Multiple-output Composite Quantile Regression through an Optimal Transport Lens

Xuzhi Yang, Tengyao Wang

Department of Statistics, LSE

Problem setup

Parameter estimation problem

Linear model: $(X, Y) \in \mathbb{R}^p \times \mathbb{R}^d$ with joint distribution $P^{(X,Y)}$ is generated from

$$Y = b^* X + \varepsilon, \tag{1}$$

with regression coefficient $b^* \in \mathbb{R}^{d\times p}$, $\mathbb{E}X = 0$, and random residue $\varepsilon \in \mathbb{R}^d$ independent with X.

Objective: given $\{(X_i, Y_i)\}_{i=1}^n \overset{\text{iid}}{\sim} P^{(X,Y)}$, consider

- heavy-tailed residue: $\varepsilon \sim P^\varepsilon$ allows ℓ-th moment (for $\ell > 2$);
- multiple-output response: $d \geq 2$.

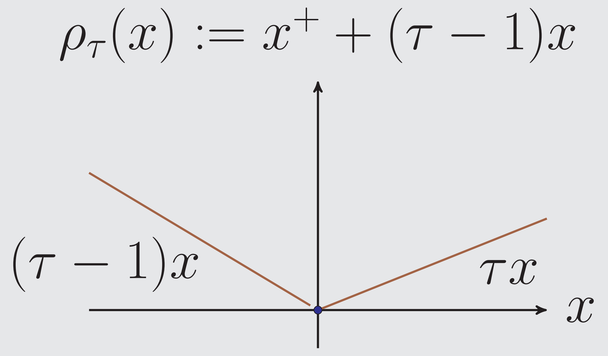
We aim to estimate b^*.

When $d = 1$

Quantile regression: For a *fixed* $\tau \in (0, 1)$, consider

$$(\hat{b}, \hat{q}_\tau) = \arg\min_{\substack{b\in\mathbb{R}^{1\times p}\\ q_\tau\in\mathbb{R}}} \sum_{i=1}^n \rho_\tau(Y_i - bX_i - q_\tau)$$

$$\rho_\tau(x) := x^+ + (\tau - 1)x$$



- Relative efficiency can be arbitrary samll!

Composite quantile regression (CQR): Let $\tau_k = k/(K+1)$, consider aggregated algorithm

$$(\hat{q}_1, \ldots, \hat{q}_K, \hat{b}) = \arg\min_{\substack{q_1,\ldots,q_K\in\mathbb{R}\\ b\in\mathbb{R}^{1\times p}}} \sum_{i=1}^n \sum_{k=1}^K \rho_{\tau_k}(Y_i - bX_i - q_k), \tag{2}$$

When $d \geq 2$: Check function ρ_τ is ill-defined!

Existing methods

- Projection method
- Spatial quantile
- The Monge-Kantorovich quantile

(Projection method, Spatial quantile: Only capture convex support!)

None of them is for robust coefficient estimation!

An optimal transport formulation

CQR in population formula Assume $q_1 \leq \ldots \leq q_K$ in (2), then the corresponding population formula is

$$(b^*, q_\varepsilon^*) \in \arg\min_{\substack{b\in\mathbb{R}^{1\times p}\\ q\in\mathcal{M}}} \mathbb{E}\int_0^1 \rho_\tau(Y - bX - q(\tau))\,\mathrm{d}\tau, \tag{3}$$

where $\mathcal{M}$ denote the set of all increasing functions on $\mathbb{R}$. Let $U \sim \mathrm{Unif}[0, 1]$, we have the following observation:

Start with CQR

$$\begin{aligned}
&\inf_{q\in\mathcal{M}} \mathbb{E}\Big\{\int_0^1 \rho_\tau\big(Y - bX - q(\tau)\big)\,\mathrm{d}\tau\Big\} + \frac{1}{2}\mathbb{E}Y \\
&= \inf_{q\in\mathcal{M}}\Big\{\mathbb{E}\int_0^1 (Y - q(\tau) - bX)^+ \mathrm{d}\tau + \int_0^1 (1-\tau)q(\tau)\mathrm{d}\tau\Big\} \\
&= \inf_{q\in\mathcal{M}}\Big\{\mathbb{E}\max_{t\in[0,1]}\int_0^t (Y - q(\tau) - bX)\,\mathrm{d}\tau + \mathbb{E}\phi(U)\Big\} \\
&= \inf_{\phi\in\mathcal{C}}\Big\{\mathbb{E}\max_{t\in[0,1]}(t(Y - bX) - \phi(t)) + \mathbb{E}\phi(U)\Big\}
\end{aligned}$$

Start with OT

$$\begin{aligned}
&\langle\langle Y - bX, U\rangle\rangle_{\mathcal{W}_2} := \sup_\pi \mathbb{E}(Y - bX)^\top U \\
&= \frac{1}{2}\mathcal{W}_2^2(Y - bX, U) - \frac{1}{2}\mathbb{E}(Y - bX)^2 - \frac{1}{2}\mathbb{E}U^2 \\
&= \inf\Big\{\mathbb{E}\psi(Y - bX) + \mathbb{E}\phi(U) : \phi(x) + \psi(y) \geq xy\Big\} \\
&= \inf\Big\{\mathbb{E}\max_{t\in[0,1]}(t(Y - bX) - \phi(t)) + \mathbb{E}\phi(U) : \phi \in \mathcal{C}\Big\}
\end{aligned}$$

Conclusion: (3) $\Leftrightarrow$ $b^* \in \arg\min_{b\in\mathbb{R}^{1\times p}}\langle\langle Y - bX, U\rangle\rangle_{\mathcal{W}_2}$.

