

Out-of-Distribution generalization of quantile regression with heavy tailed inputs: an SVM approach

Baptiste Leroux, Clément Dombry, Anne Sabourin

October 8, 2026

Table of contents

Quantile regression and the SVM approach

Extrapolation and the extreme covariate setting

Statistical guarantees

Case study on the Danube dataset

Quantile regression and the SVM approach

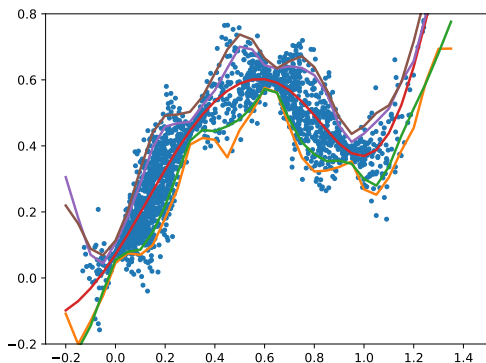
Quantile regression

Let $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ be a couple of random variables.

$(X, Y) \sim \mathbb{P}$. $\tau \in (0, 1)$.

- **Quantile regression:** Estimation of $q(x)$ the τ -**conditional quantile** of Y given $X = x$:

$$q(x) = \inf\{t \in \mathbb{R} : \mathbb{P}(Y \leq t \mid X = x) \geq \tau\}.$$

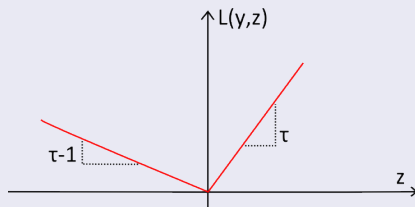


The pinball loss

Definition

The pinball loss $\ell_\tau: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ is defined by

$$\ell_\tau(y, z) = (1 - \tau)(z - y)\mathbb{1}_{y < z} + \tau(y - z)\mathbb{1}_{y \geq z}$$



Theorem

If the target Y is bounded, the τ -conditional quantile function is a minimizer of the pinball risk $\mathcal{R}(f) = \mathbb{E}[\ell_\tau(Y, f(X))]$ among all measurable bounded functions.

The modified pinball loss

To get rid of the boundedness assumption on the target Y : we consider the modified pinball loss

$$\ell'_\tau(y, z) = \ell_\tau(y, z) - \ell_\tau(y, 0)$$

The modified pinball loss

To get rid of the boundedness assumption on the target Y : we consider the modified pinball loss

$$\ell'_\tau(y, z) = \ell_\tau(y, z) - \ell_\tau(y, 0)$$

Theorem

if The τ -conditional quantile function satisfies $\mathbb{E} [|q(X)|] < \infty$, it is a minimizer of the modified pinball risk $\mathcal{R}(f) = \mathbb{E} [\ell'_\tau(Y, f(X))]$ among all measurable functions such that $\mathbb{E} [|f(X)|] < \infty$.

—→ Estimation of the conditional quantile function can be achieved by **minimizing the modified pinball loss**

Empirical Risk Minimization (ERM)

- ▶ The conditional quantile function minimizes the modified pinball risk:

$$\mathcal{R}(q) = \mathcal{R}^* = \inf\{\mathcal{R}(f) : f \text{ measurable}, \mathbb{E}[|f(X)|] < \infty\}$$

- ▶ **Natural strategy:** Minimize the risk functional over a certain class of integrable predictor.

Empirical Risk Minimization (ERM)

- ▶ The conditional quantile function minimizes the modified pinball risk:

$$\mathcal{R}(q) = \mathcal{R}^* = \inf \{ \mathcal{R}(f) : f \text{ measurable, } \mathbb{E} [|f(X)|] < \infty \}$$

- ▶ **Natural strategy:** Minimize the risk functional over a certain class of integrable predictor.
- ▶ **Problem:** The underlying probability distribution $\mathbb{P}$ is unknown!
- ▶ **Idea:** Replace the true risk by an empirical estimate based on the sample $((X_1, Y_1), \dots, (X_n, Y_n))$

$$\mathcal{R}(f) = \mathbb{E} [\ell'_\tau(Y, f(X))] \longleftrightarrow \hat{\mathcal{R}}(f) = \frac{1}{n} \sum_{i=1}^n \ell'_\tau(Y_i, f(X_i))$$

Empirical Risk Minimization (ERM)

- Choose a class of predictor. For instance

$$\mathcal{F} = \{f: x \mapsto \langle w, x \rangle, w \in \mathbb{R}^d\}$$

- The empirical modified pinball risk is

$$\widehat{\mathcal{R}}(f) = \frac{1}{n} \sum_{i=1}^n \ell'_\tau(Y_i, f(X_i))$$

- The associated predictor is

$$\widehat{q} \in \operatorname{argmin}\{\widehat{\mathcal{R}}(f): f \in \mathcal{F}\}$$

Empirical Risk Minimization (ERM)

- ▶ Choose a class of predictor. For instance

$$\mathcal{F} = \{f: x \mapsto \langle w, x \rangle, w \in \mathbb{R}^d\}$$

- ▶ The empirical modified pinball risk is

$$\widehat{\mathcal{R}}(f) = \frac{1}{n} \sum_{i=1}^n \ell'_\tau(Y_i, f(X_i))$$

- ▶ The associated predictor is

$$\widehat{q} \in \operatorname{argmin}\{\widehat{\mathcal{R}}(f): f \in \mathcal{F}\}$$

- ▶ **Problem:** ERM does not behave well in high dimensional settings (*curse of dimensionality*), and constrains the predictor class to low complexity models.

