Vector Quantile Regression

Alfred Galichon (Sciences Po and MIT)

Joint work with Guillaume Carlier (Dauphine) and Victor Chernozhukov (MIT)

University of Iowa, April 15, 2015

Section 1

INTRODUCTION

- Consider a standard hedonic model (Ekeland, Heckman and Nesheim, Heckman, Nesheim and Matzkin). A consumer of observed characteristics $x \in \mathbb{R}^k$ and latent characteristics $u \in \mathbb{R}$ choosing a good whose quality is a scalar $y \in \mathbb{R}$. Assume utility of consumer choosing y is given by

$$S(x, y) + uy$$

where $S(x, y)$ is the observed part of the consumer surplus, which is assumed to be concave in y, and uy is a preference shock.

- The indirect utility is given by

$$\varphi(x, u) = \max_y \{S(x, y) + uy\}$$

so by first order conditions, $\partial S(x, y) / \partial y + u = 0$, thus, letting $\psi(x, y) = -S(x, y)$, quality y is chosen by consumer $(x, u(x, y))$ such that

$$u(x, y) := \frac{\partial \psi(x, y)}{\partial y}$$

which is nondecreasing in y.

- The econometrician:
  - assumes U is independent from X and postulates the distribution of U (say, $\mathcal{U}([0,1])$)
  - observes the distribution of choices Y given observable characteristics $X = x$.
- Then (Matzkin), by monotonicity of $y(x,u)$ in u, one has

$$\frac{\partial \psi(x,y)}{\partial y} = F_{U|X=x}(u|x)$$

which identifies $\partial_y \psi$, and hence the marginal surplus surplus $\partial_y S(x,y)$.

- The aim of this talk is to:
  - generalize this strategy to vector y
  - obtain a meaningful notion of conditional vector quantile
  - extend quantile regression to the vector case

Section 2

Conditional Vector Quantiles

- Now assume quality is a vector $y \in \mathbb{R}^d$, and latent characteristics is $u \in \mathbb{R}^d$. Assume utility of consumer choosing y is given by

$$S(x, y) + u'y$$

where $S(x, y)$ is still assumed to be concave in y.
- As before, let $\psi(x, y) = -S(x, y)$. By first order conditions, quality y is chosen by consumer $(x, u(x, y))$ such that

$$u(x, y) := \nabla_y \psi(x, y)$$

which, conditional on x, is "vector nondecreasing" in y in a generalized sense, where vector nondecreasing=gradient of a convex function.
- As before, assume:
  - The distribution of U given $X = x$ is μ (say $\mathcal{U}\left([0, 1]^d\right)$)
  - The distribution $F_{Y|X}$ of Y given X is observed.
- **Question**: Is $\nabla_y \psi$ identified as in the scalar case? equivalently, and omitting the dependence in x, is there a convex function $\psi(y)$ such that

$$\nabla \psi(Y) \sim \mu?$$

- The answer, is yes. In fact, ψ is the solution to

$$\begin{aligned} &\min_{\psi,\varphi} \int \psi(y)\, dF_Y(y) + \int \varphi(u)\, d\mu(u) \\ &s.t.\ \psi(y) + \varphi(u) \geq u'y \end{aligned} \tag{1}$$

  which is the Monge-Kantorovich problem.
- This is the "Mass Transportation Approach" (MTA) to identification, applied to a number of contexts by G and Salanié (2012), Chiong, G, and Shum (2014), Bonnet, G, and Shum (2015), Chernozhukov, G, Henry and Pass (2015).
- Problem (1) has a primal formulation which is

$$\begin{aligned} &\max E\left[U'Y\right] \\ &Y \sim F_Y \\ &U \sim \mu \end{aligned} \tag{2}$$

- Fundamental property: both (1) and (2) have solutions, and the solutions are related by

$$U = \nabla\psi(Y) \text{ and } Y = \nabla\varphi(U).$$

- We call the "Vector Quantile" map associated to the distribution of Y (relative to distribution μ) as

$$Q_Y(u) := \nabla \varphi(u)$$

where φ is a solution to (1).

- Q_Y is the unique map which is the gradient of a convex function and which maps distribution μ onto F_Y.

