi_1}\cdot\frac{\Gamma(n+1)\,\Gamma(n+1-i-1/\xi_1)}{\Gamma(n+1-i)\,\Gamma(n+1-1/\xi_1)},\\
E^{(3)}[Y_{(i)};\xi_2] &= \mu_u - \frac{1+\xi_2\,(3-i+n+2\xi_2(n+1))}{2\xi_2^2}\cdot\frac{\Gamma(n+1)\,\Gamma(n+1-i+1/\xi_2)}{\Gamma(n+1-i)\,\Gamma(n+2+1/\xi_2)},
\end{aligned}$$

with ψ_0 denoting the digamma function.

Given these expressions, the least-squares criterion takes the form

$$\Phi_n(\boldsymbol{\theta}) = \sum_{i=1}^{n}\Big[y_{(i)} - \mu - X^\top\boldsymbol{\beta} - \sigma\left\{w_1E^{(1)}[Y_{(i)}] + w_2E^{(2)}[Y_{(i)};\xi_1] + w_3E^{(3)}[Y_{(i)};\xi_2]\right\}\Big]^2.$$

Minimisation of Φ_n is carried out by alternating between blocks of parameters: (i) update (ξ_1,ξ_2) given $(\mu,\sigma,\gamma_1,\gamma_2)$; (ii) update (γ_1,γ_2) by a gradient-based step; (iii) update $(\mu,\sigma,\boldsymbol{\beta})$ by weighted least squares. Although convergence to the global minimum cannot be guaranteed for such a nonconvex problem, in practice the algorithm

is numerically stable and converges to a stationary point for a wide range of starting values.

5 Large–Sample Theory

We study the least–squares estimator for the mixed–tail model within the framework of quasi-M–estimation. Since the data need not be generated by our model family, the estimation target is the *pseudo–true* parameter

$$\theta_0 := \arg\min_{\theta\in\Theta} \Phi(\theta), \qquad \Phi(\theta) := \mathbb{E}\big[\Phi_n(\theta)\big],$$

i.e., the minimiser of the population loss in the sense of White (1982). Its empirical counterpart is

$$\Phi_n(\theta) := \sum_{i=1}^{n} \big[y_{(i)} - E\{Y_{(i)};\theta\}\big]^2, \qquad \hat\theta := \arg\min_{\theta} \Phi_n(\theta), \tag{10}$$

where $y_{(i)}$ denotes the ith order statistic of the sample and $E\{Y_{(i)};\theta\}$ is its model-based expectation.

For clarity, we first present the results for the quantile function $Q^{(m)}(u)$ in the setting without covariates; the Appendix verifies that, under standard moment conditions, all conclusions (consistency and asymptotic normality) extend to the conditional case with covariates. Order statistics are indexed by $u_{(i)} = i/(n+1)$. The assumptions required for the asymptotic theory are listed below.

Assumptions

(A1) $Q^{(m)}(u;\boldsymbol{\theta})$ is twice continuously differentiable in $\boldsymbol{\theta}$ for every $u \in (0,1)$.
(A2) The population loss $\Phi(\boldsymbol{\theta}) = \mathbb{E}\{\Phi_n(\boldsymbol{\theta})\}$ has a unique minimiser θ_0.
(A3) Θ is compact and the shape parameters satisfy $1 < \xi_1 < a$, $0 < \xi_2 < b$.
(A4) $Q^{(m)}(u;\boldsymbol{\theta})$ is strictly increasing in u.
(A5) $Q^{(m)}$ and its gradient are uniformly bounded on Θ over $u \in (0,1)$.

The validity of (A1)–(A5) is demonstrated in the Appendix. Assumption (A3) places θ_0 in a compact parameter space. For the mixing weights, this corresponds to the simplex $\{w \in \mathbb{R}^3 : w_j \geq 0,\ \sum_j w_j = 1\}$, with the interior excluding the boundary. This compactness condition guarantees consistency of the estimator. However, if one or more estimated weights lie on the boundary, the usual asymptotic normality no longer applies; in such cases, valid inference can be obtained using the constrained bootstrap correction of Andrews (2000).

Under Assumptions (A1)–(A5), the general M–estimation theory of van der Vaart (1998, Th. 5.7) (see also Newey and McFadden, 1994) yields that the estimator is $\sqrt{n}$–consistent and asymptotically normal:

$$\sqrt{n}\,(\hat{\boldsymbol{\theta}} - \theta_0) \xrightarrow{\ d\ } \mathcal{N}\big(0,\ \nabla^2\Phi(\theta_0)^{-1}\,\Sigma\,\nabla^2\Phi(\theta_0)^{-1}\big),$$

where Σ is the covariance matrix of the score function.

Pointwise and local–uniform consistency.

Since $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$ and $Q^{(m)}(u;\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$ for each fixed $u \in (0,1)$, the continuous–mapping theorem gives

$$\hat{Q}^{(m)}(u) := Q^{(m)}\big(u;\hat{\boldsymbol{\theta}}\big) \xrightarrow{p} Q_Y(u) \qquad \text{for each fixed } u \in (0,1).$$

Moreover, this convergence is uniform over any closed sub–interval $[\varepsilon,\, 1-\varepsilon]$ with $\varepsilon > 0$: continuity of $Q^{(m)}$ in u, combined with a finite ε–net argument, implies

$$\sup_{u\in[\varepsilon,\,1-\varepsilon]} \big|\hat{Q}^{(m)}(u) - Q_Y(u)\big| \xrightarrow{p} 0.$$

Uniform convergence over the full grid $\{u_i = i/(n+1) : i = 1,\dots,n\}$ is not possible, because both $Q_Y(u)$ and $Q^{(m)}(u;\hat{\boldsymbol{\theta}})$ diverge as $u \uparrow 1$ (and possibly as $u \downarrow 0$). Therefore, uniform claims are restricted to compact subsets that remain a positive distance from the endpoints.

Extension to the conditional model.

Adding a covariate shift $X^\top\beta$ leaves the tail block unchanged. In the Appendix we show that, under $\mathbb{E}\|X\|^2 < \infty$, the consistency and asymptotic normality results continue to hold for $\boldsymbol{\theta} = (\mu, \sigma, \beta^\top, \vartheta^\top)^\top$.

