Quantile Regression:
Classical and Vector Approach with Optimal Transport

Delanoue Pierre
ENS Paris-Saclay
MVA Master
pierre.delanoue@ensae.fr

Abstract

This paper explores a proposal to extend quantile regression for vectors where optimal transport results are used. We explain the key ideas of this approach and how it relates to classical quantile regression.
After having detailed the steps of quantile regression vactor calculations, we propose an implementation of this method in the Python language. We present results obtained with the help of this package for Engel's data, which is the classical dataset used to present quantile regression.
All the codes can be found on the GitHub page https://github.com/DatenBiene/Vector_Quantile_Regression

1. Introduction

In this section we present the general framework within which our problem is framed. We then briefly indicate how the method we are studying is relevant to previously proposed approaches. Finally, we introduce some notations that will be used in this paper.

1.1. Presentation of the problem

Modelling and studying the mean is often one of the first steps a statistician or a data scientist will take to gain knowledge from a set of data provided to him. Although intuitive and essential, the mean has several limitations. Among others, the average does not give information on the heterogeneity of the data. For exemple, one cannot efficiently estimate inequalities with the mean. Furthermore, the mean is sensitive to extreme values and outliers.

Less known than linear regression, quantile regression aims at estimating the quantiles of a variable Y conditionally on a variable X. The quantiles have properties that compensate for the limits of the mean that we have mentioned. Quantile regression is thus a powerful tool for studying a data set. For example, a quantile regression can show whether the income of the last decile of a population has increased while the average remains stable. In the context of economic decisions, a quantile regression can be used to estimate the heterogeneous effects of a law or a product change on the different quantiles of the population or its customers.

Quantile regression is gaining increasing interest since it has been brought up to date by the work of Koenker in the late 70's [8]. Koenker's quantile regression has been developed in the frame where Y belongs to a space of dimension 1.

Although desired in many cases, the extension of the quantile definition for Y belonging to $\mathbb{R}^d$ with $d \leq 2$ is non-trivial. Indeed, the different classical definitions of quantiles are based on the fact that $\mathbb{R}$ is ordered.

We therefore wish to study the possibility of extending the notion of quantile regression to the framework where Y belongs to a space of dimension greater than 1. This will allow us to conduct more robust analyses using the joint law of the elements of Y.

1.2. Contributions

A notion of quantile function and quantile regression in dimension greater than 1 using an optimal transport approach has been proposed by G. Carlier, V. Chernozhukov and A. Galichon in [1]. This approach retains two important properties of the classical quantile regression:

- (i) Firstly, this form retains the monotonicity of the quantile function by being in particular the gradient of a convex function.
- (ii) In addition to this, there is the fact that a uniform on the cube $[0.1]^d$ composed by the proposed quantile function has the same law as Y.

Numerous other notions of multivariate quantiles have been proposed (see [3], [6], [11] and [15]). However, none of these proposals has both properties (i) and (ii) at the same time. These two properties are important because they provide a direct link to the classical notion of quantile regression. For more details on these other propositions and different approaches one can refer to chapter 12 of [10] by Marc Hallim.

In the remainder of this paper we will therefore distinguish respectively the conditional classical quantile function and the classical quantile regression, in the univariate case, with the conditional vector quantile function (CVQF) and the vector quantile function (VQF).

It is important to recall certain notions of the classical quantile regression because most of the modeling choices and assumptions made for the vectorized quantile regression are based on results of the classical quantile regression. In addition to the understanding of the construction of quantile regression, recalling the classical quantile regression will allow us to see what are the limits of the optimal transport approach.

1.3. Notations

The reader will note that we will tend to use lowercase letters for the univariate case (y, q_α etc.) and uppercase letters for the vectorized case (Y, Q_α etc.).

We also try to keep notations from the lecture [13] such as:

- For all $n \in \mathbb{N}$, $\mathbf{1}_n \in \mathbb{R}^n$ a vector with only ones.
- For all $(n, m) \in \mathbb{N}^2$, $a \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$, we note $U(a, b)$ the set of admissible coupling for the discrete Kantorovich relaxation such that :

$$U(a, b) := \{P \in \mathbb{R}_+^{n\times m};\ P\mathbf{1}_m = a \ and\ P^T\mathbf{1}_n = b\}$$

- $T\#$ the Push-forward operator.

For a vector $u = (u^{(1)}, ..., u^{(n)}) \in \mathbb{R}^n$, $u^{-(i)}$ represents the vector u from which the i^{th} element has been removed.

2. Univariate and Vector Quantile Functions

2.1. Univariate formulation

The quantile function is easily defined for random variables with values in $\mathbb{R}$:

Definition 2.1 (α-th quantile on $\mathbb{R}$)**.** *For $\alpha \in (0,1)$, the α-th quantile of a random variable $\mathbf{y}$ 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 $\mathbf{y}$. If $F_{\mathbf{y}}$ is strictly increasing we have $q_{\mathbf{y}}(\alpha) = F_{\mathbf{y}}^{-1}(\alpha)$.

Conditional quantile can then naturally be defined by :

$$q_{\mathbf{y}|X}(\alpha) = \inf\{z, F_{\mathbf{y}|X}(z) \geq \alpha\}$$

where conditional quantiles are random variables depending on the random variable X.

It is easy to observe the two important following properties for the quantile function:

- (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.

