Quantile Regression

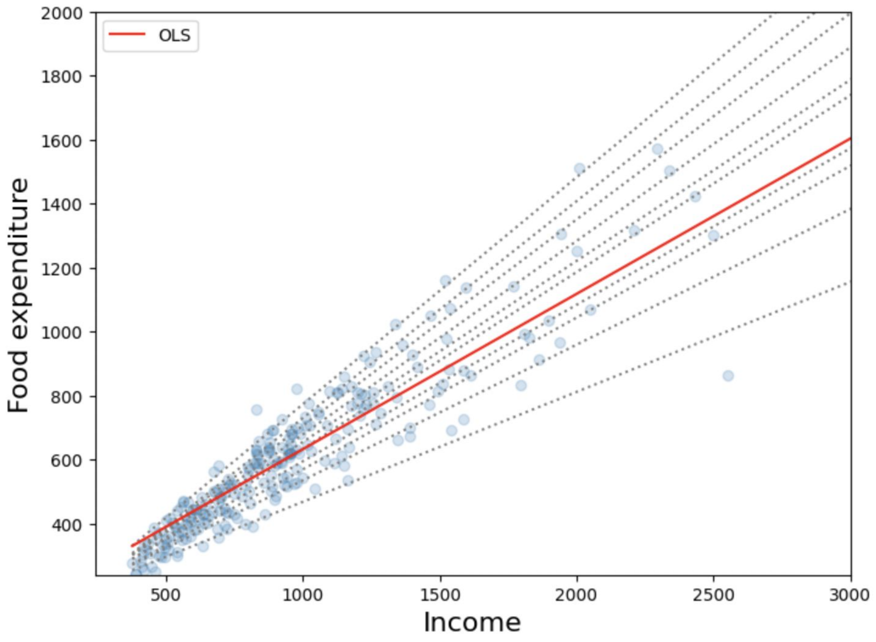
Mentor: Prof. Alfred Galichon

SURE participant: Eric Hu

Motivation

What the regression curve does is give a grand summary for the averages of the distributions corresponding to the set of of x's. We could go further and compute several different regression curves corresponding to the various percentage points of the distributions and thus get a more complete picture of the set.

–Mosteller and Tukey (1977)



Intuition

In OLS, we do not assume independency between predictors X and residual e. (linearity; strict exogeneity; no multicollinearity; spherical errors; normality).

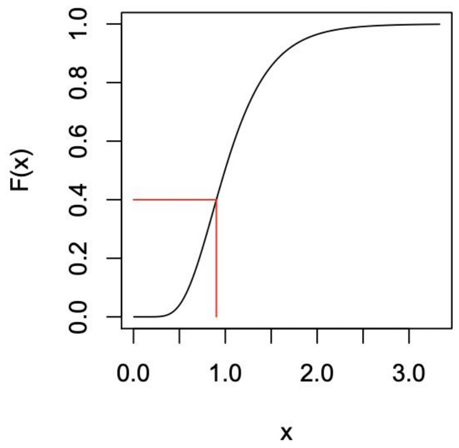
Is there a way to loosen the constraint such that the variable is independent of the error in a certain manner?

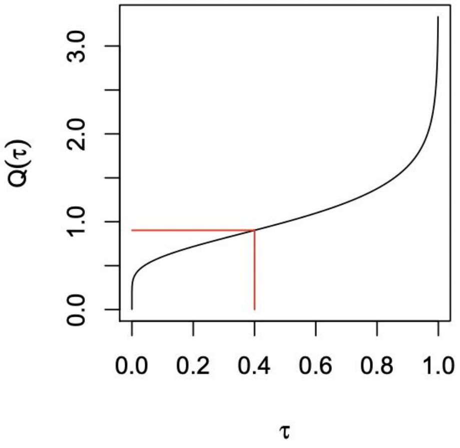
To form this question in a better way, let X be a random variable in R^d, Y be a random variable in R, and X, Y both have continuous distribution (density function exists). Is there a function Q(X, t) such that Q is increasing in t for each X and there exists another random variable U independent from X following a uniform distribution on [0,1] that makes Y = Q(X, U)?

