Vector Quantile Regression

Guillaume Carlier (Dauphine)

Victor Chernozhukov (MIT)

Alfred Galichon (ScPo)

Sciences Po,
June 5, 2014

Contents

Introduction

Let Y be a dependent variable (e.g., the random losses of a bank), and X a vector of regressors (e.g. the state of the economy). **Quantile regression** theoretically allows to represent the dependence of Y with respect to X by

$$Y = X^{\intercal}\beta\left(U\right)$$

where, if the model is correctly specified

$$U \sim Unif\left([0,1]\right),\ U \text{ and } X \text{ independent.}$$

A regression procedure (based on linear programming) allows to estimate β.

This model is very useful and convenient as it allows to represent the full conditional distribution – i.e. every conditional quantiles – of Y given X; in contrast with OLS Regression which represents only the conditional mean $E\left[Y|X\right]$.

In other words, the values of $U \in [0,1]$ are interpreted as "scenarios" ranging from $u = 1$ (most pessimistic) and $u = 0$ (most optimistic); if one is able to simulate X, $X^{\intercal}\beta(u)$ will be the value of losses in scenario u, as predicted by X.

If $F_{Y|X}$ is the conditional cdf of Y given X and $Q_{Y|X} = F_{Y|X}^{-1}$ is the corresponding conditional quantile, then if the model is specified,

$$F_{Y|X=x}^{-1}(u) = x^{\intercal}\beta(u).$$

However, Quantile Regression is restrictive in 2 respects:

- Specification of the model requires that U be independent of X – or equivalently, that the conditonal quantiles of Y given X be linear with respect to X, which is a very strong assumption. When this is not the case, $u \rightarrow x^{\intercal}\beta(u)$ is not necessarily increasing ("quantile crossing problem"), and U such that $Y = X^{\intercal}\beta(U)$ is not necessarily defined.

- Y is assumed to be univariate. Obviously there is no natural definition of the quantile in the multivariate case, but it would be of great interest to accomodate for a multivariate dependent variable.

We shall deal with these two concerns. First, let's recall the basics of Quantile Regression.

1 Classical quantile regression, old and new

1.1 Reminders and notations

Quantile: $Q_Y(t) = F_Y^{-1}(t)$

Conditional quantile: $Q_{Y|X=x}(t) = F_{Y|X}^{-1}(t|x)$

Specification of quantile regression:

$$Q_{Y|X=x}(t) = x^{\intercal}\beta(t);$$

Throughout, include a constant regressor $x_1 = 1$.

Estimation:

$$\beta(t) \leftarrow \min_{\beta} E\left[\left(Y - X^{\intercal}\beta\right)^{+}\right] + (1-t)\,\bar{x}^{\intercal}\beta,$$

where $\bar{x} := E[X]$.

By FOC, $E[X1\{Y \geq X^{\intercal}\beta\}] = (1-t)\,\bar{x}$; yields $\beta^{QR}(t)$.

Linear programming formulation. Problem

$$\min_{\beta_t} \mathbb{E}[\,(Y - X^{\intercal}\beta_t)^{+}] + (1-t)\,\bar{x}^{\intercal}\beta_t,$$

has a Linear Programming formulation as

$$\begin{aligned}
&\min_{P \geq 0, \beta} \mathbb{E}[P] + (1-t)\,\bar{x}^{\intercal}\beta_t \\
&s.t.\ Y - X^{\intercal}\beta \leq P \ \ [V_t]
\end{aligned}$$

whose dual is

$$\begin{aligned}
&\max_{V_t \geq 0} \mathbb{E}[V_t Y] \\
&V_t \leq 1 \ \ [P] \\
&\mathbb{E}[V_t X] = (1-t)\,\bar{x} \ \ [\beta_t]
\end{aligned}$$

Note (complementary slackness) that $V_t = 1\{Y \geq X^{\intercal}\beta_t\}$.