- The OT formula can be extended to the case of $d > 2$ immediately by taking $U \sim \mathrm{Unif}[0, 1]^d$!
- For any $P^\varepsilon, P^U \in \mathcal{P}_2(\mathbb{R}^d) \cap \mathcal{P}_{\mathrm{ac}}(\mathbb{R}^d)$ and P^X is not a point mass, b^* is the unique minimiser, i.e.

$$b^* = \arg\min_{b\in\mathbb{R}^{d\times p}} \mathcal{L}(b), \quad \text{where } \mathcal{L}(b) := \langle\langle Y - bX, U\rangle\rangle_{\mathcal{W}_2}$$

MCQR estimator Given $\{(X_i, Y_i)\}_{i=1}^n$ follows (1), and $\{U_i\}_{i=1}^m \overset{\text{iid}}{\sim} P^U$, the MCQR estimator is defined as

$$\hat{b} \in \arg\min_{b\in\mathbb{R}^{d\times p}} \mathcal{L}_{n,m}(b), \quad \text{where } \mathcal{L}_{n,m}(b) := \langle\langle P_n^{Y-bX}, P_m^U\rangle\rangle_{\mathcal{W}_2}. \tag{4}$$

Theoretical guarantee

Let $P^U \sim \mathcal{N}(0, I_d)$ for theoretical convenience, and assume $P^\varepsilon \in \mathcal{P}_{\mathrm{ac}}(\mathbb{R}^d)$ and P^X is an elliptical distribution.

Case 1: ε with finite ℓ-th moment ($\ell > 2$) Suppose $P^X, P^\varepsilon \in \mathcal{P}_\ell(\mathbb{R}^d)$, then with probability at least $1 - 4(\log n)^{-1}$, the MCQR estimator (4) satisfies

$$\|\hat{b} - b^*\|_\Sigma^2 \wedge 1 \leq C_{d,p}\Big(n^{-\frac{1}{4}} + n^{-\frac{1}{d\vee p}} + n^{-\frac{\ell-2}{2\ell}}\Big)\log m.$$

Case 2: ε with sub-Weibull tail:

- For $\sigma_1, \sigma_2 > 0$ and $\alpha, \beta \in (0, 2]$, $P^{\Sigma^{-1/2}X}$ is (σ_1, α)-sub-Weibull and P^ε is (σ_2, β)-sub-Weibull, i.e.

$$\mathbb{E}\exp\Big\{(\|\Sigma^{-1/2}X\|/\sigma_1)^\alpha/2\Big\} \leq 2 \quad \text{and} \quad \mathbb{E}\exp\Big\{(\|\varepsilon\|/\sigma_2)^\beta/2\Big\} \leq 2$$

- For some $\gamma_1, \gamma_2 > 0$, the density of ε, write as f_ε satisfies $f_\varepsilon(e) \geq \gamma_1\exp(-\gamma_2\|e\|^2)$, for $\|e\| \geq 1$.

Then with probability at least $1 - 33(\log n)^{-1}$, we have

$$\|b^* - \hat{b}\|_\Sigma^2 \leq M_d\Big((p/n)^{1/2} + n^{-2/d}\Big)(\log m)^{\frac{8}{2\wedge\alpha\wedge\beta}}.$$

Proof Sketch

Case 1: Consider the basic inequality

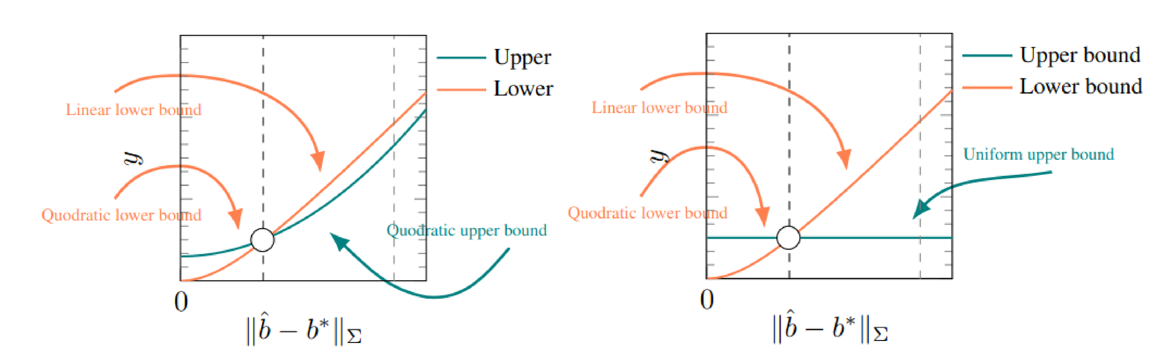
$$\mathcal{L}(\hat{b}) - \mathcal{L}(b^*) \leq \mathcal{L}(\hat{b}) - \mathcal{L}_{n,m}(\hat{b}) + \mathcal{L}_{n,m}(b^*) - \mathcal{L}(b^*).$$