Univariate Quantile

Given a real-valued random variable X, and its cumulative distribution function F, we will define the τth quantile of X as:

$$Q_X(\tau) = F_X^{-1}(\tau) = \inf\{x | F(x) \geq \tau\}$$





Quantile Loss Function

Optimization problem of quantile loss function:

$$\alpha_\tau = \arg\min_{\alpha} \mathbb{E}[\rho_\tau(Y - \alpha)]$$

$$\rho_\tau(u) = \tau u^+ + (1-\tau)u^-,\ u^+ = \max(u, 0),\ u^- = \max(-u, 0)$$

Sketch of the proof (using FOC):

Solve $\mathbb{E}[\frac{d}{d\alpha}\rho_\tau(Y - \alpha)] = 0.$

Substitute with

$$\mathbb{E}[\frac{d}{d\alpha}(Y-\alpha)^-] = \mathbb{E}[\mathbb{1}\{Y < \alpha\}] = F_Y(\alpha)$$

Conditional Quantile

Similarly, the conditional τ th quantile solves:

$$\hat{\alpha_\tau}(x) = \arg\min_{\alpha} \mathbb{E}[\rho_\tau(Y - \alpha)|X = x]$$

Assume linear formation, we have the following parametric form

$$\hat{\beta}_\tau(x) = \arg\min_{\beta} \mathbb{E}[\rho_\tau(Y - X^T\beta)|X = x]$$

$$Q_{Y|X}(\tau|x) = \inf\{y : F(y) \geq \tau|X = x\}$$

$$= \sum_k \beta_k(\tau)x_k = \beta(\tau)^T x$$

Back to Intuition – Specified Quantile Regression

Is there a function Q(X, t) such that Q is increasing in t for each X and there exists another random variable U independent from X following a uniform distribution on [0,1] that makes Y = Q(X, U)? The goal is to correctly define U.

Define U such that

$$Y = Q_{Y|X}(U|X)$$

Recall the inverse relationship

$$\begin{aligned}\mathbb{P}(U < t \mid X = x) &= \mathbb{P}(F_{Y|X}(Y|x) < t \mid X = x) \\ &= \mathbb{P}(Y < Q_{Y|X}(t|x) \mid X = x) \\ &= t\end{aligned}$$

Hence, we obtain a uniform distribution fully independent from X.

$$U = F_{Y|X}(Y|X) = \mu[0,1]$$

Regression Quantiles Computation: Linear Programming

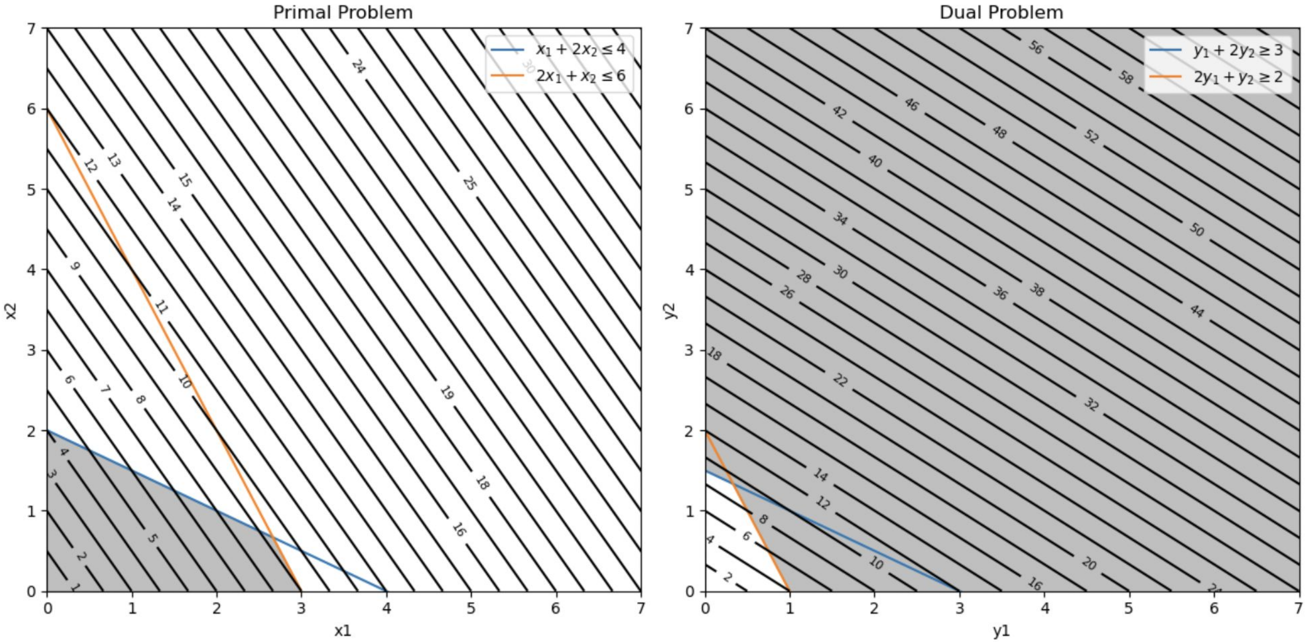
For every linear programming problem (the primal problem), there exists an associated LP problem called its dual problem (there's complementary slackness.)