Another approach: SVM regression

Key idea: Penalize the empirical estimate of the risk.

- ▶ Look for a predictor in a large, flexible Hilbert space $\mathcal{H}$:
a **Reproducing Kernel Hilbert Space (RKHS)**
- ▶ based on a sample $(X_i, Y_i)_{i \in \{1, \dots, n\}}$, the estimator is:

$$\hat{q} \in \operatorname{argmin}_{h \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^n \ell_{\tau}(Y_i, h(X_i)) + \lambda \|h\|_{\mathcal{H}}^2.$$

Another approach: SVM regression

Key idea: Penalize the empirical estimate of the risk.

- ▶ Look for a predictor in a large, flexible Hilbert space $\mathcal{H}$:
a **Reproducing Kernel Hilbert Space (RKHS)**
- ▶ based on a sample $(X_i, Y_i)_{i \in \{1, \dots, n\}}$, the estimator is:

$$\hat{q} \in \operatorname{argmin}_{h \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^n \ell_{\tau}(Y_i, h(X_i)) + \lambda \|h\|_{\mathcal{H}}^2.$$

- ▶ Computationally efficient, good empirical performance and literature on achievable learning rates. Behaves well in high dimensional settings

Extrapolation and the extreme covariate setting

Extrapolation for quantile regression

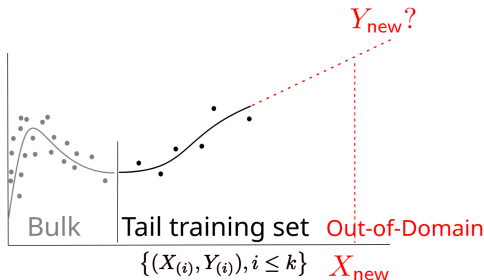
$(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ be a couple of random variables. $(X, Y) \sim \mathbb{P}$.

- **Goal:** Estimate q the τ -conditional quantile function

Extrapolation for quantile regression

$(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ be a couple of random variables. $(X, Y) \sim \mathbb{P}$.

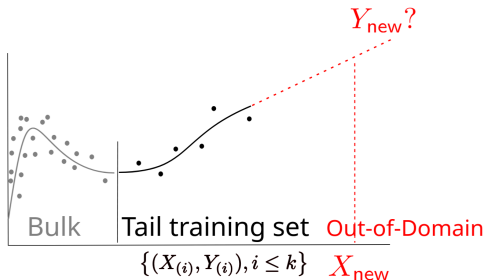
- ▶ **Goal:** Estimate q the τ -conditional quantile function
- ▶ with a twist: Estimation needs to accurate when $\|X\|$ is exceptionally large.
- ▶ Motivation: Build reliable predictions in a "crisis scenario".



Extrapolation for quantile regression

$(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ be a couple of random variables. $(X, Y) \sim \mathbb{P}$.

- ▶ **Goal:** Estimate q the τ -conditional quantile function
- ▶ with a twist: Estimation needs to be accurate when $\|X\|$ is exceptionally large.
- ▶ Motivation: Build reliable predictions in a "crisis scenario".



Several challenges to overcome:

- ▶ Few extreme examples in the training set.
- ▶ How to extrapolate beyond the range of the training set?

Why do we need a specific procedure?

- ▶ We have the following decomposition of the risk

$$\begin{aligned}\mathcal{R}(f) &= \mathbb{P}(\|X\| < t) \mathbb{E}[\ell(Y, f(X)) \mid \|X\| < t] \\ &\quad + \underbrace{\mathbb{P}(\|X\| \geq t) \mathbb{E}[\ell(Y, f(X)) \mid \|X\| \geq t]}_{\text{negligible as } t \rightarrow \infty!}\end{aligned}$$

- ▶ Minimization of the (penalized) risk may not take into account the extreme region $\{\|X\| \geq t\}$ for large values of t .