- Now let us go back to the conditional case. We have

$$\min_{\psi,\varphi} \int \psi(x,y)\, dF_{XY}(x,y) + \int \varphi(x,u)\, dF_X(x)\, d\mu(u) \qquad (3)$$
$$s.t.\ \psi(x,y) + \varphi(x,u) \geq u'y$$

which is an infinite-dimensional linear programming problem.

- The functions $\varphi(x,.)$ and $\psi(x,.)$ are conjugate in the sense that

$$\begin{array}{l} \varphi(x,u) = \sup_y \{-\psi(x,y) + u'y\} \\ \psi(x,y) = \sup_u \{-\varphi(x,u) + u'y\} \end{array} \qquad (4)$$

- Problem (1) has a primal formulation which is

$$\begin{array}{l} \max E\left[U'Y\right] \\ (X,Y) \sim F_{XY} \\ U \sim \mu \\ U,\ X \text{ indep.} \end{array} \qquad (5)$$

- Fundamental property: both (1) and (2) have solutions, and the solutions are related by

$$U = \nabla\psi(X,Y) \text{ and } Y = \nabla\varphi(X,U).$$

- We call the "Conditonal Vector Quantile" map associated to the distribution of Y conditional on X (relative to distribution μ) as

$$Q_{Y|X}\left(u|x\right) := \nabla_u \varphi\left(x,u\right)$$

where φ is a solution to (1).

- Q_Y is the unique map which is the gradient of a convex function in u and which maps distribution $F_X \otimes \mu$ onto F_{XY}.

We assume that the following condition holds:

(N) *F_U has a density f_U with respect to the Lebesgue measure on $\mathbb{R}^d$ with a convex support set $\mathcal{U}$.*

(C) *For each $x \in \mathcal{X}$, the distribution $F_{Y|X}(\cdot, x)$ admits a density $f_{Y|X}(\cdot, x)$ with respect to the Lebesgue measure on $\mathbb{R}^d$.*

(M) The second moment of Y and the second moment of U are finite, namely

$$\int\int \|y\|^2 F_{YX}(dy, dx) < \infty \text{ and } \int \|u\|^2 F_U(du) < \infty.$$

DEFINITION

The map $(u, x) \mapsto \nabla_u \varphi(u, x)$ will be called the conditional vector quantile function, namely, denoted $Q_{Y|X}(u, x)$.

THEOREM (CONDITIONAL VECTOR QUANTILES AS OPTIMAL TRANSPORT)

Suppose conditions (N), (C), and (M) hold.
(i) There exists a pair of maps $(u,x) \mapsto \varphi(u,x)$ *and* $(y,x) \mapsto \psi(y,x)$*, each mapping from* $\mathbb{R}^d \times \mathcal{X}$ *to* $\mathbb{R}$*, that solve the problem (1). For each* $x \in \mathcal{X}$*, the maps* $u \mapsto \varphi(u,x)$ *and* $y \mapsto \psi(y,x)$ *are convex and satisfy (**??**).*
(ii) The vector $U = Q^{-1}_{Y|X}(Y,X)$ *is a solution to the primal problem (2) and is unique in the sense that any other solution* U^* *obeys* $U^* = U$ *almost surely. The primal (2) and dual (1) have the same value.*
(iii) The maps $u \mapsto \nabla_u \varphi(u,x)$ *and* $y \mapsto \nabla_y \psi(y,x)$ *are inverses of each other: for each* $x \in \mathcal{X}$*, and for almost every* u *under* F_U *and almost every* y *under* $F_{Y|X}(\cdot, x)$

$$\nabla_y \psi(\nabla_u \varphi(u,x),x) = u, \quad \nabla_u \varphi(\nabla_y \psi(y,x),x) = y.$$