5.1 Testing Hypotheses

While Assumptions (A1)–(A5) hold for the interior of the parameter space, inference for the mixing weights w_j can be delicate when the true value lies on (or near) the boundary of the simplex $\{w : w_j \geq 0,\ \sum_j w_j = 1\}$. In such cases, standard Taylor expansions may fail and Wald–type inference can be unreliable, even though consistency is unaffected (it relies on uniform convergence of the empirical loss and uniqueness of the minimiser).

This motivates a bootstrap approach that respects the constraints. A classical nonparametric bootstrap may fail in the presence of boundaries (Bickel and Freedman, 1981); however, first–order validity can be recovered by generating bootstrap samples under the null and re–estimating under the same constraints (see, e.g., Andrews, 2000; Chatterjee and Bose, 2005).

Constrained bootstrap scheme.

We use a restricted (null–imposed) bootstrap:

(i) Fit the model under H_0 (restricted estimator $\hat{\theta}^{(0)}$) and generate B bootstrap samples under H_0 (e.g., parametric draws from the fitted restricted model, or a residual/pairs bootstrap recentrato) so that the null and the simplex constraints on w are enforced in the bootstrap world.

(ii) For each bootstrap sample, re–estimate both the restricted and the unrestricted model under the same constraints on (w, ξ_1, ξ_2).

Test statistic and p–value.

Let

$$T_{\text{obs}} := \Phi_n\big(\hat{\theta}^{(0)}\big) - \Phi_n\big(\hat{\theta}\big) \ \geq 0,$$

i.e., the loss difference between the restricted and unrestricted fits (see (10)). For each bootstrap replication $b = 1, \ldots, B$, compute

$$T_b^* := \Phi_n\big(\hat{\theta}_b^{(0)}\big) - \Phi_n\big(\hat{\theta}_b\big).$$

The bootstrap p–value is then

$$\hat{p} \ = \ \frac{1 + \sum_{b=1}^{B} \mathbf{1}\{\, T_b^* \geq T_{\text{obs}} \,\}}{B+1}.$$

The same constrained bootstrap replicates can be used to obtain percentile (or basic) confidence intervals for any smooth functional of θ, re–estimating under the constraints at each replication.

6 Dominance, Expansion, and Bias–Variance Decomposition

We analyse the mixed–tail quantile in the extreme upper tail. Although, for fixed n, the estimator blends several tail types, as $u \to 1$ a single component dominates—providing a quantile–scale analogue of the domain–of–attraction principle.

We focus on quantile levels $u_n = 1 - \varepsilon_n$ with an intermediate sequence $\varepsilon_n = k_n/n$, where $k_n \to \infty$ and $k_n/n \to 0$ (e.g., $k_n = n^\beta$ with $\beta \in (0,1)$).

Theorem 1 (Asymptotic dominance of a single tail component) *Let*

$$Q^{(m)}(u) \ = \ \mu \ + \ \sigma\Big\{ w_1\, \tilde{Q}^{(1)}(u) \ + \ w_2\, \tilde{Q}^{(2)}(u; \mu_l, \xi_1) \ + \ w_3\, \tilde{Q}^{(3)}(u; \mu_u, \xi_2) \Big\},$$

with $w_j \geq 0$, $\sum_{j=1}^{3} w_j = 1$, and shape parameters satisfying $1 < \xi_1 < a$ and $0 < \xi_2 < b$. Assume $\mu_l < \mu < \mu_u$ and set $u_n = 1 - \varepsilon_n$ with $\varepsilon_n \downarrow 0$. Let $j^* \in \{1,2,3\}$ be any index such that

$$\big|\tilde{Q}^{(j^*)}(u_n)\big| \ = \ \max_{j=1,2,3} \big|\tilde{Q}^{(j)}(u_n)\big| \quad \textit{for all sufficiently large } n.$$

If $w_{j^*} > 0$, then

$$\frac{Q^{(m)}(u_n) - \mu}{\sigma\, \tilde{Q}^{(j^*)}(u_n)} \ = \ w_{j^*} \ + \ o(1) \quad \textit{as } n \to \infty,$$

equivalently,

$$\frac{Q^{(m)}(u_n) - \mu - \sigma w_{j^*} \tilde{Q}^{(j^*)}(u_n)}{\tilde{Q}^{(j^*)}(u_n)} \ \longrightarrow \ 0.$$

In words, $Q^{(m)}(u_n)$ is asymptotically equivalent to the dominant component $\tilde{Q}^{(j^*)}(u_n)$ scaled by w_{j^*}.

The theorem implies that only one tail shape drives the behaviour as $u \to 1$. In particular, since $\varepsilon_n^{1/\xi_2} \ll \log(1/\varepsilon_n) \ll \varepsilon_n^{-1/\xi_1}$ as $\varepsilon_n \downarrow 0$, the Fréchet-type term dominates whenever $w_2 > 0$; otherwise the Gumbel term dominates if $w_1 > 0$; if $w_1 = w_2 = 0$ then the reversed Weibull component determines the finite endpoint. In finite samples multiple components may be influential; dominance emerges only asymptotically.

To control the bias we use a second–order expansion under Fréchet–type tails. Assume the survival function satisfies $\bar{F}(y) = y^{-\xi}L(y)$ with $\xi > 0$, L slowly varying, and a standard second–order condition; then (see, e.g., de Haan and Ferreira, 2006)

$$Q_Y(1-\varepsilon) \;=\; c\,\varepsilon^{-1/\xi}\Big(1 + C\,\varepsilon^{\rho/\xi} + o\big(\varepsilon^{\rho/\xi}\big)\Big), \qquad \varepsilon \downarrow 0,$$

for some constants $c > 0$, $C \in \mathbb{R}$ and $\rho > 0$ governing the rate. We will employ this expansion to obtain the second–order bias term in Theorem 2.

Remark 2 (Other canonical tail types) For completeness, first–order expansions yield, as $\varepsilon \downarrow 0$,

Gumbel: $Q_Y(1-\varepsilon) \;=\; c\,\log(1/\varepsilon) + o(1)$, Reversed Weibull: $Q_Y(1-\varepsilon) \;=\; \mu - c\,\varepsilon^{1/\xi} + o\big(\varepsilon^{1/\xi}\big)$.