Existing literature

- ▶ Quantile regression and extrapolation:
Buriticá and Engelke, 2024 and de Carvalho et al., 2022

Limitations:

- Only estimation of the median in the univariate case for Buriticá and Engelke, 2024
- Parametric model in de Carvalho et al., 2022

Existing literature

- ▶ Quantile regression and extrapolation:
Buriticá and Engelke, 2024 and de Carvalho et al., 2022

Limitations:

- Only estimation of the median in the univariate case for Buriticá and Engelke, 2024
 - Parametric model in de Carvalho et al., 2022
- ▶ Extrapolation also addressed in Jalalzai et al., 2018 and Huet et al., 2023 for binary classification and least-square regression under **regular variation** assumptions.

Limitations:

- Only ERM is considered (**no penalization!**)
- Guarantees rely on **VC dimension assumptions and boundedness of the target Y**

Extreme covariate setting

- ▶ Instead of minimizing the global (modified) pinball error $\mathcal{R}(f) = \mathbb{E} [\ell_\tau(Y, f(X))]$, we consider a conditional version: given $t \geq 0$,

$$\mathcal{R}_t(f) = \mathbb{E} [\ell'_\tau(Y, f(X)) \mid \|X\| \geq t]$$

- ▶ Take the limit as $t \rightarrow \infty$: we obtain the **asymptotic risk**

$$\mathcal{R}_\infty(f) = \limsup_{t \rightarrow \infty} \mathcal{R}_t(f).$$

Extreme covariate setting

- ▶ Instead of minimizing the global (modified) pinball error $\mathcal{R}(f) = \mathbb{E} [\ell_\tau(Y, f(X))]$, we consider a conditional version: given $t \geq 0$,

$$\mathcal{R}_t(f) = \mathbb{E} [\ell'_\tau(Y, f(X)) \mid \|X\| \geq t]$$

- ▶ Take the limit as $t \rightarrow \infty$: we obtain the **asymptotic risk**

$$\mathcal{R}_\infty(f) = \limsup_{t \rightarrow \infty} \mathcal{R}_t(f).$$

- ▶ We look for a minimizer of the **asymptotic risk**. Guarantees will be expressed in terms of the excess asymptotic risk

$$\mathcal{R}_\infty(\hat{q}) - \mathcal{R}_\infty^*$$

where $\mathcal{R}_\infty^* = \inf \{ \mathcal{R}_\infty(f) : \mathbb{E} [\|f(X)\|] < \infty \}.$

Regular variation

- ▶ Key assumption: **Regular variation** with respect to the covariate X

Assumption (Regular variation assumption)

The pair (X, Y) is first-component regularly varying:

$$(X/t, Y \mid \|X\| \geq t) \xrightarrow[t \rightarrow \infty]{\mathcal{L}} (X_\infty, Y_\infty)$$

- ▶ The idea is that the behaviour of (X, Y) for large X can be understood through the limiting distribution (X_∞, Y_∞) .
- ▶ An *extreme risk* can be defined

$$\mathcal{R}_{P_\infty}(f) = \mathbb{E} [\ell'_\tau(Y_\infty, f(X_\infty))]$$

and an *extreme quantile function*

$$q_{P_\infty}(x) = \text{Quantile}(Y_\infty \mid X_\infty = x)$$

The extreme quantile is of angular type

- Under the regular variation assumption,

$$\|X_\infty\| \perp (\theta(X_\infty), Y_\infty)$$

which yields the crucial lemma:

Lemma

Under the first-component regular variation assumption, the extreme conditional function q_{P_∞} is of angular type: for any $x \in \mathbb{R}^d$,

$$q_{P_\infty}(x) = q_{P_\infty}(\theta(x))$$

The extreme quantile is of angular type

- Under the regular variation assumption,

$$\|X_\infty\| \perp (\theta(X_\infty), Y_\infty)$$

which yields the crucial lemma:

Lemma

Under the first-component regular variation assumption, the extreme conditional function q_{P_∞} is of angular type: for any $x \in \mathbb{R}^d$,

$$q_{P_\infty}(x) = q_{P_\infty}(\theta(x))$$

- What's the link with the minimization of the asymptotic risk?

Minimizing the asymptotic risk

Assumption (regularity assumption)

The extreme conditional quantile function q_{P_∞} is continuous on $\mathbb{R}^d \setminus \{0\}$, and as t tends to infinity,

$$\mathbb{E} [|q(X) - q_{P_\infty}(X)| \mid \|X\| \geq t] \xrightarrow[t \rightarrow \infty]{} 0$$

Minimizing the asymptotic risk

Assumption (regularity assumption)

The extreme conditional quantile function q_{P_∞} is continuous on $\mathbb{R}^d \setminus \{0\}$, and as t tends to infinity,

$$\mathbb{E} [|q(X) - q_{P_\infty}(X)| \mid \|X\| \geq t] \xrightarrow[t \rightarrow \infty]{} 0$$

Theorem

Under the regular variation assumption and the regularity assumption,

- 1. For any angular continuous function f , $\mathcal{R}_\infty(f) = \mathcal{R}_{P_\infty}(f)$.*
- 2. The extreme conditional quantile function is also a minimizer of the asymptotic risk.*