1.2 Quantile regression: specified case

Specified case means that $Q_{Y|X=x}(t) = x^{\intercal}\beta(t)$. As a result,

$$t \to x^{\intercal}\beta(t)$$

is increasing for all x in the support, and one can invert $x^{\intercal}\beta(t)$ in t for fixed x, that is, there exists $t(x,y)$ such that

$$x^{\intercal}\beta(t(x,y)) = y.$$

Let $U = t(X,Y)$. Then $Y = X^{\intercal}\beta(U)$, and note that $U = t(X,Y) = F_{Y|X}(Y|X)$, thus the distribution of U conditional on $X = x$ is uniform. Hence, $U \sim U([0,1])$ and (U,X) independent. Therefore QR is specified in an only if one can write

$$\begin{aligned} Y &= X^{\intercal}\beta(U) \\ U &\sim U([0,1]) \text{ and } (U,X) \text{ independent} \end{aligned}$$

1.3 Quantile regression: quasi-specified case

Assume $t \longrightarrow x^{\intercal}\beta^{QR}(t)$ is increasing and continuous for any x in the support. Then, as before, one can invert $x^{\intercal}\beta^{QR}(t)$ in t for fixed x, that is, there exists $t(x,y)$ such that

$$x^{\intercal}\beta^{QR}(t(x,y)) = y.$$

Letting $U = t(X,Y)$, one gets $Y = X^{\intercal}\beta(U)$. We have no longer $t(x,y) = F_{Y|X}(y|x)$; but we have the FOC satisfied by β^{QR}

$$E\left[X1\left\{Y \geq X^{\intercal}\beta^{QR}(t)\right\}\right] = (1-t)\,\bar{x}$$

Replacing Y by $X^{\intercal}\beta^{QR}(U)$, one has

$$E\left[X1\left\{X^{\intercal}\beta^{QR}(U) \geq X^{\intercal}\beta^{QR}(t)\right\}\right] = (1-t)\,\bar{x}$$

but $1\left\{X^{\intercal}\beta^{QR}(U) \geq X^{\intercal}\beta^{QR}(t)\right\} = 1\{U \geq t\}$, so

$$E\left[X1\{U \geq t\}\right] = (1-t)\,\bar{x}$$

therefore, $U \sim U([0,1])$ and X mean-independent (MI) from X, that is $E[X|U] = \bar{x}$. Therefore QR is quasi-specified in an only if one can write

$$\begin{aligned} Y &= X^\intercal \beta(U) \\ U &\sim U([0,1]) \text{ and } E[X|U] = \bar{x} \\ t &\rightarrow x^\intercal \beta^{QR}(t) \text{ nondecr.} \end{aligned}$$

2 Vector quantile and Conditional Vector Quantile

2.1 Revisiting the notion of quantile

Assume in this section that there is no covariate X. Start with the well known fact that for a real random variable $Y \sim \nu$, F_Y its c.d.f.,

$$U = F_Y(Y)$$

is distributed as μ, where μ is the $\mathcal{U}([0,1])$ distribution. Hence, the quantile function $Q_Y = F_Y^{-1}$ satisfies the following axioms:

(A1) For any Y, there is a unique $U \sim \mu$ such that

$$Y = Q(U)$$

(A2) Q is monotone (nondecreasing).

By the **rearrangement inequality** (Hardy-Littlewood), for any $\tilde{U} \sim \mu$, one has

$$\mathbb{E}\left[|U-Y|^2\right] \leq \mathbb{E}\left[\left|\tilde{U}-Y\right|^2\right] \text{ and } \mathbb{E}\left[UY\right] \geq \mathbb{E}\left[\tilde{U}Y\right]$$

Geometrically, this means that U is the projection of Y on the set of random variables with distribution μ; or that U and Y are maximally correlated among all pairs $\left(\tilde{U}, \tilde{Y}\right)$ such that $\tilde{U}$ has the same distribution as U and $\tilde{Y}$ has the same distribution of Y.