Section 3

Vector Quantile Regression

- We can replace X by $f(X)$ denote a vector of regressors formed as transformations of X, such that the first component of X is 1 (intercept term in the model) and such that conditioning on X is equivalent to conditioning on $f(X)$. The dimension of X is denoted by p and we shall denote $X = (1, X_{-1})$ with $X_{-1} \in \mathbb{R}^{p-1}$. Set $\bar{x} = E[X]$.
- Recall that

$$Q_{Y|X}(u, x) = \nabla_u \varphi(u, x)$$

  thus we would like to impose linearity with respect to X.
- Set $\varphi(u, x) = b(u)^\top x$, so that problem (1) is changed into

$$\begin{aligned} &\min_{\psi, b} \int \psi(x, y)\, dF_{XY}(x, y) + \bar{x}' \int b(u)\, d\mu(u) \\ &s.t.\ \psi(x, y) + x' b(u) \geq u' y \end{aligned} \tag{6}$$

  and as before, we may express ψ as a function of b and get

$$\psi(x, y) = \sup_y \left\{ u' y - x' b(u) \right\}.$$

  whose first order conditions are $y = x' Db(u)$.

- As before, problem (6) has a dual formulation. The corresponding primal formulation is

$$\begin{aligned} &\max E\left[U'Y\right] \\ &(X,Y) \sim F_{XY} \\ &U \sim \mu \\ &E\left[X|U\right] = \bar{x} \end{aligned} \tag{7}$$

- Equivalently,

$$\begin{aligned} &\min E\left[\|U-Y\|^2\right]. \\ &(X,Y) \sim F_{XY} \\ &U \sim \mu \\ &E\left[X|U\right] = \bar{x} \end{aligned} \tag{8}$$

(G) The support of $W = (X_{-1}, Y)$, say $\mathcal{W}$, is a closure of an open bounded convex subset of $\mathbb{R}^{p-1+d}$, the density f_W of W is uniformly bounded from above and does not vanish anywhere on the interior of $\mathcal{W}$. The set $\mathcal{U}$ is a closure of an open bounded convex subset of $\mathbb{R}^d$, and the density f_U is strictly positive over $\mathcal{U}$.

THEOREM

Suppose that condition (G) holds. Then the dual problem (6) admits at least a solution (ψ, B) such that

$$\psi(x, y) = \sup_{u \in \mathcal{U}} \{u^\top y - B(u)^\top x\}.$$

Assume:

(QL) We have a quasi-linear representation a.s.

$$Y = \beta(\widetilde{U})'X, \quad \widetilde{U} \sim F_U, \quad \mathrm{E}(X \mid \widetilde{U}) = \mathrm{E}(X),$$

where $u \mapsto \beta(u)$ is a map from $\mathcal{U}$ to the set $\mathcal{M}_{p\times d}$ of $p \times d$ matrices such that $u \mapsto \beta(u)^\top x$ is a gradient of convex function for each $x \in \mathcal{X}$ and a.e. $u \in \mathcal{U}$:

$$\beta(u)'x = \nabla_u \Phi_x(u), \quad \Phi_x(u) := B(u)\prime x,$$

where $u \mapsto B(u)$ is C^1 map from $\mathcal{U}$ to $\mathbb{R}^d$, and $u \mapsto B(u)^\top x$ is a strictly convex map from $\mathcal{U}$ to $\mathbb{R}$.

This condition allows for a degree of misspecification, which allows for a latent factor representation where the latent factor obeys the relaxed independence constraints.

THEOREM

Suppose conditions (M), (N) , (C), and (QL) hold.
(i) The random vector $\widetilde{U}$ *entering the quasi-linear representation (QL) solves (7).*
(ii) The quasi-linear representation is unique a.s. that is if we also have $Y = \overline{\beta}(\overline{U})'X$ *with* $\overline{U} \sim F_U, \mathrm{E}(X \mid \overline{U}) = \mathrm{E}X$, $u \mapsto X'\overline{\beta}(u)$ *is a gradient of a strictly convex function in* $u \in \mathcal{U}$ *a.s., then* $\bar{U} = \tilde{U}$ *and* $X'\beta(\tilde{U}) = X'\overline{\beta}(\tilde{U})$ *a.s.*

- Assume $d = 1$. What is the connection with classical QR?
- Recall the dual formulation of classical Quantile Regression

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

- When $t \to x^\intercal \beta(t)$ is nondecreasing, thus $t \to A_t$ is nonincreasing. However, in sample, $t \to A_t$ has no reason to be nonincreasing in general. We can thus form the augmented problem, including this constraint:

$$\max_{A_t \geq 0} \int_0^1 \mathbb{E}[A_u Y] du$$
$$A_t \leq 1 \ [P]$$
$$\mathbb{E}[A_t X] = (1-t)\,\bar{x} \ \ [\beta_t]$$
$$A_t \leq A_s, \ t \geq s$$