Proposed algorithm

- ▶ **Input:** Training dataset $\mathcal{D}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$, RKHS $\mathcal{H}$, number $k \leq n$ of extreme observations, $\lambda > 0$ and $\tau \in (0, 1)$.
- ▶ **Training:** Sort the training sample: $((X_{(1)}, Y_{(1)}), \dots, (X_{(n)}, Y_{(n)}))$ where $\|X_{(1)}\| \geq \dots \geq \|X_{(n)}\|$. Learn from the k most extreme **angles**:

$$\hat{h}_\lambda \in \operatorname{argmin}_{h \in \mathcal{H}} \frac{1}{k} \sum_{i=1}^k \ell'_\tau(Y_{(i)}, h(\theta(X_{(i)}))) + \lambda \|h\|_{\mathcal{H}}^2$$

where $\theta(x) = x/\|x\|$

- ▶ **Estimator:** $\hat{q} = \hat{h}_\lambda \circ \theta$

Predicting the component of a regularly varying vector

Proposition

Let $(X, Z) \in \mathbb{R}^d \times \mathbb{R}$ with $\mathbb{P}(X = 0) = 0$.

- ▶ Assume that the pair (X, Z) has positive continuous density π on $\mathbb{R}^{d+1} \setminus \{0\}$.
- ▶ Assume also that there exists some regularly varying function $b(t)$ with tail index $\alpha > 0$ and some positive, continuous density π_∞ such that

$$\sup_{\|(x,z)\|>1} |t^{d+1}b(t)\pi(tx,tz) - \pi_\infty(x,z)| \xrightarrow[t \rightarrow \infty]{} 0.$$

- ▶ Denote $\pi_\infty(x) = \int_{\mathbb{R}} \pi_\infty(x,z) \, dz$, and assume

$$\inf_{\omega \in \mathbb{S}} \pi_\infty(\omega) \geq c > 0.$$

Then, the pair $(X, Z/\|X\|)$ satisfies the regular variation and the stability assumptions.

Statistical guarantees

The different risk functionals

Let $t_{n,k}$ be the $1 - \frac{k}{n}$ quantile of $\|X\|$. Several risk functional are involved in our analysis.

- ▶ The true conditional risk:

$$\mathcal{R}_{t_{n,k}}(h \circ \theta) = \mathbb{E} [\ell'_\tau(Y, h(\theta(X))) \mid \|X\| \geq t_{n,k}] ,$$

- ▶ the extreme/asymptotic risk:

$$\mathcal{R}_\infty(h \circ \theta) = \mathbb{E} [\ell'_\tau(Y_\infty, h(\theta(X_\infty)))] ,$$

The different risk functionals

Let $t_{n,k}$ be the $1 - \frac{k}{n}$ quantile of $\|X\|$. Several risk functional are involved in our analysis.

- ▶ The true conditional risk:

$$\mathcal{R}_{t_{n,k}}(h \circ \theta) = \mathbb{E} [\ell'_\tau(Y, h(\theta(X))) \mid \|X\| \geq t_{n,k}],$$

- ▶ the extreme/asymptotic risk:

$$\mathcal{R}_\infty(h \circ \theta) = \mathbb{E} [\ell'_\tau(Y_\infty, h(\theta(X_\infty)))] ,$$

- ▶ the empirical risk:

$$\widehat{\mathcal{R}}_k(h \circ \theta) = \frac{1}{k} \sum_{i=1}^k \ell'_\tau(Y_{(i)}, h(\theta(X_{(i)}))),$$

- ▶ its pseudo-empirical equivalent:

$$\widetilde{\mathcal{R}}_k(h \circ \theta) = \frac{1}{k} \sum_{i=1}^n \ell'_\tau(Y_i, h(\theta(X_i))) \mathbf{1}_{\|X_i\| \geq t_{n,k}}.$$

Different SVM solutions

Each risk functional is associated to a certain *SVM solution*:

- ▶ The conditional SVM solution $h_{\lambda,k}^*$

$$h_{\lambda,k}^* \in \operatorname{argmin}_{h \in \mathcal{H}} \mathcal{R}_{t_{n,k}}(h \circ \theta) + \lambda \|h\|_{\mathcal{H}}^2$$

- ▶ The asymptotic SVM solution $h_{\lambda,\infty}^*$

$$h_{\lambda,\infty}^* \in \operatorname{argmin}_{h \in \mathcal{H}} \mathcal{R}_{\infty}(h \circ \theta) + \lambda \|h\|_{\mathcal{H}}^2$$

Different SVM solutions

Each risk functional is associated to a certain *SVM solution*:

- ▶ The conditional SVM solution $h_{\lambda,k}^*$

$$h_{\lambda,k}^* \in \operatorname{argmin}_{h \in \mathcal{H}} \mathcal{R}_{t_{n,k}}(h \circ \theta) + \lambda \|h\|_{\mathcal{H}}^2$$