Proof. When $F_{\mathbf{y}}$ is strictly increasing:

$$\mathbb{P}(q_{\mathbf{y}}(U) \leq t) = \mathbb{P}(F_{\mathbf{y}}^{-1}(U) \leq t) = \mathbb{P}(U \leq F_{\mathbf{y}}(t)) = F_{\mathbf{y}}(t)$$

for all t in $\mathbb{R}$. □

We also recall the crucial property of the classical quantile function :

Property 2.1. *In the unidimentional case where $\mathbf{y}$ is a random variable in $\mathbb{R}$, we have*

$$q_{\mathbf{y}}(\alpha) \in arg\min_{a} \mathbb{E}[\rho_\alpha(Y-a)]$$

where $\rho_\alpha(.)$ is the check function with $\rho_\alpha(u) = (\alpha - 1_{\{u<0\}})u$.

2.2. Vectorized formulation

We are now in the context where Y is a random variable taking values in $\mathbb{R}^d$. We admit that the second moment of Y is finite.

The idea is to 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 we find some equivalent properties as (i) and (ii) in the univariate case. We want to have the following properties:

- (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 :

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

- (II) Having with probability one :

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

In other words, for every $z \in \mathbb{R}^q$ we want to look for a transport T from U to Y:

$$\inf_{T:\mathbb{R}^d \longrightarrow \mathbb{R}^d} \int c(x, T(x))\, F_U(\mathrm{d}x) \qquad T_{\#}F_U = F_Y,\ F_{U|Z=z} = F_{\mathcal{U}([0,1]^d)} \tag{3}$$

where we will choose the cost c to be $c(x, y) = ||x - y||^2$.

Then, we can apply the Brenier's theorem for each $z \in \mathbb{R}^d$ to prove (I).

Theorem 2.1 (Brenier's theorem, Theorem 2.32 from [2]). *Let μ, ν be two probability measures on $\mathbb{R}^d$. Then with probability one*

$$\exists! T \textit{ mesurable},\ T\#\mu = \nu \textit{ and } T = \nabla\phi$$

for some convex function ϕ.

For each $z \in \mathbb{R}^d$ we then find a unique transport T_z. We then define the map $(u, z) \longmapsto Q_{Y|Z}(u, z) = T_z(u)$ which respects the property (I).

The property (II) is proven in [1] showing that the probability law of $(Q_{Y|Z}(U, Z), Z)$ is the same than the law of (Y, Z). It should be noted that Y is not necessarily continuous.

We can reformulate 3 with a probabilistic point of view:

$$\min_U \{\mathbb{E}[||Y - U||^2] :\ U|Z \sim \mathcal{U}([0,1]^d)\} \tag{4}$$

which is equivalent to :

$$\max_U \{\mathbb{E}[U^T Y] :\ U|Z \sim \mathcal{U}([0,1]^d)\} \tag{5}$$

According to [1], the dual of the problem 5 is

$$\min_{(\psi,\phi)} \mathbb{E}(\phi(U, Z) + \psi(Y, Z)) : \phi(u, z) + \psi(y, z) \geq u^T y \qquad \forall (z, y, u) \in \mathbb{R}^p \times \mathbb{R}^d \times \mathbb{R}^d \tag{6}$$

so that for the ϕ solution of 6 gives:

$$(u, z) \longmapsto Q_{Y|Z}(u, z) = \nabla_u \phi(u, z)$$

We can see here that one of the big differences with the univariate case is that we won't be able to compute the quantile function at a single point without computing $Q_{Y|Z}(.,.)$ over its entire definition domain.

3. Regression

Here we are going to choose to set $X = f(Z)$ where f is a known function from $\mathbb{R}^p$ to $\mathbb{R}^q$. We choose f such that the first component of X is an intercept.

For both the univariate and vectorized case, we suppose the models to be linear:

- Univariate:

$$\forall \alpha \in (0, 1),\ \exists \beta_\alpha \in \mathbb{R}^q \qquad q_\alpha(\mathrm{y}|X) = \beta_\alpha^T X \tag{7}$$

- Multivariate:
$$Q_{Y|X}(U,X) = \beta(U)^T X, \qquad U|X \sim \mathcal{U}([0,1]^d) \tag{8}$$
where $u \longmapsto \beta(u)$ is a function from $[0,1]^d$ to $\mathbb{R}^{q\times d}$.

This last condition will actually be relaxed in the multivariate case to have the following condition:

$$Q_{Y|X}(U,X) = \beta(U)^T X, \qquad U \sim \mathcal{U}([0,1]^d) \; and \; \mathbb{E}[X|U] = \mathbb{E}[X] \tag{9}$$

However, 8 and 9 are equivalent if $\forall g \in L^2(F_Z),\ \exists\delta_g,\ g(Z) = X^T\delta_g$.

3.1. Univariate case

Following the linear modeling and the formula of the property 2.1, we then want to solve in the univariate case:

$$\beta_\alpha \in arg\min_\beta \mathbb{E}[\rho_\alpha(\mathrm{y} - X'\beta)] \tag{10}$$

Given a dataset $\{(\mathrm{y}_1, X_1), ..., (\mathrm{y}_n, X_n)\}$, the quantile regression estimator is then naturally built as :

$$\widehat{\beta}_\alpha \in arg\min_\beta \frac{1}{n}\sum_{i=1}^{n} \rho_\alpha(\mathrm{y}_i - X_i'\beta) \tag{11}$$

This estimator possesses several very interesting properties (see [7] for in depth proofs or [5] for insights of the proofs) :

- Identification
- Consistency
- Asymptotic normality
- Enables the building of confidence intervals and statistical tests.