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

$$U = \int_0^1 A_\tau d\tau$$

$$\psi(x, y) = \int_0^1 \left(y - x^\top \beta_\tau\right)^+ d\tau$$

$$b(t) = \int_0^t \beta_\tau d\tau.$$

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

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

  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

$$Tr\left(U^\intercal \pi Y\right) = vec\left(I_d\right)^\intercal \left(Y \otimes U\right)^\intercal vec\left(\pi\right)$$
$$vec\left(1_m^\intercal \pi\right) = \left(I_n \otimes 1_m^\intercal\right) vec\left(\pi\right)$$
$$vec\left(\pi X\right) = \left(X^\intercal \otimes I_m\right) vec\left(\pi\right)$$

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

11

A. Budgets der Ausgaben mitgetheilt von Ducpetiaux.

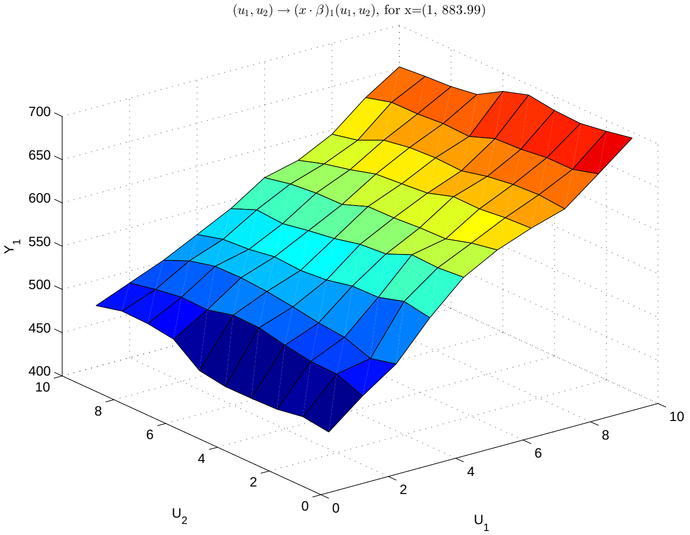
Localitäten. — Bezeichnung der Arbeiter.	Kategorie der Arbeiter-Familien.	Ausgaben für 1. Nahrung.	2. Kleidung.	3. Wohnung.*	4. Heizung und Beleuchtung.	5. Geräthe etc.	6. Erziehung etc.	7. öffentliche Sicherheit etc.	8. Gesundheitspflege etc.	9. persönliche Dienstleistungen.	Ueberhaupt.
		Fr.	Fr.	Fr.	Fr.	Fr.	Fr.	Fr.	Fr.	Fr.	Fr.
I. Provinz Brabant.											
lt Nivelles:											
Tagelöhner	1	576,28	129,34	78,32	46,80	.	.	.	5,20	.	835,94
Weber	2	631,80	83,00	79,80	73,84	.	5,00	.	.	.	873,74
Schieferbrecher	3	608,64	134,00	108,00	78,00	.	5,00	.	17,80	.	951,44

FIGURE : An excerpt from Engel's compilation of Ducpetiaux's data on household expenditure. The total expendable income is broken down into nine categories: 1 Food (Nahrung) 2. Clothing (Kleidung) 3. Housing (Wohnung) 4. Heating and Lighting (Heizung und Beleuchtung) 5. Tools (Geräthe) 6. Education (Erziehung) 7. Public safety (Öffentliche Sicherheit) 8. Health (Gesundheitspflege) 9. Personal service (persönliche Dienstleistungen).

TABLE : Descriptive statistics

	Minimum	Maximum	Median	Average
Food	242,32	2032,68	582,54	624,15
Shelter+Fuel	11,17	660,24	113,36	136,62
Clothing	5,00	520,00	111,76	135,54
Else	0,00	1184,40	39,50	69,21
Total	377,06	4957,83	883,99	982,47

VQR: APPLICATION



VQR: APPLICATION