- ▶ The asymptotic SVM solution $h_{\lambda,\infty}^*$

$$h_{\lambda,\infty}^* \in \operatorname{argmin}_{h \in \mathcal{H}} \mathcal{R}_{\infty}(h \circ \theta) + \lambda \|h\|_{\mathcal{H}}^2$$

- ▶ The empirical SVM solution $\hat{h}_{\lambda}$ (our estimator!)

$$\hat{h}_{\lambda} \in \operatorname{argmin}_{h \in \mathcal{H}} \hat{\mathcal{R}}_k(h \circ \theta) + \lambda \|h\|_{\mathcal{H}}^2,$$

- ▶ The pseudo-empirical SVM solution $\tilde{h}_{\lambda,k}$

$$\tilde{h}_{\lambda,k} \in \operatorname{argmin}_{h \in \mathcal{H}} \tilde{\mathcal{R}}_k(h \circ \theta) + \lambda \|h\|_{\mathcal{H}}^2$$

The SVM solutions are well-defined

Lemma

Assume the kernel K is bounded with $\sup_{x \in \mathbb{R}^d} K(x, x) = 1$. Then, the SVM solutions $h_{\lambda, k}^*$, $h_{\lambda, \infty}^*$, $\hat{h}_\lambda$, $\tilde{h}_{\lambda, k}$ are well defined. Moreover,

$$\max \left(\|\hat{h}_\lambda\|_{\mathcal{H}}, \|h_{\lambda, k}^*\|_{\mathcal{H}}, \|h_{\lambda, \infty}^*\|_{\mathcal{H}} \right) \leq \frac{\max(\tau, 1 - \tau)}{\lambda}.$$

The SVM solutions are well-defined

Lemma

Assume the kernel K is bounded with $\sup_{x \in \mathbb{R}^d} K(x, x) = 1$. Then, the SVM solutions $h_{\lambda, k}^*$, $h_{\lambda, \infty}^*$, $\hat{h}_\lambda$, $\tilde{h}_{\lambda, k}$ are well defined. Moreover,

$$\max \left(\|\hat{h}_\lambda\|_{\mathcal{H}}, \|h_{\lambda, k}^*\|_{\mathcal{H}}, \|h_{\lambda, \infty}^*\|_{\mathcal{H}} \right) \leq \frac{\max(\tau, 1 - \tau)}{\lambda}.$$

Idea of proof: Strict convexity of the penalized risk + can restrict the search of the minimizer to a ball of radius $\max(\tau, 1 - \tau)/\lambda$.

Decomposition of the excess asymptotic risk

Theorem

Assume both the regular variation assumption and the regularity assumption hold. Assume that the kernel K associated to the RKHS $\mathcal{H}$ is continuous and satisfies $\sup_{x \in \mathbb{R}^d} K(x, x) \leq 1$. Then,

$$\mathcal{R}_\infty(\hat{h}_\lambda \circ \theta) + \lambda \|\hat{h}_\lambda\|_{\mathcal{H}}^2 - \mathcal{R}_\infty^* \leq B(n, k, \lambda) + \mathcal{A}_2^\infty(\lambda) + D(n, k, \lambda)$$

1. $B(n, k, \lambda) = 2 \sup_{\|h\|_{\mathcal{H}} \leq \max(\tau, 1-\tau)\lambda^{-1}} |\mathcal{R}_{t_{n,k}}(h \circ \theta) - \mathcal{R}_\infty(h \circ \theta)|$
is the **threshold bias**,
2. $\mathcal{A}_2^\infty(\lambda) = \inf_{h \in \mathcal{H}} (\mathcal{R}_\infty(h \circ \theta) + \lambda \|h\|_{\mathcal{H}}^2) - \mathcal{R}_\infty^*$
is the **asymptotic approximation error** of the RKHS,
3. $D(n, k, \lambda) = \mathcal{R}_{t_{n,k}}^{(\lambda)}(\hat{h}_\lambda \circ \theta) - \mathcal{R}_{t_{n,k}}^{(\lambda)}(h_{\lambda,k}^* \circ \theta)$
is a **deviation term**.

Control of the bias term $B(n, k, \lambda)$

Regime of interest:

- ▶ $k(n) \xrightarrow{n \rightarrow \infty} +\infty$ (enough extreme observations)
- ▶ $k(n)/n \xrightarrow{n \rightarrow \infty} 0$ (but not too many!)
- ▶ $\lambda = \lambda(n) \xrightarrow{n \rightarrow \infty} 0$: ensures good approximation in the RKHS.

Control of the bias term $B(n, k, \lambda)$

Regime of interest:

- ▶ $k(n) \xrightarrow{n \rightarrow \infty} +\infty$ (enough extreme observations)
- ▶ $k(n)/n \xrightarrow{n \rightarrow \infty} 0$ (but not too many!)
- ▶ $\lambda = \lambda(n) \xrightarrow{n \rightarrow \infty} 0$: ensures good approximation in the RKHS.