These follow directly from the respective quantile forms.

Building on the dominance result and a standard deviation bound, we now decompose the mean–squared error (MSE) of the extreme–quantile estimator into a variance term, driven by the effective sample size $n\varepsilon_n$, and a bias term whose rate depends on the tail type of the true distribution. To invoke classical $\sqrt{n}$ theory for misspecified M–estimators (van der Vaart, 1998; Newey and McFadden, 1994), we add the following interior–regularity conditions.

Additional regularity for $\sqrt{n}$–consistency

(A6) $\theta_0 \in \operatorname{int}(\Theta)$.
(A7) $\nabla_\theta \Phi(\theta)$ and $\nabla^2_\theta \Phi(\theta)$ are continuous on a neighbourhood of θ_0.
(A8) $\mathbb{E}\|X_i\|^2 < \infty$ (when covariates are present) and $\mathcal{I}(\theta_0) := \nabla^2_\theta \Phi(\theta_0)$ is non–singular.

Under (A1)–(A4) and (A6)–(A8), the quasi–M–estimator is $\sqrt{n}$–consistent and asymptotically normal:

$$\sqrt{n}\,(\hat{\theta} - \theta_0) \;\xrightarrow{\;d\;}\; \mathcal{N}\big(0,\, \mathcal{I}^{-1}(\theta_0)\big).$$

Theorem 2 (Adaptive bias–variance decomposition) *Fix $u_n = 1 - \varepsilon_n$ with $\varepsilon_n = k_n/n$, $k_n = n^\beta$, $\beta \in (0,1)$, and let $\hat{Q}^{(m)}(u_n) = Q^{(m)}(u_n; \hat{\theta})$. Assume the true tail admits a second–order representation (cf. de Haan and Ferreira, 2006): for some $Q_0(u_n)$ and $d_n \to 0$,*

$$Q_Y(u_n) = Q_0(u_n)\{1 + d_n\}, \qquad Q_0(u_n) = \begin{cases} c\,\varepsilon_n^{-1/\xi} & \textit{(Fréchet)}, \\ c\,\log(1/\varepsilon_n) & \textit{(Gumbel)}, \\ -\,c\,\varepsilon_n^{1/\xi} & \textit{(Weibull)}. \end{cases}$$

Suppose (A6)–(A8) hold so that $\|\hat{\theta} - \theta_0\| = O_p(n^{-1/2})$*, and that the dominant component in* $Q^{(m)}$ *matches the true tail type with positive weight* ($w_{j^*} > 0$)*. Then*

$$\mathbb{E}\Big[\{\hat{Q}^{(m)}(u_n) - Q_Y(u_n)\}^2\Big] \;=\; B_n^2 + V_n + o(B_n^2 + V_n),$$

with variance $V_n = O\{(n\varepsilon_n)^{-1}\}$ *and squared bias*

$$B_n^2 = \{\sigma Q_0(u_n) d_n\}^2\{1 + o(1)\}, \qquad B_n = \begin{cases} O(\varepsilon_n^{-1/\xi}\, d_n) & \textit{(Fréchet)}, \\ O(\log(1/\varepsilon_n)\, d_n) & \textit{(Gumbel)}, \\ O(\varepsilon_n^{1/\xi}\, d_n) & \textit{(Weibull)}. \end{cases}$$

Theorem 2 provides a unified decomposition of the estimation error into variance and bias components across tail types. The variance follows the classic effective–sample–size law $V_n = O\{(n\varepsilon_n)^{-1}\}$, reflecting that only the top ε_n fraction informs $u_n = 1 - \varepsilon_n$. The bias scales as $B_n \asymp Q_0(u_n)\, d_n$ and therefore depends sensitively on both the tail shape (Q_0) and the second–order regularity (d_n): heavy tails (Fréchet) amplify the bias through $Q_0(u_n) \sim \varepsilon_n^{-1/\xi}$ unless the second–order term decays fast, whereas for short tails (reversed Weibull) $Q_0(u_n) \sim \varepsilon_n^{1/\xi} \to 0$ and the bias attenuates more readily. These expressions enable principled trade–offs when choosing ε_n; in particular, Corollary 1 yields a rate–optimal choice balancing B_n^2 and V_n.

Corollary 1 (Rate–optimal choice of ε_n) *Under Theorem 2, suppose the second–order term satisfies* $d_n = A\varepsilon_n^{\rho/\xi^*}$ *with* $\rho > 0$*. Then*

$$\mathrm{MSE}(\varepsilon_n) = B_n^2 + V_n = C_1\, \varepsilon_n^{-2/\xi^* + 2\rho/\xi^*} \;+\; \frac{C_2}{n\,\varepsilon_n},$$

is minimised, to first order, by

$$\varepsilon_n^\star \;\asymp\; n^{-\,1/\,(1 - 2/\xi^* + 2\rho/\xi^*)},$$

provided $1 - 2/\xi^* + 2\rho/\xi^* > 0$*.*

The proof of Corollary 1 follows by differentiating $\mathrm{MSE}(\varepsilon_n) = B_n^2 + V_n = C_1\, \varepsilon_n^{-2/\xi^* + 2\rho/\xi^*} + C_2/(n\varepsilon_n)$ with respect to ε_n, setting the derivative to zero, and solving for ε_n. This yields, to first order, $\varepsilon_n^\star \asymp n^{-1/(1 - 2/\xi^* + 2\rho/\xi^*)}$ (provided $1 - 2/\xi^* + 2\rho/\xi^* > 0$).

Corollary 1 highlights that the optimal choice of ε_n depends critically on tail heaviness (ξ^*) and second–order regularity (ρ). In practice, this guideline can inform the selection of extreme quantile levels when estimating risk measures (e.g., Value–at–Risk) or return levels using the mixed–tail quantile model.