Primal:

$$\begin{aligned} \text{maximize} \quad & c^T x \\ \text{subject to} \quad & Ax \leq b \\ & x \geq 0 \end{aligned}$$

Dual:

$$\begin{aligned} \text{minimize} \quad & b^T y \\ \text{subject to} \quad & A^T y \geq c \\ & y \geq 0 \end{aligned}$$



Linear Programming - Continued

The goal is to minimize

$$\mathbb{E}[\rho_\tau(Y - X^\top \beta)]$$

We consider its sample version for the ease of coding

$$\arg\min_{\beta \in \mathbb{R}^k} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^\top \beta)$$

Let $u_i = (y_i - x_i^\top \beta)^+$ and $v_i = (y_i - x_i^\top \beta)^-$ with $u_i, v_i \geq 0$. $u_i - v_i = y_i - x_i^\top \beta$.

$$\begin{aligned}
\arg\min_{\beta \in \mathbb{R}^k} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^\top \beta) &= \arg\min_{\beta \in \mathbb{R}^k} \sum_{i=1}^{n} \rho_\tau(u_i - v_i) \\
&= \arg\min_{\beta \in \mathbb{R}^k} \sum_{i=1}^{n} \tau u_i + (1-\tau) v_i \\
&= \arg\min_{\beta \in \mathbb{R}^k} [u^\top \mathbf{1} \tau + v^\top \mathbf{1} (1-\tau)]
\end{aligned}$$

Primal - Dual Problem

Original:

$$\underset{\beta\in\mathbb{R}^k}{\arg\min}[u^\top \mathbf{1}\tau + v^\top \mathbf{1}(1-\tau)]$$

$$s.t.\ y - X^\top\beta - (u-v) = 0,\ (u, v \geq 0)$$

After Lots of Linear Algebra…

Primal: Let $w = (\beta, u, v)$, $c = (0, \mathbf{1}\tau, \mathbf{1}(1-\tau))$, $Z = (X^\top, I_n, -I_n)$

$$\underset{w\in\mathbb{R}^k}{\arg\min}(w^\top c)$$

$$s.t.\ Zw = y$$

Dual:

$$\underset{z}{\arg\max}(y^\top z)$$

$$s.t.\ Xz = (1-\tau)X\mathbf{1}, z \in [0,1]^n$$

Another Way

$$\mathbb{E}[\tau P + (1-\tau)N]$$

$$s.t.\ \ P - N = Y - X^\top\beta$$

where $P = (Y - X^\top\beta)^+$ and $N = (Y - X^\top\beta)^-$ Call - Put Parity
We may rewrite and eliminate N to get the dual problem

Dual →

$$\min_{P\geq 0,\beta} \mathbb{E}\left[P + (1-\tau)X^\top\beta\right]$$

$$s.t.\ \ P + X^\top\beta \geq Y$$

Add a slack variable V to the dual problem

$$\min_{P\geq 0,\beta} \mathbb{E}[P + (1-\tau)X^\top\beta] + \max_{V\geq 0}[V(Y - P - X^\top\beta)]$$

We may rewrite it by combining the min max

$$V_D = \min_{P\geq 0,\beta}\max_{V\geq 0}\mathbb{E}[P + (1-\tau)X^\top\beta + V(Y - P - X^\top\beta)]$$