Theorem

1. *Under the previous assumptions, for any fixed $\lambda > 0$, we have, whenever $k(n)/n \rightarrow 0$,*

$$B(n, k(n), \lambda) \xrightarrow{n \rightarrow \infty} 0.$$

2. *Moreover, there exists a sequence $\lambda = \lambda(n)$ with $\lambda(n) \rightarrow 0$ such that*

$$B(n, k(n), \lambda(n)) \xrightarrow{n \rightarrow \infty} 0.$$

Control of the asymptotic approximation error $\mathcal{A}_2^\infty(\lambda)$

Assumption

There exists $c > 0$ and $\beta \in (0, 1]$ such that, for any $\lambda > 0$,

$$\mathcal{A}_2^\infty(\lambda) \leq c\lambda^\beta.$$

Control of the asymptotic approximation error $\mathcal{A}_2^\infty(\lambda)$

Assumption

There exists $c > 0$ and $\beta \in (0, 1]$ such that, for any $\lambda > 0$,

$$\mathcal{A}_2^\infty(\lambda) \leq c\lambda^\beta.$$

- ▶ If the extreme quantile belongs to the RKHS $\mathcal{H}$: $\beta = 1$.
- ▶ This assumption holds in many cases, depending on the regularity of the extreme quantile function and the kernel of the RKHS $\mathcal{H}$ (Eberts and Steinwart, 2012)

Control of the deviation term $D(n, k, \lambda)$

Theorem

With the previous assumptions, there exists a constant $C > 0$ such that, with probability no less than $1 - \delta$,

$$D(n, k, \lambda) \leq \frac{C \log(1/\delta)}{\lambda k}$$

Control of the deviation term $D(n, k, \lambda)$

Theorem

With the previous assumptions, there exists a constant $C > 0$ such that, with probability no less than $1 - \delta$,

$$D(n, k, \lambda) \leq \frac{C \log(1/\delta)}{\lambda k}$$

The control of all the error terms establish **consistency** of the procedure in the regime $k(n) \rightarrow \infty$, $k(n)/n \rightarrow 0$ for a certain sequence $\lambda = \lambda(n)$.

Sketch of proof for the deviation term

- **Step 1:** Leverage the **stability** of SVM solutions: if two probability distribution P_1 and P_2 are "close", then the corresponding SVM solutions h_{λ, P_1} and h_{λ, P_2}

Sketch of proof for the deviation term

- ▶ **Step 1:** Leverage the **stability** of SVM solutions: if two probability distribution P_1 and P_2 are "close", then the corresponding SVM solutions h_{λ, P_1} and h_{λ, P_2}
- ▶ **Step 2** Bound $D(n, k, \lambda)$ by a supremum over a **localized** subset centered in $h_{\lambda, k}^*$:

$$D(n, k, \lambda) \leq \sup_{h \in \mathbb{B}_{\mathcal{H}}(h_{\lambda, k}^*, \rho)} \left| \Delta \mathcal{R}_{t_{n, k}}(h \circ \theta) - \Delta \mathcal{R}_{t_{n, k}}(h_{\lambda, k}^* \circ \theta) \right|$$

where $\Delta \mathcal{R}_{t_{n, k}}(h \circ \theta) = \mathcal{R}_{t_{n, k}}(h \circ \theta) - \widehat{\mathcal{R}}_k(h \circ \theta)$.

Sketch of proof for the deviation term

- ▶ **Step 1:** Leverage the **stability** of SVM solutions: if two probability distribution P_1 and P_2 are "close", then the corresponding SVM solutions h_{λ, P_1} and h_{λ, P_2}
- ▶ **Step 2** Bound $D(n, k, \lambda)$ by a supremum over a **localized** subset centered in $h_{\lambda, k}^*$:

$$D(n, k, \lambda) \leq \sup_{h \in \mathbb{B}_{\mathcal{H}}(h_{\lambda, k}^*, \rho)} \left| \Delta \mathcal{R}_{t_{n, k}}(h \circ \theta) - \Delta \mathcal{R}_{t_{n, k}}(h_{\lambda, k}^* \circ \theta) \right|$$

where $\Delta \mathcal{R}_{t_{n, k}}(h \circ \theta) = \mathcal{R}_{t_{n, k}}(h \circ \theta) - \widehat{\mathcal{R}}_k(h \circ \theta)$.

- ▶ **Step 3:** Use **concentration** tools (Talagrand inequality) to bound the supremum with high probability.