These last two properties are two major advantages of the classical quantile regression. The constitution of statistical tests and confidence intervals provide a lot of information on the data studied.

3.2. Vectorized case

With the condition of linearity 8, we recall that in this framework, where we wish to solve an optimal transport problem, we try to estimate the function $u \longmapsto \beta_0(u)$ from $\mathbb{R}^d$ to the matrix space $\mathbb{R}^{q\times d}$ such that $\forall x \in \mathbb{R}^q$, $u \longmapsto \beta_0(u)^T x$ is monotonous, smooth map, being the gradient of a certain convex function Φ:

$$\forall (u,x) \in \mathbb{R}^d \times \mathbb{R}^q, \qquad \beta_0(u)^T x = \nabla_u \Phi_x(u) \qquad \Phi_x(u) = B_0(u)^T x$$

where $u \longmapsto B_0(u)$ is continuously differentiable form $\mathbb{R}^d$ to $\mathbb{R}^q$. Therefore, finding B_0 would allow us to approximate β_0.

To construct a vector quantile regression model, [1] is based on 5 with a relaxed form of the condition $U|X \sim \mathcal{U}([0,1]^d)$:

$$\max_U \{\mathbb{E}[U^T Y] : \; U \sim \mathcal{U}([0,1]^d) \; and \; \mathbb{E}[X|U] = \mathbb{E}[X]\} \tag{12}$$

Then, under 9, the dual program of 12 is proven by [1] to be the following:

$$\inf_{(\psi,b)} \mathbb{E}[\psi(X,Y)] + \mathbb{E}[b(U)]^T \mathbb{E}[X] : \psi(x,y) + b(u)^T x \geq u^T y \qquad \forall (y,x,u) \in \mathbb{R}^d \times \mathbb{R}^q \times \mathbb{R}^d \tag{13}$$

Carlier, Chernozhukov and Galichon then prove the following result which will allow us to implement a vector quantile estimation method:

Theorem 3.1 (Dual solutions, Theorem 3.2 in [1])**.** *Under the linearity condition, if the second moment of Y and the second moment of U are nite then the solution* (ψ^*, b^*) *of the dual problem 13 meet the following equalities :*

$$b^*(u) = B_0(u) \qquad \psi^*(x,y) = \sup_u \{u^T y - B_0(u)^T x\}$$

More explicitly, thanks to the theorem 3.1 we have the following equality:

$$\forall (u,x) \in \mathbb{R}^d \times \mathbb{R}^q, \qquad \beta_0(u)^T x = \nabla_u \Phi_x(u) \qquad \Phi_x(u) = B_0(u)^T x = b^*(u)^T x \tag{14}$$

thus,

$$\forall (u,x) \in \mathbb{R}^d \times \mathbb{R}^q, \qquad \beta_0(u)^T x = \nabla_u (b^*(u)^T x) \tag{15}$$

This is where one of the big differences between the classical version and the optimal transport approach of quantile regression emerges. There does not seem to be a clear framework within which to create statistical tests of significance of the β parameters. Similarly, it does not seem obvious at first glance to construct confidence intervals for β.

4. Computation

We are now given a dataset $D_n = \{(Y_1, Z_1), ..., (Y_n, Z_n)\}$ of n observations where $Y_1, ..., Y_n$ are iid sample following F_{y} (univariate) or F_Y (vectorized version) and $Z_1, ..., Z_n$ are iid sample following F_Z.

Here we are going to choose to set $X_i = f(Z_i)$ for all i in $[\![1, n]\!]$ where f is the same as defined in the previous section.

We also set X and Y such that:

$$\mathsf{X} = \begin{pmatrix} X_1^T \\ \vdots \\ X_n^T \end{pmatrix} \in \mathbb{R}^{n \times q} \qquad \mathsf{Y} = \begin{pmatrix} Y_1^T \\ \vdots \\ Y_n^T \end{pmatrix} \in \mathbb{R}^{n \times d} \tag{16}$$

4.1. Univariate case

Without any explicit solution to 11, we need to solve the program numerically. Specifically, the check function ρ_α is non differentiable so we canno't use Newton-Raphson algorithms.

However, we can reformulate 11 as a linear programming problem

$$\min_{(\beta,u,v) \in \mathbb{R}^p \times \mathbb{R}_+^{2n}} \alpha \mathbf{1}_n^T u + (1-\alpha) \mathbf{1}_n^T v \qquad s.t. \mathsf{X}\beta + u - v - \mathsf{Y} = 0 \tag{17}$$

where $\mathsf{X} = (X_1, ..., X_n)^T$, $\mathsf{Y} = (Y_1, ..., Y_n)^T$.

In the case where n is small (approximately $n \leq 1000$), we can solve 17 by simplex methods. With n large, we can solve 17 by interior point methods.

We thus directly obtain $\widehat{\beta_\alpha}$ having the properties described in section 3.1.

4.2. Vectorized case

We will estimate the uniform distribution over $[0.1]^d$ by a finite grid of m points $(U_i)_{i\in[\![1,m]\!]}$ of $[0.1]^d$ spaced evenly.