Using the minimax inequality we have $V_P \geq V_D$ with

$$\begin{aligned} V_P &= \max_{V\geq 0}\min_{P\geq 0,\beta}\mathbb{E}[P + (1-\tau)X^\top\beta + V(Y - P - X^\top\beta)] \\ &= \max_{V\geq 0}\mathbb{E}[VY] + \min_{P\geq 0,\beta}\mathbb{E}[(1-V)P + (1-\tau-V)X^\top\beta + VY] \end{aligned}$$

Primal →

Then, we have obtained the primal problem

$$\max_{V\geq 0}\ \mathbb{E}[YV]$$

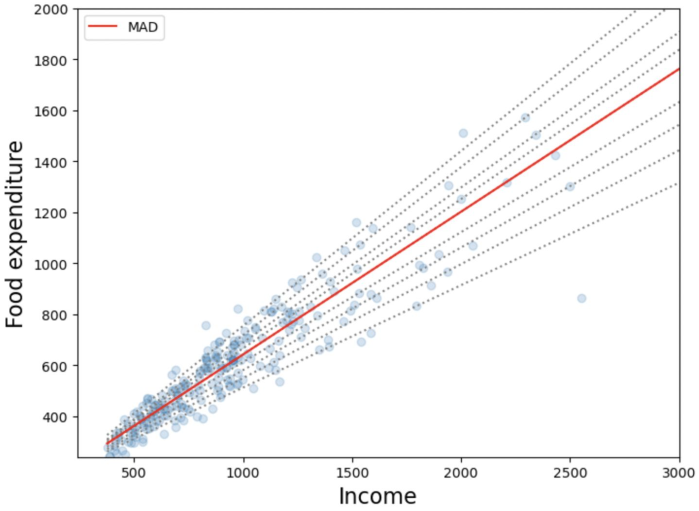
$$s.t.\ \ V \leq 1\ [P_t \geq 0]$$

$$\mathbb{E}[VX] = (1-\tau)\mathbb{E}[X]$$

Realization

Solve the LP problem

using Gurobi




```
income = np.array(engle_data['income'])
food = np.array(engle_data['food'])
housing = np.array(engle_data['housing'])
nbi=len(income)
X_i_k = np.array([np.ones(nbi),income]).T
Y = np.array([food,housing]).T
_,nbk = X_i_k.shape
qr_lp=grb.Model()
τ = 0.5
P = qr_lp.addMVar(shape=nbi, name="P")
β = qr_lp.addMVar(shape=nbk, name="β", lb=-grb.GRB.INFINITY )
qr_lp.setObjective(np.ones(nbi) @ P + (1-τ) * (np.ones(nbi) @ X_i_k) @ β, grb.GRB.MINIMIZE)
qr_lp.addConstr(P  + X_i_k @ β >= food)
qr_lp.optimize()

βhat = qr_lp.getAttr('x')[-nbk:]
βhat
```

```
Gurobi Optimizer version 10.0.2 build v10.0.2rc0 (mac64[arm])

CPU model: Apple M1
Thread count: 8 physical cores, 8 logical processors, using up to 8 threads

Optimize a model with 235 rows, 237 columns and 705 nonzeros
Model fingerprint: 0x2e1ba335
Coefficient statistics:
  Matrix range     [1e+00, 5e+03]
  Objective range  [1e+00, 1e+05]
  Bounds range     [0e+00, 0e+00]
  RHS range        [2e+02, 2e+03]
Presolve time: 0.00s
Presolved: 235 rows, 237 columns, 705 nonzeros

Iteration    Objective       Primal Inf.    Dual Inf.      Time
       0    handle free variables                          0s
     223    8.2118037e+04   0.000000e+00   0.000000e+00      0s

Solved in 223 iterations and 0.01 seconds (0.00 work units)
Optimal objective  8.211803659e+04

[81.48614818083601, 0.560174688436782]
```

Compare the beta values

```
In [17]: import statsmodels.api as sm
         import statsmodels.formula.api as smf
         import matplotlib.pyplot as plt
         import scipy.sparse as spr
```

```
In [18]: τ = 0.5
         model = smf.quantreg('food ~ income', engle_data)
         print(model.fit(q=τ).summary())
```

