

OPTIMAL TRANSPORT METHODS IN ECONOMICS: AN INTRODUCTION

Alfred Galichon (NYU Econ+Courant)

Invited lectures at UCLA, Spring 2018
Lecture 2, April 20, 2018

- ▶ Univariate quantiles: properties and uses; quantile regression
- ▶ Rosenblatt's quantiles
- ▶ Vector quantiles; vector quantile regression

Section 1

UNIVARIATE QUANTILES

- ▶ [OTME], Ch. 6.3 and 9.5
- ▶ Hedonic models: Ekeland, Heckman and Nesheim, Heckman, Nesheim and Matzkin
- ▶ Quantile regression: Koenker and Bassett (1978), Koenker (2005)
- ▶ Vector quantiles: Ekeland, G and Henry (2012), Carlier, G and Santambrogio (2010), Carlier, G. and Chernozhukov (2016) Chernozhukov, G, Hallin and Henry (2016).

In dimension one, the following statements equivalently define quantiles of a distribution $Y \sim \nu$:

- ▶ The quantile map is the (generalized) inverse of the cdf of ν : F_ν^{-1} .
- ▶ The quantile map is the nondecreasing map T such that if $U \sim \mathcal{U}([0, 1])$, then $T(U) \sim \nu$.
- ▶ The quantile at t $F_\nu^{-1}(t)$ is the solution of $\min_q \mathbb{E}[\rho_t(Y - q)]$, where $\rho_t(z) = tz^+ + (1 - t)z^-$.
- ▶ The quantile map is the solution to the Monge problem between distribution $\mathcal{U}([0, 1])$ and ν relative to cost $\Phi(u, y) = uy$.

Quantiles have a number of enjoyable properties that make them easy to work with.

- ▶ They fully characterize the distribution ν .
- ▶ They allow to construct a representation of ν : $F_\nu^{-1}(U)$, $U \sim \mu := \mathcal{U}([0, 1])$ has distribution ν .
- ▶ They embed the median ($F_\nu^{-1}(1/2)$) and the extreme values ($F_\nu^{-1}(0)$ and $F_\nu^{-1}(1)$).
- ▶ They allow to provide a construction of distance between distributions: for $p \geq 1$,

$$\left(\int \left| F_\nu^{-1}(t) - F_\mu^{-1}(t) \right|^p dt \right)^{1/p}$$

is the p -Wasserstein distance between μ and ν .

- ▶ They allow for a natural construction of robust statistics by trimming the interval $[0, 1]$.
- ▶ They lend themselves to a natural notion of regression: quantile regression (Koenker and Bassett, 1978; Koenker 2005). (See later).

Quantiles are widely used in economics, finance and statistics.

- ▶ Comonotonicity: $(F_{\nu_1}^{-1}(U), F_{\nu_2}^{-1}(U))$ for $U \sim \mu := \mathcal{U}([0, 1])$ is a comonotone representation of ν_1 and ν_2 .
- ▶ Measures of risk: Value-at-risk $F_{\nu}^{-1}(1 - \alpha)$; CVaR $\int_{1-\alpha}^1 F_{\nu}^{-1}(t) dt$.
- ▶ Non-expected utility: Yaari's rank-dependent EU (Choquet integral) $\int_0^1 F_{\nu}^{-1}(t) w(t) dt$.
- ▶ Demand theory: Matzkin's identification of hedonic models. (See later)
- ▶ Income and inequality: Chamberlain (1994)'s study of the effect of unionization on wages.
- ▶ Biometrics: growth charts.

Section 2

MULTIVARIATE QUANTILES

- ▶ In dimension one, recall the equivalent ways to define a quantile map $Q_\nu : [0, 1] \rightarrow \mathbb{R}$, among which:
 - ▶ the inverse of a cdf: $Q_\nu(t) = F_\nu^{-1}(t) = \inf \{y : F_Y(y) > t\}$;
 - ▶ the minimizer of $\mathbb{E}[\rho_t(Y - q)]$ with $\rho_t(w) = tw^+ + (1 - t)w^-$;
 - ▶ the map T that maximizes $\mathbb{E}[UT(U)]$ subject to $T\#\mathcal{U}([0, 1]) = P$.
- ▶ While equivalent in dimension one, the first two definitions cannot be generalized to the case when Y is multivariate, while the last one can, thanks to Brenier's theorem.
- ▶ The *vector quantile* associated with definition ν is the (unique) gradient of a convex function $Q = \nabla \varphi$ such that if $U \sim \mu = \mathcal{U}([0, 1]^d)$, then $\nabla \varphi(U) \sim \nu$. This concept is rooted in two important results in optimal transport: Brenier's theorem and its generalization, McCann's theorem.

