# VECTOR QUANTILE REGRESSION

## Computational Optimal Transport

**Delanoue Pierre**

ENS Paris Saclay, MVA Master

January 7, 2020

# Mean vs Quantile



**Breaking out of the dictatorship of the average.**

**Mean**

- No information on the heterogeneity of the data
- Sensitive to extreme values and outliers

**Quantile**

- Distinguishes the impacts on each decile
- Robust

# Quantile Function in $\mathbb{R}$



For $\alpha \in (0, 1)$, the $\alpha$-th quantile of a random variable **y** on $\mathbb{R}$ is defined by:

$$q_{\mathbf{y}}(\alpha) = \inf\{x \in \mathbb{R}, F_{\mathbf{y}}(x) \geq \alpha\}$$

where $F_{\mathbf{y}}$ is the distribution function of **y**.

# Optimal Transport Approach

Optimal transport approach proposed by G. Carlier, V. Chernozhukov and A. Galichon:

[1] G. Carlier, V. Chernozhukov, and A. Galichon. Vector quantile regression: An optimal transport approach. *The Annals of Statistics*, 44(3):1165–1192, 2016.

# Quantile Function Properties

- (i) $\alpha \longmapsto q_{\mathbf{y}}(\alpha)$ is non-decreasing
- (ii) If $U \sim \mathcal{U}([0,1])$, then $q_{\mathbf{y}}(U) = \mathbf{y}$ with probability one.

# Extension Idea



Built a deterministic function $(u, z) \longmapsto Q_{Y|Z}(u, z)$ from $[0, 1]^d \times \mathbb{R}^q$ to $\mathbb{R}^d$ where :

- (I) $(u, z) \longmapsto Q_{Y|Z}(u, z)$ being monotone with respect to u, in the sense of being a gradient of a convex function :

$$\left(Q_{Y|Z}(u, z) - Q_{Y|Z}(u', z)\right)^T (u - u') \geq 0 \qquad \forall (u, u') \in [0, 1]^d \times [0, 1]^d, z \in \mathbb{R}^q \tag{1}$$

- (II) Having with probability one :

$$Y = Q_{Y|Z}(U, Z), \qquad U|Z \sim \mathcal{U}([0, 1]^d) \tag{2}$$

# Linear Model



- Univariate:

$$\forall \alpha \in (0,1),\ \exists \beta_\alpha \in \mathbb{R}^q \qquad q_\alpha(\mathrm{y}|X) = \beta_\alpha^T X \tag{3}$$

- Multivariate:

$$Q_{Y|X}(U,X) = \beta_0(U)^T X, \qquad U|X \sim \mathcal{U}([0,1]^d) \tag{4}$$

where $u \longmapsto \beta(u)$ is a function from $[0,1]^d$ to $\mathbb{R}^{q\times d}$.

# Problem to be solved



$$\max_{U}\{\mathbb{E}[U^T Y] : \; U \sim \mathcal{U}([0,1]^d) \text{ and } \mathbb{E}[X|U] = \mathbb{E}[X]\} \tag{5}$$

**Dual**

$$\inf_{(\psi,b)} \mathbb{E}[\psi(X,Y)] + \mathbb{E}[b(U)]^T \mathbb{E}[X] : \psi(x,y) + b(u)^T x \geq u^T y \qquad \forall (y,x,u) \in \mathbb{R}^d \times \mathbb{R}^q \times \mathbb{R}^d \tag{6}$$

**Solution of Dual Gives**

$$\forall (u,x) \in \mathbb{R}^d \times \mathbb{R}^q, \qquad \beta_0(u)^T x = \nabla_u (b^*(u)^T x) \tag{7}$$

# Discretization

$D_n = \{(Y_1, Z_1), \ldots, (Y_n, Z_n)\}$ and $m$ points $(U_i)_{i \in [\![1,m]\!]}$ of $[0.1]^d$ spaced evenly.

Discrete form of our transportation problem:

$$\max_{P \succeq 0} \sum_{i,j} P_{i,j} Y_j^T U_i \qquad s.t. \qquad P^T \mathbf{1}_m = \nu[\psi],\ PX = \mu\nu^T X[b] \tag{8}$$

where the square brackets indicate the associated Lagrange multiplier.