We then set $\nu \in \mathbb{R}^n$ and $\mu \in \mathbb{R}^m$ such that:

$$\nu = \begin{pmatrix} \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{pmatrix} \qquad \mu = \begin{pmatrix} \frac{1}{m} \\ \vdots \\ \frac{1}{m} \end{pmatrix} \tag{18}$$

The discrete form of our transportation problem is therefore:

$$\max_{P\succeq 0} \sum_{i,j} P_{i,j} Y_j^T U_i \qquad s.t. \qquad P^T \mathbf{1}_m = \nu[\psi],\ PX = \mu\nu^T X[b] \tag{19}$$

where the square brackets indicate the associated Lagrange multiplier.

We solve the problem 19 to get $\widehat{b^*}$, the Lagrange multiplier associated with the second equality constraint.

By calculation, we can estimate that $\widehat{b^*}$ satisfies the following equality:

$$\widehat{b^*} = \begin{pmatrix} b^*(U_1) \\ \vdots \\ b^*(U_m) \end{pmatrix} = \begin{pmatrix} b_1^*(U_1) \dots b_q^*(U_1) \\ \vdots \\ b_1^*(U_m) \dots b_q^*(U_m) \end{pmatrix} \tag{20}$$

For convenience we write $\widehat{b^*}_i^{(j)} = b_j^*(U_i)$.

We wish to calculate an estimator of $u \longmapsto \beta_0(u)$.

We recall that :

$$\beta_0(u) = \nabla b^*(u) \approx \left(\frac{b_j^*(u^{(i)} + \epsilon, u^{-(i)}) - b_j^*(u^{(i)}, u^{-(i)})}{\epsilon}\right)_{i\in[\![1,d]\!],j\in[\![1,q]\!]} \tag{21}$$

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

Remember that we constructed $(U_i)_{i\in[\![1,m]\!]} = ((u_i^{(1)}, ..., u_i^{(d)}))_{i\in[\![1,m]\!]}$ so as to form a grid of $[0,1]^d$ with steps equal to ϵ.

We introduce the following notation to express the neighbor of a sample U_i on its dimension k:

$$U_i^{(n:k)} = \begin{cases} (u_i^{(1)}, ..., u_i^{(k)} + \epsilon, ..., u_i^{(d)}) & \text{if } 0 \leq u_i^{(k)} < 1 \\ (u_i^{(1)}, ..., u_i^{(k)} - \epsilon, ..., u_i^{(d)}) & \text{if } u_i^{(k)} = 1 \end{cases}$$

We have the following property on our dataset:

$$\forall (i,k) \in [\![1,m]\!] \times [\![1,d]\!],\ \exists j \in [\![1,m]\!], \qquad U_j = U_i^{(n:k)}$$

Thus we can create an estimator $\widehat{\beta}$ of the function $u \longmapsto \beta_0(u)$ at each point U_i at our disposal while calculating:

$$\forall i \in [\![1,m]\!], \widehat{\beta}(U_i) := (\frac{b_j^*(U_i^{(n:k)}) - b_j^*(U_i)}{\epsilon})_{k \in [\![1,d]\!], j \in [\![1,q]\!]} \tag{22}$$

A linear estimator of the vectorized version of quantile regression was thus highlighted.

It is important to note that several points of this implementation will suffer from the curse of dimentionality.

Indeed, the number m of samples of the Uniform Act $(U_i)_{i \in [\![1,m]\!]}$ increases exponentially with d the dimension of the definition space of Y. It will therefore be important to discuss the computational limitations of such a method.

5. Application

The method presented above is now being put into practice. For this we first briefly present the Python implementation we have done. This implementation is then used on the dataset classically used to study quantile regression: Engel's data on household expenditures used since [9]. The use of these data has been mostly done in the univariate framework whereas the target variable is multivariate. We use this dataset to show that we find similar results with our method in dimension 1 and new results in dimension 2.

5.1. Implementation details

We have fully implemented the vector quantile regression method for the Python language. This is the only Python implementation available on GitHub. There are two other implementations of vector quantile regression on GitHub in Matlab and R respectively. These implementations were developed by the authors of the method. However, these implementations use Gurobi (which requires a license) and do not accept random variables Y with dimensions greater than 2. So it is not easy to use them quickly.

To overcome this, we have created our code using the same conventions used by the scikit-learn package ([12]). Thus the quantile regression is performed in two steps "fit" then "predict" which correspond to two lines of code. In addition, our code automatically accepts as arguments Y of any size and in particular random variables with values in $\mathbb{R}^d$ with $d \leq 2$.

Our implementation uses CVXPY ([4]) to solve its optimization problems. All the codes can be found on the GitHub page `https://github.com/DatenBiene/Vector_Quantile_Regression`

5.2. Engel's Data: One dimensional Case

We wish to use our package implementing the vector quantile regression when the variable Y is of dimension 1. By taking the median wage of our data set, we first look at how the value of the α^{th} quantile evolves as a function of α. We see our results in figure 1.

We also compare our results with the classical quantile regression. For this we use the statsmodel package [14].

We can see in figure 2 that both methods give similar results. Using bootstrap allows statsmodel to have a more stable estimation. This is an improvement prospect for our implementation.