THEOREM (BRENIER)

Assume that μ and ν have finite second moments, and μ has a density. Then the solution $(U, Y) \sim \pi \in \mathcal{M}(\mu, \nu)$ to the primal problem is represented by

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

where (φ, φ^) is a solution to the dual problem. Such φ is unique up to a constant.*

Intuition of the proof: if φ is differentiable, then y is matched with u that maximizes $\{u^\top y - \varphi(u)\}$ over $u \in \mathbb{R}^d$. By first order conditions, such u satisfy $\nabla \varphi(u) = y$. It turns out, however, that differentiability is not a serious concern (at least, almost never).

The previous result allows to provide a representation of a large class of probability distributions ν over \mathbb{R}^d as the probability distribution of $\nabla\varphi(U)$, for U with a fixed distribution μ . There is however a limitation, in the sense that it requires that ν has finite second moments, which is needed for the optimal transport problem to have a finite value. Fortunately, McCann's theorem addresses this issue:

THEOREM (McCANN)

Assume that μ and ν are probability distributions such that μ has a density. Then there is a unique (up to a constant) function u such that

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

holds almost surely with $U \sim \mu$ and $Y \sim \nu$.

The latter is denoted $\nabla\varphi\#\mu = \nu$.

- Definition. Let μ be a distribution on \mathbb{R}^d . For every distribution ν on \mathbb{R}^d , the μ -vector quantile associated with distribution ν is the map

$$Q_\nu(u) = \nabla \varphi(u)$$

where φ is a convex function such that $\nabla \varphi \# \mu = \nu$.

- The existence and uniqueness of this object are provided by McCann's theorem.
- We will take $\mu = \mathcal{U}([0, 1]^d)$. In that case, when $\nu = \nu_1 \otimes \nu_2 \otimes \dots \otimes \nu_d$, we have

$$\varphi(u) = \sum_{i=1}^d \varphi_i(u_i),$$

so $Q_\nu(u) = (\varphi'_1(u_1), \varphi'_2(u_2), \dots, \varphi'_d(u_d)) = (F_{\nu_1}^{-1}(u_1), F_{\nu_2}^{-1}(u_2), \dots, F_{\nu_d}^{-1}(u_d))$.

- Let P be a continuous distributions over \mathbb{R}^2 with finite second moments, and let $\mu = \mathcal{U}([0, 1]^2)$. The *Rosenblatt quantile* of distribution P is defined by

$$\bar{T}(u_1, x_2) = (\bar{T}_1(u_1), \bar{T}_2(u_1, u_2))$$

where \bar{T} is given by

$$\bar{T}_1(u_1, u_2) = F_{Y_1}^{-1}(u_1), \text{ and}$$

$$\bar{T}_2(u_1, u_2) = F_{Y_2|Y_1}^{-1}(u_2 | Y_1 = F_{Y_1}^{-1}(u_1))$$

- The fundamental property of this map is that $\bar{T} \# \mu = P$, and the Jacobian $D\bar{T}$ is lower triangular. The construction extends: the Rosenblatt quantile is the map \bar{T} such that $\bar{T} \# \mu = P$, and such that

$$\begin{cases} Y_1 = T_1(U_1) \\ Y_2 = T_2(U_1, U_2) \\ \dots \\ Y_M = T_M(U_1, U_2, \dots, U_M) \end{cases},$$

has $Y \sim P$ where $U \sim \mu = \mathcal{U}([0, 1]^d)$, and $T_i(u)$ depends only on u_1, \dots, u_i and is a nondecreasing function of u_i .