To find :

$$\widehat{b^*} = \begin{pmatrix} b^*(U_1) \\ \vdots \\ b^*(U_m) \end{pmatrix} = \begin{pmatrix} b_1^*(U_1) \ldots b_q^*(U_1) \\ \vdots \\ b_1^*(U_m) \ldots b_q^*(U_m) \end{pmatrix} \tag{9}$$

# Computation

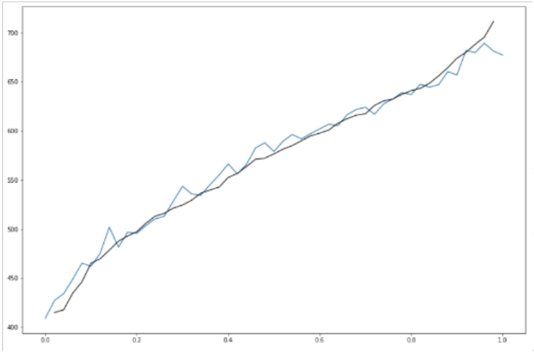


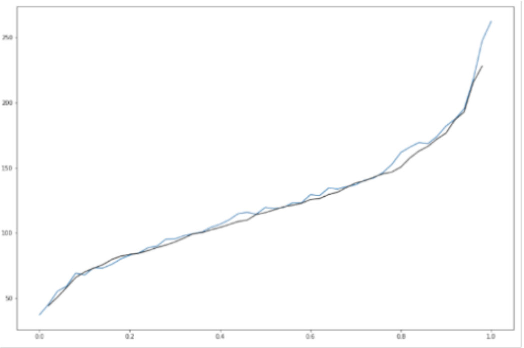
$$\beta_0(u) = \nabla b^*(u) \approx \left(\frac{b_j^*(u^{(i)} + \epsilon, u^{-(i)}) - b_j^*(u^{(i)}, u^{-(i)})}{\epsilon}\right)_{i\in[\![1,d]\!],j\in[\![1,q]\!]} \tag{10}$$

where $u = (u^{(1)}, ..., u^{(d)})$ and $\epsilon > 0$

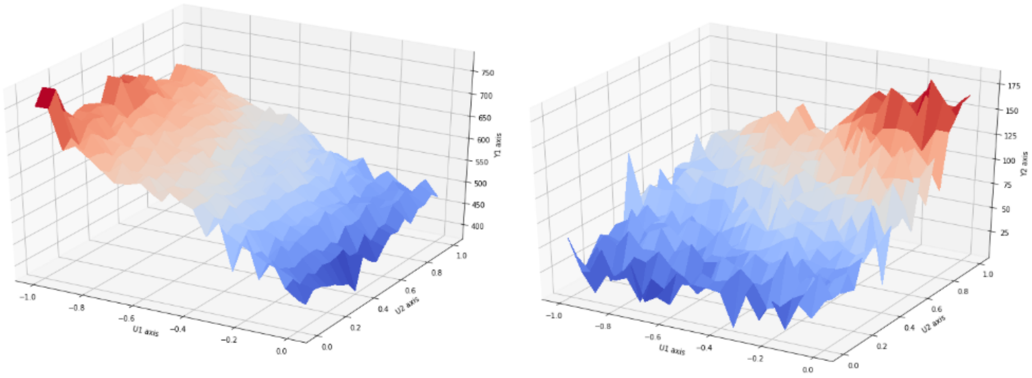
$$\forall i \in [\![1, m]\!], \widehat{\beta}(U_i) := \left(\frac{b_j^*(U_i^{(n:k)}) - b_j^*(U_i)}{\epsilon}\right)_{k\in[\![1,d]\!],j\in[\![1,q]\!]} \tag{11}$$

# Engel's Data: One dimensional Case



# Engel's Data: Two dimensional Case



# Research Perspectives

- Tests and confidence intervals
- Examples of relevant applications in large dimensions
- Overcome the dimension curse

*Thank You*

*for Listening.*