The figures 3 and 4 are the usual representations of the results of a classical regression. We thus represent the evolution of 5 key quantiles (for α equal to 0.1, 0.25, 0.5, 0.75 and 0.9) as a function of the income for respectively food and housing expenses. This representation testifies notably the power of quantile regression against the ordinary last square to study the heterodestacidity of the data.

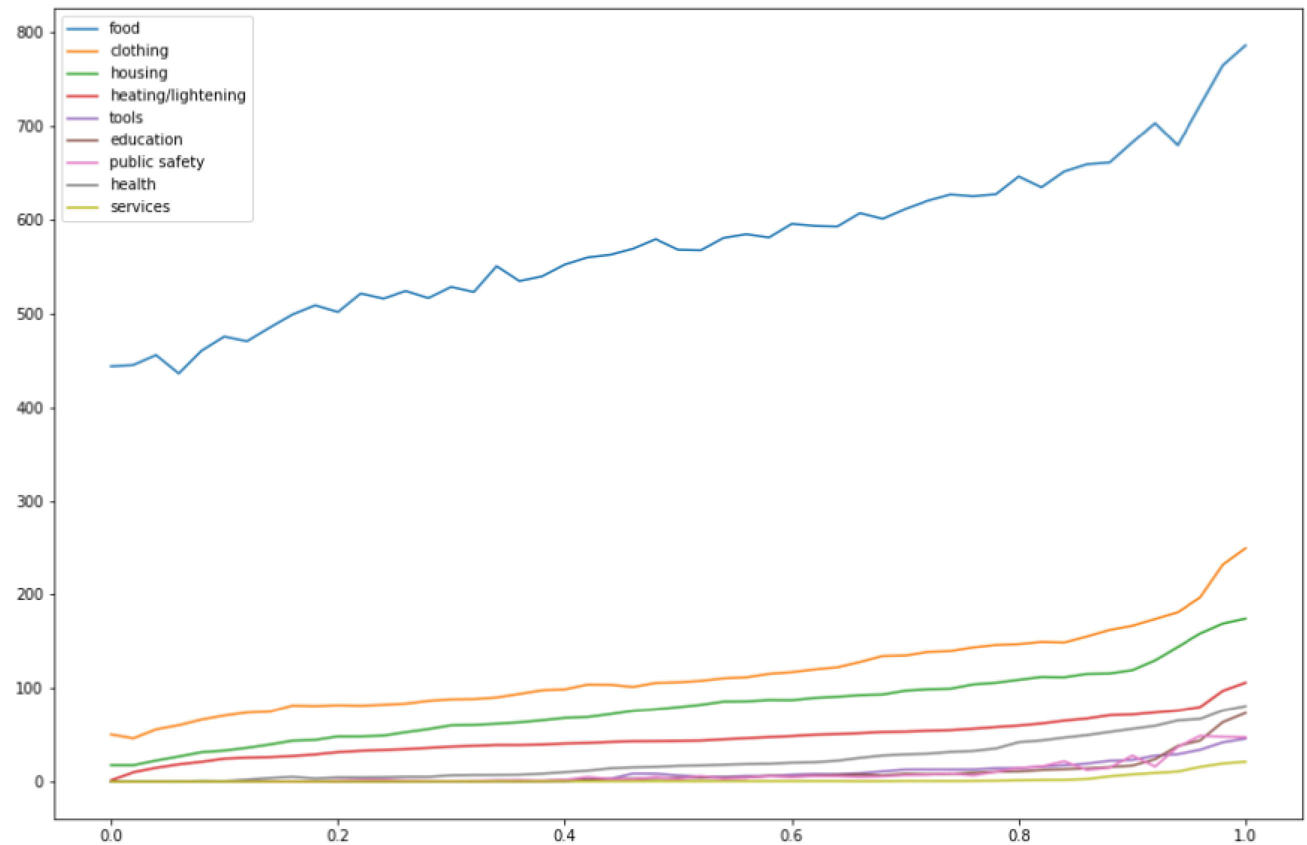

Figure 1. For each dimension of Engel's data observations, evolution of the quantile values for median income X: 883.99

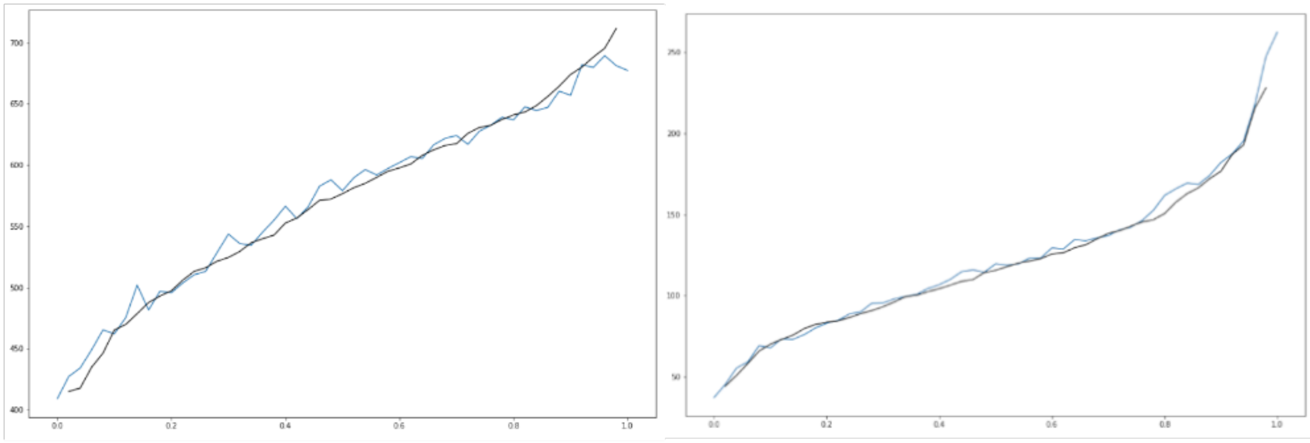

Figure 2. Comparison with classical quantile regression (statsmodel). Left: food expenditure, Right: housing expenditure