Hence the variational problem for the single quantile is

$$\begin{aligned} & \max \mathbb{E}\left[UY\right] \\ s.t. \quad & U \sim \mu \end{aligned}$$

and this is an optimal transportation (Monge-Kantorovich) problem. Its dual can be computed as

$$\begin{aligned} & \min \int \varphi d\mu + \int \psi d\nu \\ s.t. \quad & \varphi\left(u\right) + \psi\left(y\right) \geq uy \end{aligned}$$

whose minimizers φ and ψ are known to exist, to be convex, and to satisfy

$$\psi\left(y\right) = \sup_u \left\{uy - \varphi\left(u\right)\right\}.$$

By complementary slackness, the solutions (U, Y) to the primal problem, and (φ, ψ) to the dual problem are related by

$$Y = \varphi'\left(U\right)$$

that is, $\varphi'\left(u\right) = Q_Y\left(u\right)$ the quantile function of Y. We see that the quantile function is obtained as (the derivative of) the minimizer of the dual variational problem.

This notion generalizes into a notion of Vector Quantile.

2.2 Vector Quantiles

By analogy with the 1D case, when $d \geq 2$ and $\mu = U\left([0,1]^d\right)$, it is natural to consider

$$\begin{aligned} \min \quad & \mathbb{E}\left[\|U - Y\|^2\right] \\ s.t. \quad & U \sim \mu \end{aligned}$$

or equivalently

$$\begin{aligned} & \max \mathbb{E}\left[U^\intercal Y\right] \\ s.t. \quad & U \sim \mu \end{aligned}$$

whose dual is

$$\begin{aligned} & \min \int \varphi d\mu + \int \psi d\nu \\ s.t. \quad & \varphi(u) + \psi(y) \geq u^\intercal y \end{aligned}$$

whose minimizers φ and ψ are known to exist, to be convex, and to satisfy

$$\psi(y) = \sup_u \left\{u^\intercal y - \varphi(u)\right\}.$$

By complementary slackness, the solutions (U, Y) to the primal problem, and (φ, ψ) to the dual problem are related by

$$Y = \nabla \varphi(U)$$

that is, $\nabla \varphi(u) = Q_Y(u)$ the **vector quantile** of Y.

The Vector Quantile is the only map such that:

(A'1) For any Y, there is a unique $U \sim \mu$ such that

$$Y = Q(U)$$

(A'2) Q is monotone (=gradient of a convex function).

2.3 Conditional Vector Quantiles

It is natural to consider

$$
\begin{aligned}
& \max \mathbb{E}\left[UY\right] \\
s.t. \quad & U \sim \mathcal{U}\left([0,1]\right) \\
& (X,Y) \sim \nu \\
& (X,U) \text{ independent.}
\end{aligned}
$$

However, this program rewrites

$$
\begin{aligned}
& \max \int \mathbb{E}\left[UY|X=x\right] d\nu_X\left(x\right) \qquad (1) \\
s.t. \quad & Y|X=x \sim \nu\left(.|x\right) \ \forall x \\
& U|X=x \sim \mathcal{U}\left([0,1]\right) \ \forall x
\end{aligned}
$$

and thus the solution is

$$U = F_{Y|X}\left(Y\right),$$

thus

$$Y = F_{Y|X}^{-1}\left(U\right).$$

This notion generalizes into a notion of Conditional Vector Quantile. Consider

$$\begin{aligned} & \max \mathbb{E}\left[U.Y\right] \\ s.t. \quad & U \sim \mu \\ & (X,Y) \sim \nu \\ & (X,U) \text{ independent.} \end{aligned}$$

whose dual is

$$\begin{aligned} & \inf_{\varphi,\psi} \int \varphi d\mu + \int \psi d\nu \\ & s.t.\ \varphi\left(x,u\right) + \psi\left(x,y\right) \geq u^{\intercal} y \end{aligned}$$

thus by first order conditions in the constraint of the dual, if (U,X,Y) is a solution of the primal and (φ,ψ) is a solution of the dual, then

$$Y = \nabla_u \varphi\left(X,U\right)$$

thus $\nabla_u \varphi\left(x,u\right) = Q_{Y|X=x}\left(u\right)$ is the conditional vector quantile of Y conditional on x and u. Of course, φ has no reason to be linear in u.