- For $\lambda > 0$, let $T^\lambda(u)$ be the optimal transport map between μ and P relative to surplus $\Phi^\lambda(u, y) = u_1 y_1 + \lambda u_2 y_2$. One has

$$\bar{T}(u) = \lim_{\lambda \rightarrow 0^+} T^\lambda(u).$$

- Intuition: because $\lambda \rightarrow 0$, the solution will tend to maximize $\mathbb{E}[U_1 Y_1]$ which yields $Y_1 = F_{Y_1}^{-1}(U_1)$, and over set of couplings (U, Y) that verify this relation, will pick those maximizing $\mathbb{E}[U_2 Y_2]$. Thus the $Y_2 = F_{Y_2|Y_1}^{-1}(u_2 | F_{Y_1}^{-1}(u_1))$.
- See a rigorous proof in Carlier, G and Santambrogio <https://arxiv.org/abs/0810.4153>.

Section 3

QUANTILE REGRESSION

- ▶ Hedonic model: A producer of observed characteristics $z \in \mathbb{R}^k$ and latent characteristics $u \in \mathbb{R}$ must choose to produce a good whose quality is a scalar $y \in \mathbb{R}$.
- ▶ The price of a unit of quality y is $p(y)$ (observed) and the cost is $C(z, y)$ (unobserved) so that the profit of choosing quality y is given by

$$p(y) - C(z, y) + uy = -\psi(z, y) + uy$$

where $\psi(z, y) = C(z, y) - p(y)$ is the observed part of minus the profit, which is assumed to be convex in y , and uy is a technology shock (high u 's produce high quality at less cost).

- ▶ The indirect utility is given by

$$\varphi(z, u) = \max_y \{-\psi(z, y) + uy\}$$

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

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

which is nondecreasing in y .

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

$$\frac{\partial \psi(z, y)}{\partial y} = F_{Y|Z}(y|z)$$

which identifies $\partial_y \psi$, and hence the marginal cost $\partial_y C(z, y)$.

- ▶ By the same token,

$$\frac{\partial \varphi(z, u)}{\partial u} = F_{Y|Z}^{-1}(u|z)$$

identifies $\partial_u \varphi(z, u)$ to $F_{Y|Z}^{-1}$.

- ▶ However, the conditional cdf $F_{Y|Z}(y|z)$ or the conditional quantile $F_{Y|Z}^{-1}(u|z)$ are not very easy to estimate nonparametrically. Indeed, the observations are given under the form (Z_i, Y_i) and if Z is continuous, there is not two units i and i' such that $Z_i = Z_{i'}$.
- ▶ Quantile regression therefore adopts a parameterization of the conditional quantile which is linear in Z . That is

$$Q_{Y|Z}(u|z) = z^\top \beta_u$$

(note that one can always augment z with nonlinear functions of z , so this parameterization is quite general).

- ▶ Note that this amounts to taking a linear parameterization of the indirect utility

$$\varphi(z, u) = z^\top b_u \text{ with } b_u = \int_0^u \beta_t dt.$$

- In order to estimate β_u , first note that

$$Q_{Y|Z}(u|z) = \arg \min_q \mathbb{E} [\rho_u(Y - q) | Z = z]$$

where $\rho_u(w) = tw^+ + (1 - t)w^-$.

- Therefore, if the conditional quantile has the specified form, β_u is the solution to

$$\min_{\beta \in \mathbb{R}^k} \mathbb{E} [\rho_u(Y - Z^\top \beta) | Z = z]$$

for each z , and therefore it is the solution to the quantile regression problem introduced by Koenker and Bassett (1978)

$$\min_{\beta \in \mathbb{R}^k} \mathbb{E} [\rho_u(Y - Z^\top \beta)].$$

- Koenker and Bassett showed that this problem has a linear programming formulation. Indeed, consider its sample version

$$\min_{\beta \in \mathbb{R}^k} \sum_{i=1}^n \rho_u(Y_i - Z_i^\top \beta)$$

- Introducing $Y_i - Z_i^\top \beta = P_i - N_i$ with $P_i, N_i \geq 0$, we have

$$\begin{aligned} \min_{\substack{\beta \in \mathbb{R}^k \\ P_i \geq 0, N_i \geq 0}} \quad & \sum_{i=1}^n u P_i + (1 - u) N_i \\ \text{s.t.} \quad & P_i - N_i = Y_i - Z_i^\top \beta \end{aligned}$$

therefore β can be obtained by simple linear programming.