Stability argument: Some results

Lemma

Assume that the kernel K is bounded by 1. Then,

$$\|\hat{h}_\lambda - \tilde{h}_{\lambda,k}\|_{\mathcal{H}} \leq \frac{\max(\tau, 1 - \tau)}{2\lambda k} \left| \sum_{i=1}^n \mathbb{1}_{\|X_i\| \geq t_{n,k}} - k \right|.$$

→ The distance between $\hat{h}_\lambda$ and $\tilde{h}_{\lambda,k}$ is determined by **how well the k -th statistic order estimates the true quantile $t_{n,k}$**

Stability argument: Some results

Lemma

Assume that the kernel K is bounded by 1. Then,

$$\|\hat{h}_\lambda - \tilde{h}_{\lambda,k}\|_{\mathcal{H}} \leq \frac{\max(\tau, 1 - \tau)}{2\lambda k} \left| \sum_{i=1}^n \mathbb{1}_{\|X_i\| \geq t_{n,k}} - k \right|.$$

→ The distance between $\hat{h}_\lambda$ and $\tilde{h}_{\lambda,k}$ is determined by **how well the k -th statistic order estimates the true quantile $t_{n,k}$**

Corollary

If the kernel K is bounded by 1, with probability $1 - \delta$,

$$\|\hat{h}_\lambda - \tilde{h}_{\lambda,k}\|_{\mathcal{H}} \leq \frac{\max(\tau, 1 - \tau)}{\lambda\sqrt{k}} \left(\frac{\sqrt{2\log(1/\delta)}}{2} + \frac{\log(1/\delta)}{6\sqrt{k}} \right).$$

Stability argument: a bound on

Lemma

With probability greater than $1 - \delta$,

$$\|\tilde{h}_{\lambda,k} - h_{\lambda,k}^*\|_{\mathcal{H}} \leq \frac{\max(\tau, 1 - \tau)}{\lambda\sqrt{k}} \left(\sqrt{2 \log(1/\delta)} + 1 + \frac{4 \log(1/\delta)}{3\sqrt{k}} \right).$$

Stability argument: a bound on

Lemma

With probability greater than $1 - \delta$,

$$\|\tilde{h}_{\lambda,k} - h_{\lambda,k}^*\|_{\mathcal{H}} \leq \frac{\max(\tau, 1 - \tau)}{\lambda\sqrt{k}} \left(\sqrt{2 \log(1/\delta)} + 1 + \frac{4 \log(1/\delta)}{3\sqrt{k}} \right).$$

Corollary

With probability $1 - 3\delta$, if the kernel K is bounded by 1,

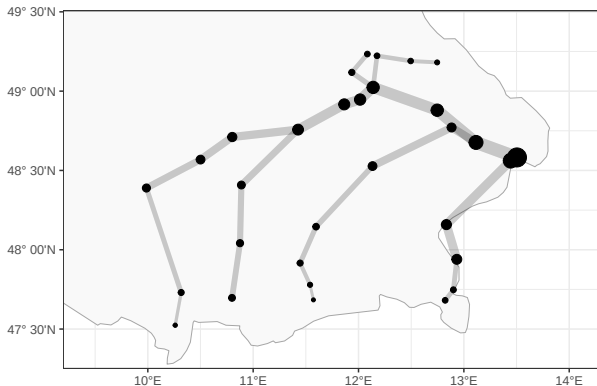
$$\|\hat{h}_{\lambda} - h_{\lambda,k}^*\|_{\mathcal{H}} \leq \frac{\max(\tau, 1 - \tau)}{\lambda\sqrt{k}} \left(1 + \frac{3\sqrt{2 \log(1/\delta)}}{2} + \frac{3 \log(1/\delta)}{2\sqrt{k}} \right)$$

→ The localization argument holds!

Case study on the Danube dataset

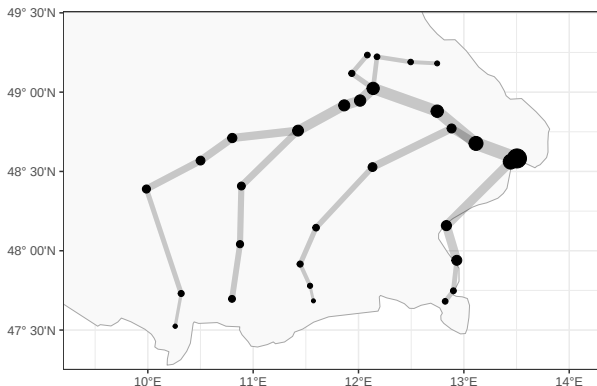
Case study

- ▶ 31 gauging stations measuring a river flow in the Danube



Case study

- ▶ 31 gauging stations measuring a river flow in the Danube



- ▶ Choose one station: Estimation of a conditional quantile for this station given the measurement of the others?
- ▶ Accurate prediction in case of a flood?

Proposed procedure

- ▶ Each observation corresponds to a vector $\tilde{X} \in \mathbb{R}^{31}$.
Goal : Estimate a **missing component** of the vector $\tilde{X}$.