7 Empirical Analysis

7.1 Simulation Study

We conducted two simulation studies to evaluate extreme-quantile estimators under different data-generating processes (DGPs) and sample sizes. We focused on quantile

levels $p \in \{0.95, 0.99, 0.999\}$ and used the root mean square error (RMSE) as the performance metric over $B = 100$ replications:

$$\text{RMSE} = \sqrt{\frac{1}{B}\sum_{i=1}^{B}\big(\hat{q}_i - q_{\text{true}}\big)^2}.$$

Design.

In both studies we generate a covariate $x \sim \text{Unif}(0, 1)$ and a response $y = 3x + \varepsilon$ with:

1. *Pareto errors (Study 1):* $\varepsilon \sim \text{Pareto}(\text{shape} = 2,\ \text{scale} = 0.5) - 1$ (heavy tails).
2. *Log-normal errors (Study 2):* $\varepsilon \sim \text{LogNormal}(0, 1)$ (lighter tails).

Sample sizes are $n \in \{50, 100, 500\}$. For each replication and x, the *oracle* conditional quantile is $q_p(x) = 3x + q_p(\varepsilon)$; we therefore compute q_{true} analytically from the error distribution.[1]

We compare:

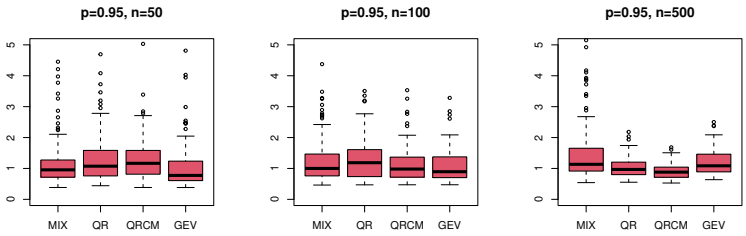
1. MIXQ: mixed-tail parametric quantile (this paper),
2. QR: classical quantile regression (check loss),
3. QRCM: quantile–coefficient model (monotone by construction),
4. GEV: parametric GEV regression.

Table 2: RMSE for the Pareto DGP (lower is better).

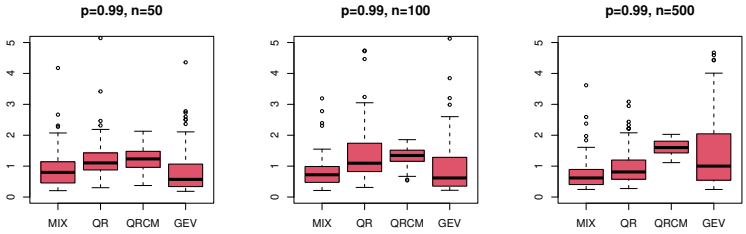
n	p	mixQ	QR	QRCM	GEV
50	0.999	1.927	3.141	2.651	2.412
100	0.999	11.716	14.329	16.663	8217715.434
500	0.999	9.671	13.168	16.443	23.985
50	0.99	1.927	3.141	2.651	2.412
100	0.99	1.559	3.646	2.525	6293.482
500	0.99	1.173	1.546	2.541	2.302
50	0.95	0.829	0.914	0.941	0.737
100	0.95	0.749	0.698	0.603	40.139
500	0.95	0.765	0.484	0.414	0.558

Across settings, MIXQ is competitive at $p = 0.95$ and improves relative to the baselines as p increases, with clear gains at $p \in \{0.99, 0.999\}$. QR and QRCM perform well for moderate quantiles but deteriorate in the extreme tail. GEV shows good performance in some moderate cases but exhibits instability at $p = 0.999$; the very large RMSE values indicate numerical failures or severe misspecification. For transparency, we recommend reporting the fraction of failed fits and robust summaries (e.g., median RMSE with IQR) alongside means.

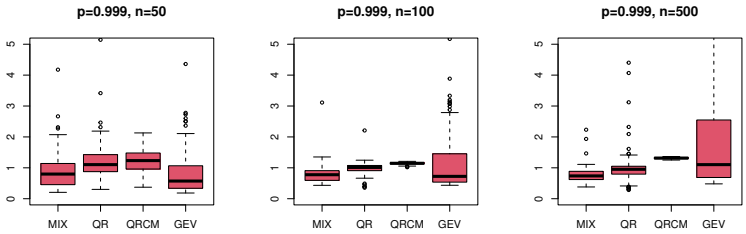
[1] If an analytic form is unavailable, replace "true" with "benchmark" and document the estimator (e.g., 'cmidecdf' from `Qtools`).



(a) RMSE for $p = 0.95$



(b) RMSE for $p = 0.99$



(c) RMSE for $p = 0.999$

Fig. 1: RMSE boxplots for the Pareto DGP across quantiles and sample sizes.

Performance improves with larger n for all methods; MIXQ maintains an advantage even at $n = 50$. The boxplots in Figures 1–2 confirm lower dispersion for MIXQ at high quantiles, indicating stability in rare-event regimes.

7.2 Data Analysis of Global Ice Extent Levels

To illustrate the approach, we investigate trends in the extremal levels of global sea-ice extent. We use $n = 898$ monthly measurements (January 1950–October 2024), and model the (scaled) ice extent (in $10^{13}\,\mathrm{m}^2$) using a quadratic time trend.

Table 3: RMSE for the Log-normal DGP (lower is better).

n	p	mixQ	QR	QRCM	GEV
50	0.999	3.196	4.732	5.013	4.622
100	0.999	7.437	9.701	13.033	24.080
500	0.999	4.209	6.717	13.214	30.433
50	0.99	3.196	4.732	5.013	4.622
100	0.99	2.488	4.483	4.894	3.164
500	0.99	1.446	1.936	5.009	3.569
50	0.95	1.421	1.857	1.848	1.166
100	0.95	1.260	1.313	1.230	0.761
500	0.95	0.774	0.629	0.944	0.359

Model choice.

Table 4 reports the loss (10) at convergence for each combination of tail components (G=Gumbel, F=Fréchet, W=Weibull). The preferred model is the Gumbel+Weibull combination; note that the equal loss for $(G, F, W) = (1, 1, 1)$ and $(1, 0, 1)$ indicates the Fréchet weight shrinks to zero.

Table 4: Ice data. Loss (10) at convergence for each tail-component combination. G = Gumbel, F = Fréchet, W = Weibull.