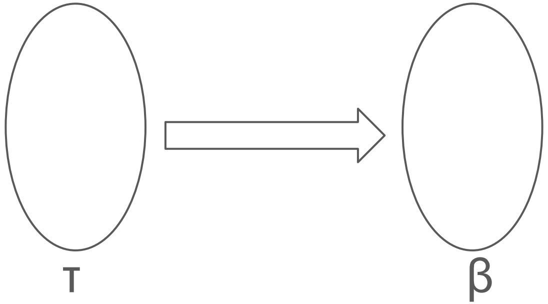
```
                         QuantReg Regression Results
==============================================================================
Dep. Variable:                   food   Pseudo R-squared:               0.6205
Model:                       QuantReg   Bandwidth:                       64.51
Method:                 Least Squares   Sparsity:                        209.3
Date:                Thu, 13 Jul 2023   No. Observations:                  235
Time:                        15:40:38   Df Residuals:                      233
                                        Df Model:                            1
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     81.4862     14.635      5.568      0.000      52.652     110.321
income         0.5602      0.013     42.513      0.000       0.534       0.586
==============================================================================

The condition number is large, 2.38e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Quantile Curve

The previous optimization problem has provided a way to compute β for pointwise values of τ.

If we want to compute the whole curve that maps τ to β, the pointwise method surely makes it an impossible task to finish…

The solution is not difficult, we “integrate” the infinite amount of optimization problem into one!



Continued

Since we know that β_τ solves the primal problem

$$\max_{V_{\tau_i} \in [0,1]} \mathbb{E}[Y V_{\tau_i}]$$

$$s.t.\ \mathbb{E}[V_{\tau_i} X] = (1 - \tau_i)\mathbb{E}[X]\ [\beta]$$

for $\tau_i = \frac{i}{n}$, $0 \leq i \leq n$
Combine them by taking the sum of these problems,

$$\max_{V_{\tau_i} \in [0,1]} \sum_{\tau_i} \mathbb{E}[Y V_{\tau_i}]$$

$$s.t.\ \forall i,\ \mathbb{E}[V_{\tau_i} X] = (1 - \tau_i)\mathbb{E}[X]$$

Taking the limit and we may convert the sum into an integral

Primal →

$$\max_{V(\cdot) \geq 0} \int_0^1 \mathbb{E}[V_\tau Y] d\tau$$

$$s.t.\ V(\tau) \leq 1$$

$$\mathbb{E}[V(\tau) X] = (1 - \tau)\mathbb{E}[X]$$

Similarly, its dual is constructed as

Dual →

$$\min_{P \geq 0, \beta} \int_0^1 \mathbb{E}[P(\tau) + (1 - \tau) X^T \beta(\tau)] d\tau$$

$$s.t.\ P(\tau) \geq Y - X^T \beta(\tau)$$

Monotonicity Constraint

Recall the definition of quantile as the inverse of a CDF function. There is a natural constraint of monotonicity imposed on the dual (Koenker and Ng):

$$\min_{P \geq 0, N \geq 0, \beta} \int_0^1 \mathbb{E}\left[P(\tau) + (1-\tau) X^\top \beta(\tau)\right] d\tau$$

$$\text{s.t. } P(\tau) - N(\tau) = Y - X^\top \beta(\tau)$$

$$\boxed{X^\top \beta(\tau) \geq X^\top \beta\left(\tau'\right), \tau \geq \tau'}$$

To solve it more easily, we consider its primal formulation (Carlier, Chernozhukov and Galichon). Given that

$$V(\tau) = 1\left\{Y \geq X^\top \beta(\tau)\right\},$$

We have $X^\top \beta(\tau)$ nondecreasing in $\tau \Longrightarrow V(\tau)$ nonincreasing.
The primal problem is

$$\max_{V(\tau)} \int_0^1 \mathbb{E}[Y V(\tau)] d\tau$$

$$\text{s.t. } 0 \leq V(\tau) \leq 1$$

$$\mathbb{E}[V(\tau) X] = (1-\tau) \mathbb{E}[X]$$

$$\boxed{V(\tau) \leq V\left(\tau'\right), \tau \geq \tau'}$$

Sampled Version