- LHS Lower bound:

Lemma . *Let $Z \perp\!\!\!\perp \varepsilon$ random vectors in $\mathbb{R}^d$ and $U \sim \mathcal{N}(0, I_d)$. If P^ε and P^Z are atomless with finite-second moments, then*

$$\langle\langle Z + \varepsilon, U\rangle\rangle_{\mathcal{W}_2}^2 \geq \langle\langle Z, U\rangle\rangle_{\mathcal{W}_2}^2 + \langle\langle \varepsilon, U\rangle\rangle_{\mathcal{W}_2}^2.$$

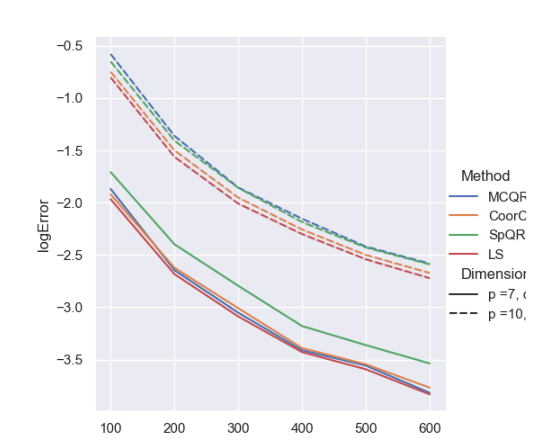
- RHS Upper bound



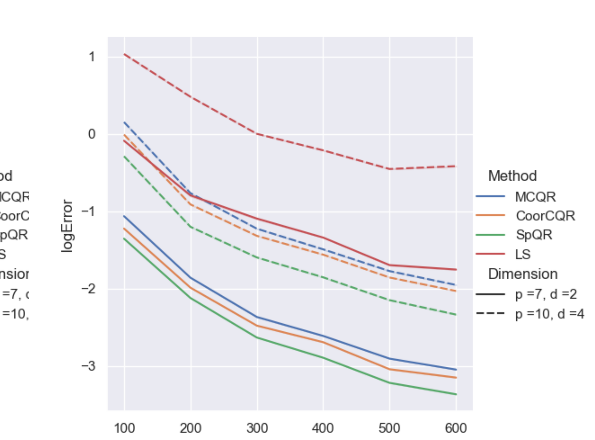
Case 2: Upper bound involves the following *uniform* error bound for empirical 2-Wasserstein distance:

$$\sup_{b\in\mathcal{B}}\Big|\mathcal{W}_2^2(P^{Y-bX}, P^U) - \mathcal{W}_2^2(P_n^{Y-bX}, P_m^U)\Big|.$$

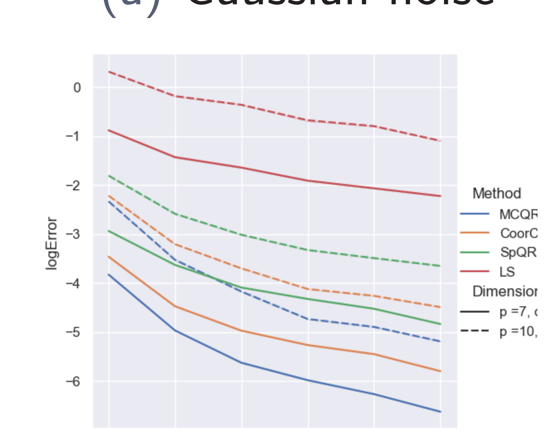
Experiments



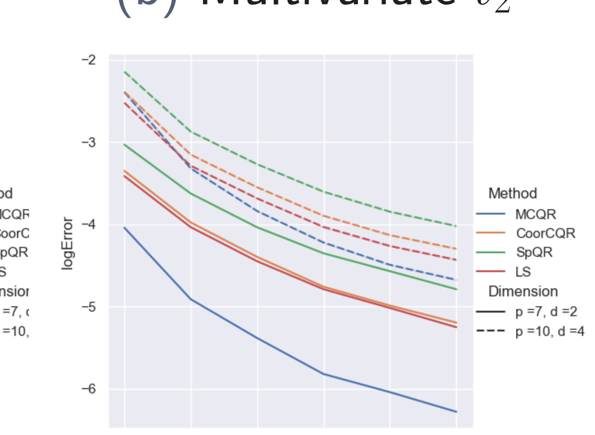
(a) Gaussian noise



(b) Multivariate t_2



(c) Pareto copula



(d) Banana-shaped