G	F	W	Loss
1	1	1	0.218
1	0	1	0.218
0	1	1	0.380
1	1	0	15.377
0	0	1	0.895
1	0	0	15.377
0	1	0	45.288

Estimates.

Parameter estimates and 95% bootstrap confidence intervals for the chosen model are in Table 5. Observed data and predicted 99% and 99.9% conditional quantiles (with 95% intervals) are shown in Figure 3.

Trend inference.

We test $H_0 : \beta_1 = \beta_2 = 0$ using the constrained bootstrap; the null is rejected ($p < 0.01$), indicating a time effect despite individual coefficients being non-significant. In a linear-only model, the slope is $\hat{\beta} = -0.02$ (bootstrap $p < 0.01$), consistent with a decreasing trend over time.[2]

[2]We recommend centring and scaling time (e.g., decades since 1987.5) to reduce collinearity between t and t^2.

Table 5: Ice data. Parameter estimates (Gumbel+Weibull model). CIs and p-values from the constrained bootstrap.

Parameter	Estimate	2.5% CI	97.5% CI	p-value
w_1 (Gumbel)	0.18	0.14	0.28	—
w_3 (Weibull)	0.82	0.71	0.86	—
μ	1.60	1.41	1.79	< 0.01
β_1 (time)	**0.83**	**-0.06**	**0.15**	0.167
β_2 (time^2)	-0.01	-0.01	0.01	0.464
σ	0.47	0.43	0.52	—
ξ_2 (Weibull shape)	0.45	0.38	0.58	—

Goodness of fit.

We predicted conditional quantiles for $u \in \{0.01, \ldots, 0.99\}$ and computed, at each u, the fraction of observations exceeding the prediction. Under correct calibration, this fraction should be close to $1-u$. Figure 4 shows good agreement across intermediate u.

Overall, the results indicate a clear decreasing trend in peak ice extent, with some evidence of acceleration in recent decades. The relatively small gap between the predicted 99% and 99.9% levels suggests a short (though non-negligible) upper tail under the chosen model. Physical tipping points have been discussed in the literature; our quantile forecasts offer an extreme-focused statistical corroboration and can be extrapolated to near-term horizons with appropriate uncertainty bands.

8 Concluding remarks

We proposed a fully parametric mixed–tail quantile model that blends the three classical domains of attraction (Gumbel, Fréchet, reversed Weibull) within a single, smoothly weighted quantile function. Estimation uses a closed–form least–squares fit to expected order statistics, so one global curve is learned at once rather than solving many check–loss problems; this avoids crossing and is computationally light.

Beyond standard M–estimation guarantees (consistency and asymptotic normality), two facts matter in practice. First, a *dominance* result shows that as $u \to 1$ the mixed tail behaves like a single component; this provides a quantile–scale counterpart to the classical domain–of–attraction picture. Second, a *bias–variance* decomposition—via a second–order tail expansion—yields a simple rule for choosing the extrapolation level $u_n = 1 - \varepsilon_n$ by balancing B_n^2 and $V_n = O\{(n\varepsilon_n)^{-1}\}$. Inference for mixing weights is handled with a constrained bootstrap, which is robust to boundary solutions.

Empirically, the model adapts across tail regimes: in simulations it is competitive at moderate quantiles and improves as p approaches one, especially when the true tail departs from any single canonical form. In the global sea–ice application, a Gumbel–Weibull blend provides a parsimonious description of the upper tail and confirms a clear downward trend in extreme ice extent.

The approach is most useful when tail behaviour is heterogeneous or uncertain and one seeks a single, monotone family for all quantiles. A practical workflow is: (i) fix μ_l, μ_u at empirical quantiles (e.g., 0.05/0.95) for identifiability; (ii) constrain $\xi_1 \in$

$(1, a)$, $\xi_2 \in (0, b)$ to limit confounding; (iii) report $k = n(1-u)$ alongside u to make the effective sample size explicit; (iv) use the constrained bootstrap for intervals/tests on weights and shapes; and (v) check calibration with the proportion–above plot against the $1 - u$ reference line.

Our specification is a quantile model, not a full density; covariates enter through a location shift, leaving the tail block common across X. Allowing covariate–dependent weights, handling serial dependence or censoring, and developing high–order refinements for the variance constant are natural next steps.

Overall, the mixed–tail quantile offers the flexibility of a mixture with the parsimony of a single family, making it a practical option for risk, environmental extremes, and other settings where tails may mix or evolve across regimes.

A Proofs of Theorems and Lemmas

This appendix collects all proofs and additional lemmas referenced in the main text. Section A.1 verifies Assumptions (A1)–(A5); Section A.2 extends the results to the covariate case; Sections A.3–A.4 contain second-order expansions and other auxiliary material.

A.1 Verification of Assumptions (A1)–(A5)

Lemma 1 *For every $\boldsymbol{\theta} \in \Theta$ with $w_j \geq 0$, $\sum_j w_j = 1$ and shape parameters $1 < \xi_1 < a$, $0 < \xi_2 < b$, the mixed-tail quantile $Q^{(m)}(u; \boldsymbol{\theta})$ in (8) satisfies Assumptions* (A1)–(A5) *in the interior of Θ.*

Proof (A1)–(A2) Each component $\tilde{Q}^{(j)}(u)$ is C^∞ on $(0, 1)$; any finite convex combination therefore remains C^2 in $\boldsymbol{\theta}$ and strictly increasing in u. The population minimiser is unique because the weights are identifiable ($w_j \neq w_{j'}$ for tails of different order) and the location/scale pair (μ, σ) is unrestricted.

(A3) The parameter space can be confined to the compact product set $[\underline{\mu}, \overline{\mu}] \times [\underline{\sigma}, \overline{\sigma}] \times [1, a] \times [0, b] \times \mathcal{W}$, where $\mathcal{W} := \{\boldsymbol{\gamma} \in \mathbb{R}^3\}$ is made compact via the soft-max map, which is continuous and bounded-to-one.

(A4) $\partial_u Q^{(m)}(u; \boldsymbol{\theta}) = \sigma \sum_j w_j \, \partial_u \tilde{Q}^{(j)}(u) > 0$ because each $\partial_u \tilde{Q}^{(j)}$ is positive on $(0, 1)$.