Proposed procedure

- ▶ Each observation corresponds to a vector $\tilde{X} \in \mathbb{R}^{31}$.
Goal : Estimate a **missing component** of the vector $\tilde{X}$.
- ▶ To ensure that the assumptions are verified: We begin by a **standardization to unit Pareto margin** of the vector $\tilde{X}$: $\tilde{X}_j$ the coordinate corresponding to the station j , F_j the corresponding cumulative distribution function, we set

$$V_j = \left(\frac{1}{1 - F_j(\tilde{X}_{ij})} \right)_{i \in \{1, \dots, n\}}.$$

The resulting vector is $V = (X, Z)$.

Proposed procedure

- ▶ Each observation corresponds to a vector $\tilde{X} \in \mathbb{R}^{31}$.
Goal : Estimate a **missing component** of the vector $\tilde{X}$.
- ▶ To ensure that the assumptions are verified: We begin by a **standardization to unit Pareto margin** of the vector $\tilde{X}$: $\tilde{X}_j$ the coordinate corresponding to the station j , F_j the corresponding cumulative distribution function, we set

$$V_j = \left(\frac{1}{1 - F_j(\tilde{X}_{ij})} \right)_{i \in \{1, \dots, n\}}.$$

The resulting vector is $V = (X, Z)$.

- ▶ Proposed procedure:
 1. Set $Y = Z/\|X\|$
 2. Use the extreme quantile regression algorithm to predict a quantile $\hat{q}(x)$ of Y given $X = x$
 3. prediction for Z : $\|x\|\hat{q}(x)$

Our approach versus a naive SVM approach

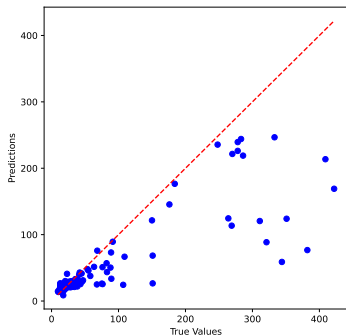
Our approach:

- ▶ Standardize the vector and obtain $\hat{V} = (X, Z)$
- ▶ Set $Y = Z/\|X\|$
- ▶ Learn the conditional quantile function of Y given the value of X .

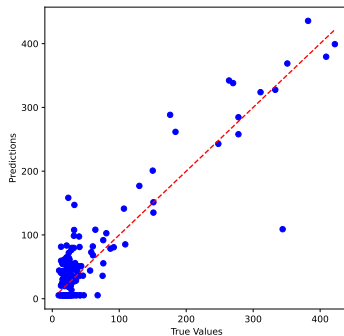
Naive approach:

- ▶ Standardization is still needed to ensure extreme in every direction is explored. Here, each component is centered and scaled with the empirical mean and standard deviation.
- ▶ Standard quantile SVM to learn directly the quantile of the missing component on the most extreme observations.

Results



(a) Baseline approach



(b) Our approach

Figure: Predictions versus true values for the two approaches

ERM vs SVM

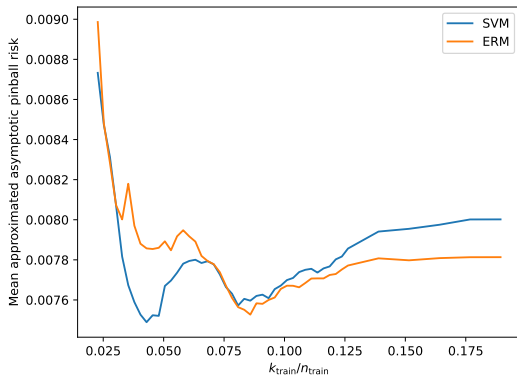


Figure: Bias/variance tradeoff for the choice of k_{train} using and ERM or an SVM approach.

- SVM has better performance for small sample sizes. Especially relevant for extreme quantile regression.

Thank you!

Bibliography I

- Asadi, P., Davison, A. C., and Engelke, S. (2015). Extremes on river networks.
- Buriticá, G. and Engelke, S. (2024). Progression: an extrapolation principle for regression. *arXiv preprint arXiv:2410.23246*.
- Christmann, A. and Steinwart, I. (2008). Support vector machines.
- Cléménçon, S. and Sabourin, A. (2025). Weak signals and heavy tails: Machine-learning meets extreme value theory. *arXiv preprint arXiv:2504.06984*.
- de Carvalho, M., Kumukova, A., and Dos Reis, G. (2022). Regression-type analysis for multivariate extreme values. *Extremes*, 25(4):595–622.
- Eberts, M. and Steinwart, I. (2012). Optimal regression rates for svms using gaussian kernels. *Electronic Journal of Statistics*, 6:2627–2668.

Bibliography II

- Huet, N., Clémentçon, S., and Sabourin, A. (2023). On regression in extreme regions. *arXiv preprint arXiv:2303.03084*.
- Jalalzai, H., Clémentçon, S., and Sabourin, A. (2018). On binary classification in extreme regions. *Advances in Neural Information Processing Systems*, 31.
- Steinwart, I. and Christmann, A. (2011). Estimating conditional quantiles with the help of the pinball loss.