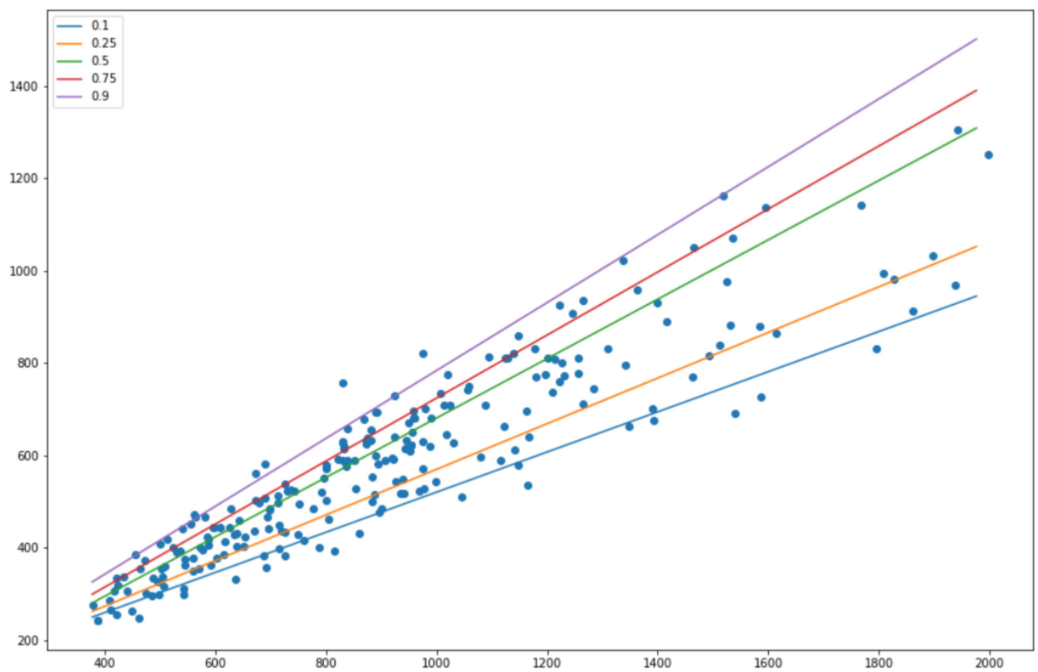

Figure 3. Results of our implementation of quantile regression for food expenses as a function of salary for 5 values of quantiles: 0.1, 0.25, 0.5, 0.75 and 0.9

5.3. Engel's Data: Two dimensional Case

Our implementation now allows us to use quantile regression for vectors of dimension 2. We propose two illustrations of our results on Engel's data.

A first illustration consists in displaying the values of the quantiles of food and housing expenditures as a function of U1 and U2. We find this in the figure 5. As explained in the original paper [1], U1 and U2 can be interpreted

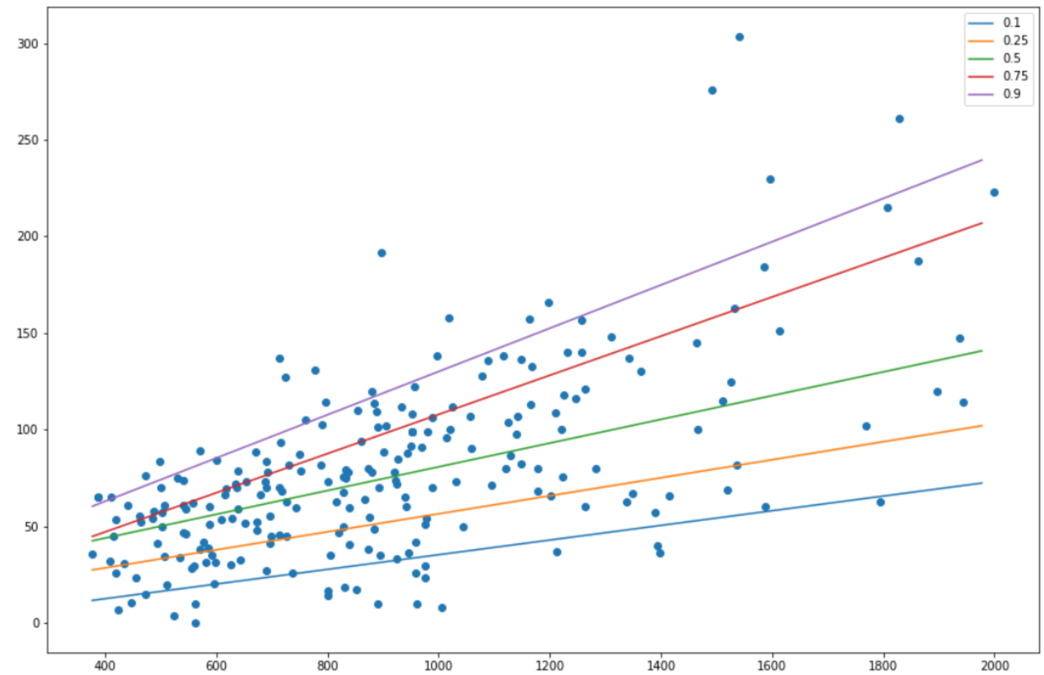

Figure 4. Results of our implementation of quantile regression for housing expenses as a function of salary for 5 values of quantiles: 0.1, 0.25, 0.5, 0.75 and 0.9