(A5) On a compact domain, $\max_\Theta \|\nabla_{\boldsymbol{\theta}} Q^{(m)}\|$ is finite: for Gumbel/Fréchet components the derivative grows at most like $C(1-u)^{-1/\xi_1}$, for Weibull like $C(1-u)^{1/\xi_2 - 1}$; both are integrable on $(0, 1)$ given $1 < \xi_1 < a$ and $0 < \xi_2 < b$. □

A.2 Impact of Covariates on the Large–Sample Theory

Our empirical model incorporates covariates through an additive linear shift $X^\top \boldsymbol{\beta}$ applied to the mixed–tail quantile function $Q^{(m)}(u)$. This modification affects only the location of the distribution while leaving the tail structure unchanged.

Let $Q^{(m)}_{Y|X}(u; X, \boldsymbol{\theta})$ denote the conditional quantile function with covariates, as defined in (8), and consider $\boldsymbol{\theta} = (\mu, \sigma, \boldsymbol{\beta}, \boldsymbol{\vartheta})$. Suppose the data (Y_i, X_i) are i.i.d., the covariates satisfy $\mathbb{E}\|X_i\|^2 < \infty$, and the regularity conditions in Lemma 1 hold for the tail block.

Under these assumptions, the least–squares estimator

$$\widehat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\vartheta}})$$

is consistent and asymptotically normal. Specifically,

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathcal{N}\left(0, \mathcal{I}^{-1}(\boldsymbol{\theta}_0)\right),$$

where $\mathcal{I}(\boldsymbol{\theta}_0)$ denotes the Godambe information matrix associated with the smooth M–estimator.

The result follows from standard theory: the score equations for $\boldsymbol{\beta}$ are

$$\frac{1}{n}\sum_{i=1}^{n} X_{(i)}\left\{Y_{(i)} - X_{(i)}^{\top}\boldsymbol{\beta} - Q^{(m)}(u_{(i)};\boldsymbol{\theta})\right\} = 0,$$

which satisfy the law of large numbers and central limit theorem under the finite–moment condition on X_i. The overall score function is smooth, and the Hessian is nonsingular due to the separability of the location and tail blocks.

In summary, including covariates provides a flexible, data–adaptive shift that can improve the fit in the central region of the distribution, without altering the large–sample properties of the tail block: consistency and asymptotic normality remain as in the no–covariate case.

A.3 Proof of Theorem 1

Let $u_n = 1 - \varepsilon_n$ with $\varepsilon_n \downarrow 0$. Using the local forms,

$$\tilde{Q}^{(1)}(u_n) = -\log(\varepsilon_n) + O(\varepsilon_n), \qquad \tilde{Q}^{(2)}(u_n) = \mu_l + \varepsilon_n^{-1/\xi_1}\Big(1 + O(\varepsilon_n)\Big), \qquad \tilde{Q}^{(3)}(u_n) = \mu_u - \varepsilon_n^{1/\xi_2}\Big(1 + O(\varepsilon_n)\Big).$$

As $\varepsilon_n \downarrow 0$,

$$\varepsilon_n^{1/\xi_2} \ll \log(1/\varepsilon_n) \ll \varepsilon_n^{-1/\xi_1},$$

hence, for all large n, one index $j^* \in \{1, 2, 3\}$ satisfies $\left|\tilde{Q}^{(j^*)}(u_n)\right| = \max_j \left|\tilde{Q}^{(j)}(u_n)\right|$. Define $r_j(u_n) := \tilde{Q}^{(j)}(u_n)/\tilde{Q}^{(j^*)}(u_n)$. Then $|r_j(u_n)| \le 1$ by construction and $r_j(u_n) \to 0$ for all $j \neq j^*$. Write

$$\frac{Q^{(m)}(u_n) - \mu}{\sigma\,\tilde{Q}^{(j^*)}(u_n)} = \sum_{j=1}^{3} w_j\, r_j(u_n) = w_{j^*} + \sum_{j \neq j^*} w_j\, r_j(u_n) \;\longrightarrow\; w_{j^*},$$

provided $w_{j^*} > 0$. This is equivalent to

$$\frac{Q^{(m)}(u_n) - \mu - \sigma w_{j^*}\tilde{Q}^{(j^*)}(u_n)}{\tilde{Q}^{(j^*)}(u_n)} \to 0,$$

which proves the claim. □

A.4 Proof of Theorem 2

Let $u_n = 1 - \varepsilon_n$, $\varepsilon_n = k_n/n \to 0$. Define the true extreme quantile

$$q_n := \mu + \sigma\, Q_Y(u_n),$$

where the tail admits the unified second–order representation

$$Q_Y(u_n) = Q_0(u_n)\{1 + d_n\}, \qquad d_n \to 0,$$

with

$$Q_0(u_n) = \begin{cases} c\,\varepsilon_n^{-1/\xi} & \text{(Fréchet)}, \\ c\,\log(1/\varepsilon_n) & \text{(Gumbel)}, \\ -\,c\,\varepsilon_n^{1/\xi} & \text{(Weibull)}. \end{cases}$$

(The constants μ_l, μ_u are fixed and do not affect the rates.)

Let $\hat{q}_n := \hat{\mu} + \hat{\sigma}\,\tilde{Q}^{(w)}(u_n; \hat{\vartheta})$. A first–order Taylor expansion at ϑ_0 gives

$$\tilde{Q}^{(w)}(u_n; \hat{\vartheta}) = \tilde{Q}^{(w)}(u_n; \vartheta_0) + \nabla_\vartheta \tilde{Q}^{(w)}(u_n; \vartheta_0)^\top (\hat{\vartheta} - \vartheta_0) + r_n, \quad r_n = o_p(n^{-1/2}).$$

Since $\hat{\mu} - \mu,\ \hat{\sigma} - \sigma,\ \hat{\vartheta} - \vartheta_0 = O_p(n^{-1/2})$ and the remainder is $o_p(n^{-1/2})$, we obtain

$$\hat{q}_n = \mu + \sigma\,\tilde{Q}^{(w)}(u_n; \vartheta_0) + O_p(n^{-1/2}).$$

Taking expectations and using L2–boundedness from asymptotic normality,

$$\mathbb{E}[\hat{q}_n] = \mu + \sigma\,\tilde{Q}^{(w)}(u_n; \vartheta_0) + O(n^{-1/2}) = \mu + \sigma\,\tilde{Q}^{(w)}(u_n; \vartheta_0) + o\big(Q_0(u_n)\big),$$

because $Q_0(u_n) \to \infty$ as $u_n \to 1$.