Section 4

VECTOR QUANTILE REGRESSION

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

$$-\psi(z, y) + u'y$$

where $\psi(z, y) = C(z, y) - P(y)$ is still assumed to be convex in y .

- By first order conditions, quality y is chosen by consumer $(z, u(z, y))$ such that

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

which, conditional on z , 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 $Z = z$ is μ (say $\mathcal{U}([0, 1]^d)$)
 - The distribution $F_{Y|Z}$ of Y given Z is observed.

- By Brenier's theorem, for each z , $\psi(z, y)$ and $\varphi(z, u)$ are solution to

$$\begin{aligned} \min_{\psi, \varphi} & \mathbb{E}[\psi(Z, Y)] + \mathbb{E}[\varphi(Z, U)] \\ \text{s.t. } & \psi(z, y) + \varphi(z, u) \geq y^\top u \end{aligned}$$

- The solution potential $\psi(z, y)$ is convex in y and is such that for $(Z, Y) \sim F_{ZY}$,

$$U = \nabla_y \psi(Z, Y) \sim \mu \text{ and is independent from } Z,$$

or equivalently, for $U \sim \mu$ independent from Z , one has

$$(Z, Y = \nabla_u \varphi(Z, U)) \sim F_{ZY}.$$

- 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).

- In vector quantile regression, one would like to get a more parametric way to write down the dependence of Y in Z ; more precisely, linear in Z as in classical quantile regression.
- One way to do this is to set $\varphi(Z, U) = Z^\top b(U)$. The M-K problem becomes

$$\begin{aligned} \min_{\psi, \varphi} & \mathbb{E} [\psi(Z, Y)] + \mathbb{E} [Z^\top b(U)] \\ \text{s.t.} & \psi(z, y) + z^\top b(u) \geq y^\top u \end{aligned}$$

whose primal is

$$\begin{aligned} \max_{U, Z, Y} & \mathbb{E} [U^\top Y] \\ \text{s.t.} & U \sim \mu \\ & (Z, Y) \sim F_{ZY} \\ & \mathbb{E} [Z|U] = \mathbb{E} [Z] \end{aligned}$$

Section 5

CODING

- ▶ Following Koenker, we use Engel's dataset. The package 'quantreg' performs classical quantile regression.

- ▶ Do:

```
thedata = data.frame(X0,Y)
```

```
QRres = rq(Y ~X0, data=thedata, tau = t )  
print(summary(QRres))
```

- Alternatively, we can code it ourselves using Gurobi

$$\begin{aligned} \min_{\substack{\beta \in \mathbb{R}^k \\ P_i \geq 0, N_i \geq 0}} \quad & \sum_{i=1}^n u P_i + (1 - u) N_i \\ \text{s.t.} \quad & (I_n P)_i + (-I_n N)_i + (Z\beta)_i = Y_i \end{aligned}$$

- Do:

```
X=cbind(1,X0)
k=dim(X)[2]
obj=c(rep(t,n),rep(1-t,n),rep(0,k))
A=cbind(sparseMatrix(1:n,1:n),-sparseMatrix(1:n,1:n),X)
result = gurobi
(list(A=A,obj=obj,modelsense="min",rhs=c(Y),lb=c(rep(0,2*n),rep(
thebeta = result$x[(2*n+1):(2*n+k)]
```

- For vector quantile regression, we solve the problem using

```
A1 = kronecker(sparseMatrix(1:n,1:n),matrix(1,1,m))
A2 = kronecker(t(X),sparseMatrix(1:m,1:m))
f1 = matrix(t(nu),nrow=n)
f2 = matrix(mu %*% xbar,nrow=m*r)
e = matrix(1,m*n,1)
A = rbind2(A1,A2)
f = rbind2(f1,f2)
result = gurobi
(list(A=A,obj=c,model sense="min",rhs=f,ub=e,sense="="),
params=NULL )
```