as a propensity to consume the good with which they are associated. The interest of figure 5 is to see how Y1 evolves with U2. Indeed we know by construction that Y1 is increasing with U1, but we learn that Y1 covaries strongly with U2. It is the same for Y2 with U1. This shows us that for a median income, these two goods are locally substitutable goods.

We can find this phenomenon in our second illustration with the graph 6. We wish to see the evolution of the consumption of certain goods in the total wage as a function of the evolution of the income. Thus we represent for 9 pairs (U1,U2) the percentage of the salary spent on food in relation to the percentage spent on clothing as a function of the salary. The interpretations are similar to those made in the previous paragraph.

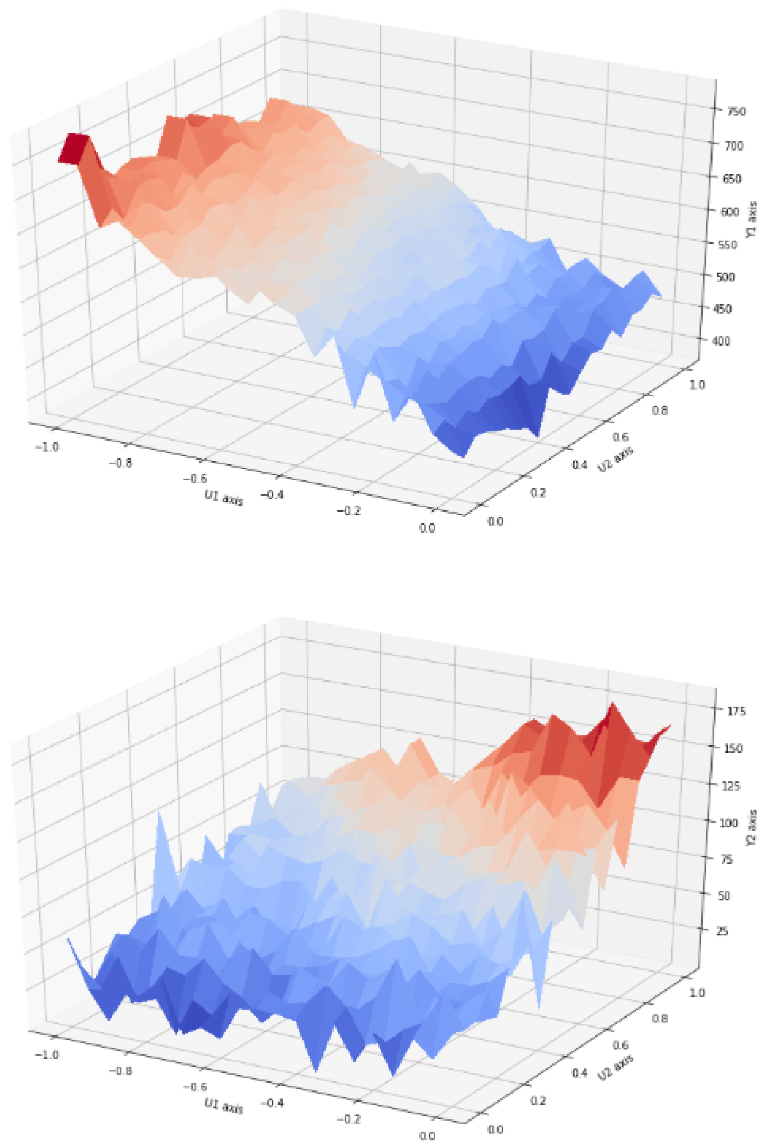

Figure 5. Values predicted by the vectorized quantile regression for X = 883.99 (median value of the Engel's data). The top graph shows food and the bottom graph shows house expenses.