By the dominance result, the fitted tail satisfies

$$\tilde{Q}^{(w)}(u_n; \vartheta_0) = Q_0(u_n)\{1 + o(1)\},$$

hence

$$\mathbb{E}[\hat{q}_n] - q_n = \sigma\, Q_0(u_n)\Big(\{1 + o(1)\} - \{1 + d_n\}\Big) + o\big(Q_0(u_n)\big) = -\,\sigma\, Q_0(u_n)\, d_n\{1 + o(1)\}.$$

Therefore the squared bias is

$$B_n^2 = \{\sigma Q_0(u_n) d_n\}^2\{1 + o(1)\}.$$

For the variance, a delta–method argument yields

$$\mathbb{V}(\hat{q}_n) = \frac{1}{n}\,\nabla_\theta Q^{(m)}(u_n; \theta_0)^\top\, \mathcal{I}(\theta_0)^{-1}\, \nabla_\theta Q^{(m)}(u_n; \theta_0)\ \{1 + o(1)\}.$$

Under tail dominance, the sensitivity to the data is concentrated in the top $k_n = n\varepsilon_n$ fraction, leading to the effective–sample–size order

$$V_n := \mathbb{V}(\hat{q}_n) = O\Big(\frac{1}{n\varepsilon_n}\Big),$$

see, e.g., de Haan and Ferreira (2006) for the standard scaling heuristic.

Finally, by the identity $\mathrm{MSE}(\hat{q}_n) = \mathbb{V}(\hat{q}_n) + \{\mathbb{E}(\hat{q}_n) - q_n\}^2$,

$$\mathbb{E}[(\hat{q}_n - q_n)^2] = B_n^2 + V_n + o(B_n^2 + V_n),$$

with the stated tail–specific forms of B_n obtained from $Q_0(u_n)$. □

Funding. This work was partially supported by the following projects: European Commission's NextGeneration EU programme, PNRR – M4C2 – Investimento 1.3, Partenariato Esteso, PE00000013 - "FAIR - Future Artificial Intelligence Research" – Spoke 8 "Pervasive AI".

References

Andrews, D. W. K. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica 68*(2), 399–405.

Andriyana, Y., I. Gijbels, and A. Verhasselt (2016). Quantile regression in varying-coefficient models: non-crossing quantile curves and heteroscedasticity. *Statistical Papers 59*, 1589 – 1621.

Bickel, P. J. and D. A. Freedman (1981). Some asymptotic theory for the bootstrap. *Annals of Statistics 9*(6), 1196–1217.

Bondell, H. D., B. J. Reich, and H. Wang (2010, 08). Noncrossing quantile regression curve estimation. *Biometrika 97*(4), 825–838.

Cai, Y. (2010). Polynomial power-pareto quantile function models. *Extremes 13*, 291–314.

Cai, Y. (2013). A quantile survival model for censored data. *Australian & New Zealand Journal of Statistics 55*(2), 155–172.

Cai, Y. (2016). A general quantile function model for economic and financial time series. *Econometric Reviews 35*(7), 1173–1193.

Cai, Y. and D. E. Reeve (2013). Extreme value prediction via a quantile function model. *Coastal Engineering 77*, 91–98.

Cai, Y. and J. Stander (2008). Quantile self-exciting threshold autoregressive time series models. *Journal of Time Series Analysis 29*(1), 186–202.

Chatterjee, A. and A. Bose (2005). Bootstrap for parameter on the boundary. *Statistica Sinica 15*, 913–927.

Chernozhukov, V. (2005). Extremal quantile regression. *The Annals of Statistics 33*(2), 806–839.

Chernozhukov, V., I. Fernandez-Val, and T. Kaji (2017, December). Extremal quantile regression: an overview. CeMMAP working papers 65/17, Institute for Fiscal Studies.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika 96*(3), 559–575.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010). Quantile and Probability Curves without Crossing. *Econometrica 78*(3), 1093–1125.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer London.

David, F. N. and E. J. Gumbel (1960). Statistics of extremes. *Biometrika 47*, 209.

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer.

Dette, H. and S. Volgushev (2008). Non-crossing non-parametric estimates of quantile curves. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 70*(3), 609–627.

Dombry, C. (2013). Maximum likelihood estimators for the extreme value index based on the block maxima method. *arXiv: Probability*.

Frumento, P. and M. Bottai (2016). Parametric modeling of quantile regression coefficient functions. *Biometrics 72*.

Gilchrist, W. (2000). *Statistical Modelling with Quantile Functions*. Chapman and Hall/CRC.

He, X. (1997). Quantile curves without crossing. *The American Statistician 51*(2), 186–192.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of Royal Meteorological Society 81*(348), 158–171.

Jenkinson, A. F. (1969). Estimation of maximum floods. *World Meteorological Organization Technical Note 98*(5), 183–227.

Karvanen, J. (2006). Estimation of quantile mixtures via L-moments and trimmed L-moments. *Computational Statistics and Data Analysis 51*(2), 947–959.

Koenker, R. (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press.

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica 46*(1), 33–50.

Koenker, R. W. and V. D'Orey (1987). Computing regression quantiles. *Journal of the Royal Statistical Society: Series C 36*(3), 383–393.

Liu, Y. and Y. Wu (2011). Simultaneous multiple non-crossing quantile regression estimation using kernel constraints. *Journal of Nonparametric Statistics 23*, 415 – 437.

Neocleous, T. and S. Portnoy (2008). On monotonicity of regression quantile functions. *Statistics & Probability Letters 78*(10), 1226–1229.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4 of *Handbook of Econometrics*, Chapter 36, pp. 2111–2245. Amsterdam: Elsevier.