Assume $\tau_1 = 0 < \ldots < \tau_T \leq 1$ and let $\bar{x}$ be a $1 \times K$ row vector whose k-th entry is $\mathbb{E}[X_k]$. I is the dimension of Y. T is the size of the partition set of τ. The sampled version of the previous primal problem is

$$\max_{V_{ti} \geq 0} \frac{1}{I} \sum_{\substack{1 \leq i \leq I \\ 1 \leq t \leq T}} V_{ti} Y_i$$

$$\begin{aligned} s.t. \quad & V_{ti} \leq 1 \\ & \frac{1}{I} \sum_{1 \leq i \leq I} V_{ti} X_{ik} = (1 - \tau_t)\, \bar{x}_k \\ & V_{(t+1)i} \leq V_{ti} \end{aligned}$$

Since $\tau_1 = 0$, we have $V_{1i} = 1$ from the second constraint. The program becomes

$$\max_{V_{ti}} \frac{1}{I} \sum_{\substack{1 \leq i \leq I \\ 1 \leq t \leq T}} V_{ti} Y_i$$

$$\begin{aligned} s.t. \quad & V_{1i} = 1 \\ & \frac{1}{I} \sum_{1 \leq i \leq I} V_{ti} X_{ik} = (1 - \tau_t)\, \bar{x}_k \\ & V_{1i} \geq V_{2i} \geq \ldots \geq V_{(T-1)i} \geq V_{Ti} \geq 0 \end{aligned}$$

After all these matrix algebra…

We defined a joint probability density, π!

Next, let τ be the $T \times 1$ row matrix with entries τ_k and D be a $T \times T$ matrix defined as

$$D = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \ddots & \vdots & \vdots \\ 0 & -1 & 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 & 0 \\ \vdots & & 0 & -1 & 1 & 0 \\ 0 & & 0 & 0 & -1 & 1 \end{pmatrix}$$

we have $V^\top D \geq 0$ if and only if

$$V_{1i} \geq V_{2i} \geq \ldots \geq V_{(T-1)i} \geq V_{Ti} \geq 0$$

← replace with D

We can write the sampled version into the following matrix form

$$\begin{aligned} & \frac{1}{I} \max_V 1_T^\top V Y \\ s.t. \; & \frac{1}{I} V X = (1_T - \tau)\, \bar{x} \\ & V^\top D 1_T = 1_I \\ & V^\top D \geq 0 \end{aligned}$$

Suppose $\pi = \frac{D^\top V}{I}$ and $U = D^{-1}\mathbf{1}_I$, $\mu = D^\top (1_T - \tau)$ and $p = \frac{\mathbf{1}_I}{I}$. It is equivalent to

$$\begin{aligned} & \max_\pi U^\top \pi Y \\ s.t. \; & \pi X = \mu \bar{x} \\ & \pi^\top 1_T = p \\ & \pi \geq 0 \end{aligned}$$

Assume that the first entry of X is one for the ease of computation. If π satisfies the constraints

$$\sum_{i=1}^{I} \pi_{ti} = \mu_t \text{ and } \sum_{t=1}^{T} \pi_{ti} = p_i$$

then π can be thought of as a joint probability on τ and X given the marginal probability constructed above.

1D Vector Quantile Regression

1D-VQR is equivalent to classical quantile regression by construction!

It can be rewritten into the one-dimensional vector quantile regression construction in the continuous case (Carlier, Chernozhukov and Galichon)

$$
\begin{aligned}
&\max_{\pi} \mathbb{E}_{\pi}[UY] \\
s.t.\ \ & U \sim \mu \\
& (X,Y) \sim P \\
& \boxed{\mathbb{E}[X \mid U] = \mathbb{E}[X]}
\end{aligned}
$$

If we replace the mean-independence between X and U by independence using conditional probability (scalar VQR, by $Thm.3.3$), we have

$$
\begin{aligned}
&\max_{\pi} \mathbb{E}_{\pi}[UY] \\
&\text{s.t.U} \ \sim \mu \\
&(X,Y) \sim P \\
&X \perp\!\!\!\perp U
\end{aligned}
$$

The solution to the latter problem is simply $U = F_{Y|X}(Y \mid X)$ and the nonparametric conditional quantile representation is $Y = F^{-1}_{Y|X}(U \mid X)$. This conforms with the classical quantile function we constructed above!

Computation