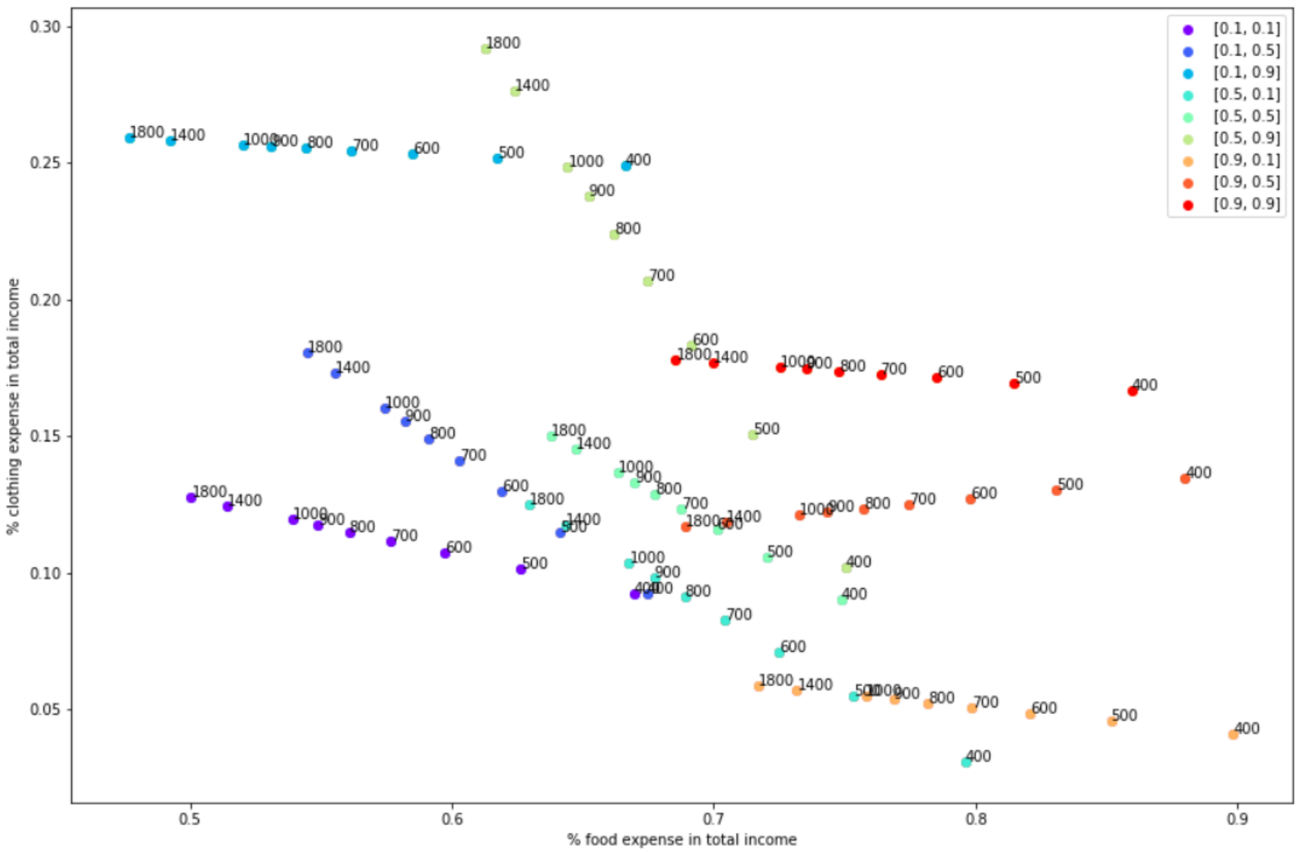

Figure 6. For several pairs of quantiles (left: food, right: clothing) we predict the value of the vector quantile regression for expenditure on food and clothing conditioned by income and we look at the evolution of the share of consumption in relation to income.

6. Conclusion and perspectives

This paper was the occasion to explain the approach of Carlier, Chernozhukov and Galichon to use optimal transport results for the extension of the quantile function to vectors. We insisted on the method to compute a linear estimator in order to make a vector quantile regression.
Our implementation opens perspectives of applications that it would be interesting to use for more advanced econometric surveys.
Many paths of research can be considered. First, for this method to have a real advantage over classical quantitative regression, statistical tests equivalent to those in the univariate case would have to be available. Significance tests and confidence intervals would make it possible to find a crucial approach in the work of an econometrician. In a second step, it is important to question the computational performances of this algorithm. We have seen that this approach suffers from the curse of the dimension. An approach wishing to investigate the use of Sinkhorn's algorithm would need to recalculate new results because the dual would be changed. Finally, it would be valuable if an innovative study could give examples of uses where the vector quantile regression provides results that are crucial for the understanding of a phenomenon.

References

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

[2] V. Cedric. *Topics in optimal transportation*. American mathematical society, 2016.

[3] P. Chaudhuri. On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, 91(434):862–872, 1996.

[4] S. Diamond and S. Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

[5] X. D'Haultfœuille. *Semi and Nonparametric Econometrics, Lectures Note*. 2017.

[6] M. Hallin, D. Paindaveine, and M. Šiman. Multivariate quantiles and multiple-output regression quantiles: From l 1 optimization to halfspace depth. *The Annals of Statistics*, 38(2):635–669, 2010.

[7] R. Koenker. *Quantile Regression*. Cambridge University Press, 2005.

[8] R. Koenker and G. Bassett. Regression quantiles. *Econometrica*, 46(1):33, 1978.

[9] R. Koenker and G. Bassett. Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, 50(1):43, 1982.

[10] R. Koenker, V. Chernozhukov, X. He, and L. Peng. *Handbook of quantile regression*. CRC Press, Taylor Francis Group, 2018.

[11] V. I. Koltchinskii. M -estimation, convexity and quantiles. *The Annals of Statistics*, 25(2):435–477, 1997.

[12] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[13] G. Peyré and M. Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5-6):355–206, 2019.

[14] S. Seabold and J. Perktold. Statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

[15] Y. Wei. An approach to multivariate covariate-dependent quantile contours with application to bivariate conditional growth charts. *Journal of the American Statistical Association*, 103(481):397–409, 2008.