3 Vector Quantile Regression

3.1 The LP problems

The previous discussions suggests that if we should hope to have linearity with respect to X, we should look $U \sim \mu$ such that X is *mean-independent* (instead of independent) from U, that is $\mathbb{E}[X|U] = \bar{x}$. Among these choices, it is natural to look for U with maximum correlation with Y. This leads us to

$$\begin{aligned} & \max \mathbb{E}\left[U^{\intercal} Y\right] \\ s.t. \quad & U \sim \mu, \; (X, Y) \sim \nu \\ & \mathbb{E}[X|U] = \bar{x} \end{aligned} \tag{2}$$

which has dual

$$\begin{aligned} & \inf_{b,\psi} \bar{x}' \int b d\mu + \int \psi d\nu \\ & s.t. \; x^{\intercal} b(u) + \psi(x, y) \geq u^{\intercal} y \end{aligned} \tag{3}$$

The constraint in the dual rewrites as

$$\psi(x,y) = \sup_u \{u^\intercal y - x^\intercal b(u)\}$$

hence, by first order conditions (under differentiability of b),

$$Y = X^\intercal \beta^{VQR}(U)$$

where

$$\beta^{VQR}(u) = Db(u).$$

3.2 Connection with classical QR

When $\mathrm{d} = 1$, what is the connection with classical QR?

1. Unsurprisingly, when the model is quasi-specified (and hence, when it is correctly specified),

$$\beta^{VQR} = \beta^{QR}.$$

By complementary slackness, $V_t = 1\left\{Y \geq X^{\intercal}\beta\left(t\right)\right\}$.

2. There is much more. Recall the dual formulation of classical Quantile Regression

$$\begin{aligned}
&\max_{V_t \geq 0} \mathbb{E}[V_t Y] \\
&V_t \leq 1 \;\; [P] \\
&\mathbb{E}[V_t X] = (1 - t)\,\bar{x} \;\; [\beta_t]
\end{aligned}$$

Hence, under quasispecification $t \longrightarrow x^{\intercal}\beta\left(t\right)$ is nondecreasing, thus $t \longrightarrow V_t$ is nonincreasing. However, $t \longrightarrow V_t$

has no reason to be nonincreasing in general. We can thus form the augmented problem, including this constraint:

$$\begin{aligned}
&\max_{V_t \geq 0} \mathbb{E}[V_t Y] \\
&V_t \leq 1 \;\; [P] \\
&\mathbb{E}[V_t X] = (1-t)\,\bar{x} \;\; [\beta_t] \\
&V_t \leq V_s, \;\; t \geq s
\end{aligned}$$

It turns out that this problem is now fully equivalent to VQR, with

$$V_t = 1\{U \geq t\},$$

and the Lagrange multiplier of $\mathbb{E}[V_t X] = (1-t)\,\bar{x}$ is β_t^{VQR}.

3.3 Computation

Sample (X_i, Y_i) of size n. Discretize U into m sample points. Let p be the number of regressors. Program is

$$\begin{array}{l} \max_{\pi \geq 0} Tr(U^\intercal \pi Y) \\ 1_m^\intercal \pi = \nu^\intercal \;\; [\psi^\intercal] \\ \pi X = \mu \bar{x} \;\; [b] \end{array}$$

where X is $n \times p$, Y is $n \times d$, ν is $n \times 1$ such that $\nu_i = 1/n$; U is $m \times d$, μ is $m \times 1$; π is $m \times n$.

To run this optimization problem, need to vectorize matrices. Very easy using Kronecker products. We have

$$\begin{aligned} Tr\left(U^\intercal \pi Y\right) &= vec\left(I_d\right)^\intercal \left(Y \otimes U\right)^\intercal vec\left(\pi\right) \\ vec\left(1_m^\intercal \pi\right) &= \left(I_n \otimes 1_m^\intercal\right) vec\left(\pi\right) \\ vec\left(\pi X\right) &= \left(X^\intercal \otimes I_m\right) vec\left(\pi\right) \end{aligned}$$

Program is implemented in Matlab; optimization phase is done using state-of-the-art LP solver (Gurobi).