Perepolkin, D., B. Goodrich, and U. Sahlin (2023). The tenets of quantile-based inference in bayesian models. *Computational Statistics & Data Analysis 187*, 107795.

Redivo, E., C. Viroli, and A. Farcomeni (2023). Quantile-distribution functions and their use for classification, with application to naïve Bayes classifiers. *Statist. Comput. 33*(2).

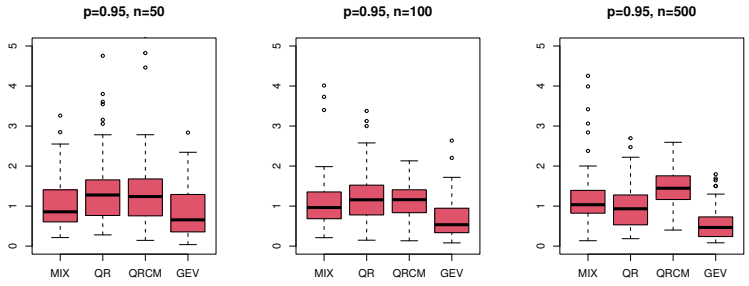
Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer-Verlag.

Sottile, G. and P. Frumento (2021). Parametric estimation of non-crossing quantile functions. *Statistical Modelling 23*, 173 – 195.

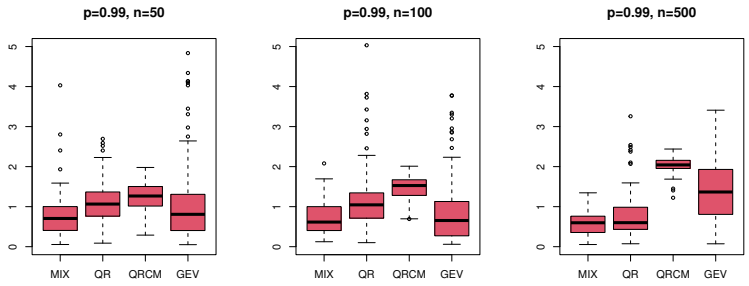
van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica 50*, 1–25.

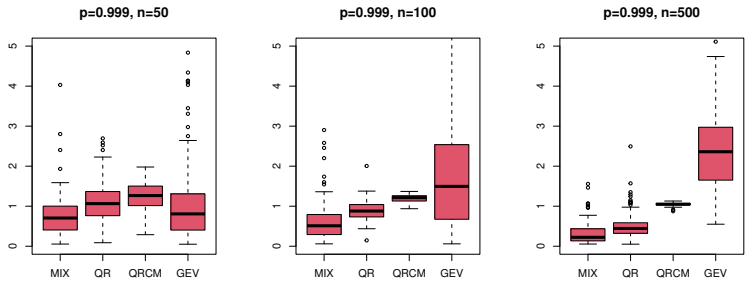
Yu, K. and R. A. Moyeed (2001). Bayesian quantile regression. *Statistics & Probability Letters 54*(4), 437–447.



(a) RMSE for $p = 0.95$



(b) RMSE for $p = 0.99$



(c) RMSE for $p = 0.999$

Fig. 2: RMSE boxplots for the Log-normal DGP across quantiles and sample sizes.

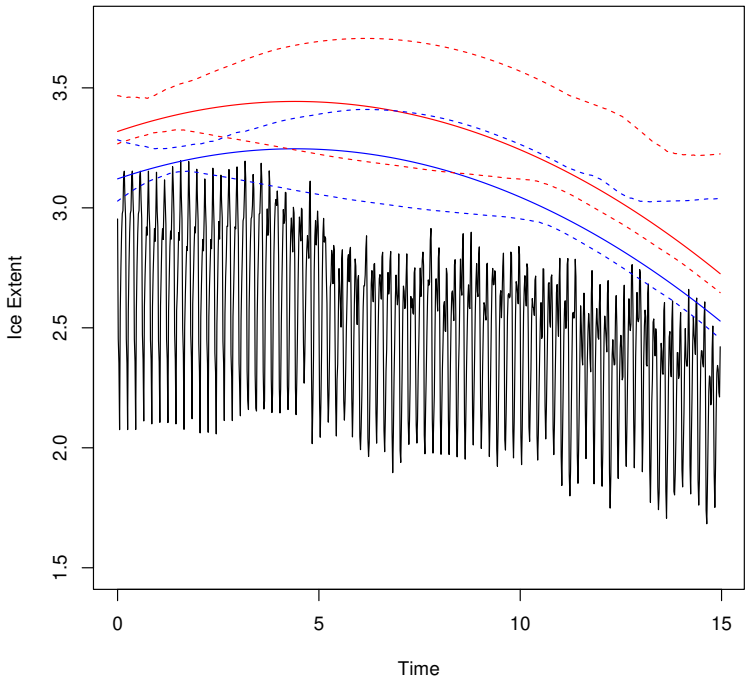


Fig. 3: Ice data. Observed series and predicted 99% (blue) and 99.9% (red) conditional quantiles with 95% bootstrap intervals (dashed).

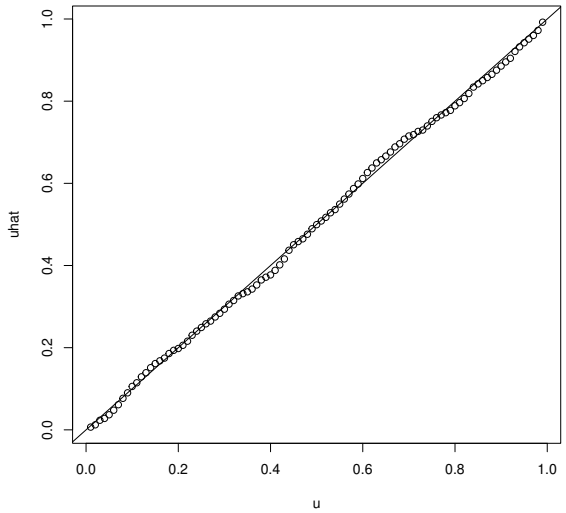


Fig. 4: Ice data. Proportion above the predicted u-quantile versus u (target reference line $1 - u$ shown).