```
In [21]: D_t_t = spr.diags([1, -1], [ 0, -1], shape=(nbt, nbt))

         U_t_1 = np.linalg.inv(D_t_t.toarray()) @ np.ones( (nbt,1))
         μ_t_1 = D_t_t.T @ (np.ones((nbt,1)) - τ_t_1)

         A1 = spr.kron(spr.identity(nbt),X_i_k.T)
         A2 = spr.kron(np.array(np.repeat(1,nbt)),spr.identity(nbi))
         A = spr.vstack([A1, A2])
         rhs = np.concatenate( [(μ_t_1 * xbar_1_k).flatten(), np.ones(nbi)/nbi])
         obj = np.kron(U_t_1, Y_i_1).T
         vqr_lp=grb.Model()
         pi = vqr_lp.addMVar(shape=nbi*nbt, name="pi")
         vqr_lp.setParam( 'OutputFlag', False )
         vqr_lp.setObjective( obj @ pi, grb.GRB.MAXIMIZE)
         vqr_lp.addConstr(A @ pi == rhs)
         vqr_lp.optimize()

         ϕ_t_k = np.array(vqr_lp.getAttr('pi'))[0:(nbt*nbk)].reshape((nbt,nbk))

         βvqr_t_k = D_t_t.toarray() @ ϕ_t_k

         βvqr_t_k[10,:]

Out[21]: array([81.48614818,  0.56017469])
```

$$\max_{\pi} U^\top \pi Y$$

$$s.t. \quad \pi X = \mu \bar{x}$$

$$\boxed{\begin{aligned} \pi^\top 1_T &= p \\ \pi &\geq 0 \end{aligned}}$$

Vectorize

$$\begin{pmatrix} I_T \otimes X^\top \\ 1_T^\top \otimes I_I \end{pmatrix} \mathrm{vec}(\pi) = \begin{pmatrix} \mathrm{vec}(\mu\bar{x}) \\ p \end{pmatrix}$$

Vector Quantile Regression - General Case

In some fixed nonatomic probability space, $(\Omega, F, \mathbb{P})$, given a random vector Z with values in $\mathbb{R}^k$ defined on this space, we will denote by $\mathscr{L}(Y)$ the law of Z.

We fix as a reference measure the uniform measure on the unit cube $[0,1]^d$

$$\mu_d := \mathcal{U}\left([0,1]^d\right)$$

Given Y, an integrable $\mathbb{R}^d$-valued random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, Brenier's Theorem states that there exists a unique $U \sim \mu_d$ and a unique convex function defined on $[0,1]^d$ such that

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

The map $\nabla\varphi$ is called the **Brenier's map** between μ_d and $\mathscr{L}(Y)$.
The **vector quantile** of Y is defined to be the Brenier's map between μ_d and $\mathscr{L}(Y)$.
In one-dimensional space, the optimal transport map of Brenier is given by $\nabla\varphi = Q$, where Q is the quantile of Y. Monotonicity persists in both one-dimensional and higher dimensions.

Brenier's Continued

Theorem 2.1. *(Brenier's theorem) If Y is a squared-integrable random vector valued in $\mathbb{R}^d$, there is a unique map of the form $T = \nabla\varphi$ with φ convex on $[0,1]^d$ such that $\nabla\varphi_{\#}\mu = \mathrm{Law}(Y)$, this map is by definition the vector quantile function of Y.*

More topics to explore:

Regularized Vector Quantile Regression

From Quasi-specified QR to Unspecified QR

…

Reference

Carlier, G., Chernozhukov, V., De Bie, G., Galichon, A. (2020). *Vector Quantile Regression and Optimal Transport, from Theory to Numerics.* Empirical Economics. https://doi.org/10.48550/arXiv.2102.12809

Carlier, G., Chernozhukov, V., De Bie, G., Galichon, A. (2016). *Vector quantile regression beyond correct specification.* https://doi.org/10.48550/arXiv.1610.06833

Carlier, G., Chernozhukov, V., De Bie, G., Galichon, A. (2014). *Vector Quantile Regression: An Optimal Transport Approach* https://doi.org/10.48550/arXiv.1406.4643

R. Koenker, G. Bassett. (1978). *Regression Quantiles,* Econometrica, 46,33-50 http://links.jstor.org/sici?sici=0012-9682%28197801%2946%3A1%3C33%3ARQ%3E2.0.CO%3B2-J