HOMEWORK 5: QUANTILE REGRESSION

PAUL SCHRIMPF
DUE: MONDAY, NOVEMBER 21ST
UNIVERSITY OF BRITISH COLUMBIA
ECONOMICS 628: TOPICS IN ECONOMETRICS

In this assignment, you will analyze Graddy's data on the Fulton fish market using quantile regression.

Part 1. Download the daily Fulton fish market data in worksheet format from `http://people.brandeis.edu/~kgraddy/data.html`. You can load the data into Matlab/Octave with

```
data=dlmread('fish.out.txt','\t',2,0);
dayOfWeek = data(:,1:4); % monday, tuesday, wednesday, thursday
stormy = data(:,6); % stormy weather
mixed = data(:,7);  % rough weather
logp = data(:,8); % log price
logq = data(:,9); % log quantity sold
rainy = data(:,10); % rainy weather
cold = data(:,11); % cold weather
windspd = data(:,12); % wind speed
windspd2 = data(:,13); % wind speed squared
price = data(:,14); % price
qRecieved = data(:,15); % quantity of fish received from fishermen
qSold = data(:,16); % quantity of fish sold
```

Even though this is time series data, you can assume that the observations are independent for this assignment.

(1) Estimate a least squares regression of log quantity on log price and a constant. Report the coefficient on log price and its standard error.
(2) Estimate a two stage least squares regression of log quantity on log price and a constant. Report the coefficient on log price and its standard error.

Part 2. The problem of minimizing a linear function subject to linear constraints,

$$\min_{x\in\mathbb{R}^n} c^T x \quad \text{subject to} \quad Ax \leq b$$

is called a linear program. Note that the constraints can include equality constraints or upper and lower bound constraints on x by defining A appropriately. Linear programs have many applications, so there are efficient algorithms to compute their solutions. Chapter 6 of Koenker (2005) is a good introduction to linear programming and how it relates to quantile regression.

Recall that quantile regression minimizes the average of the check function,

$$\min_{\beta \in \mathbb{R}^d} \mathbb{E}_n \left[\rho_\tau (y_i - x_i \beta) \right].$$

This can be written in the form of a linear program as follows. Let X be the $n \times d$ matrix of regressors, and Y be the $n \times 1$ vectors of outcomes.

$$\begin{aligned} \min_{(u,v,\beta) \in \mathbb{R}^{2n+d}} \; & \tau u + (1-\tau) v \\ \text{subject to} \; & \\ & Y - X\beta = u - v \\ & u \geq 0 \\ & v \geq 0 \end{aligned}$$

or

$$\min_{(u,v,\beta) \in \mathbb{R}^{2n+d}} \begin{pmatrix} \tau \iota_n \\ (1-\tau)\iota_n \\ 0\iota_d \end{pmatrix}^T \begin{pmatrix} u \\ v \\ \beta \end{pmatrix}$$

subject to

$$\begin{pmatrix} I_n & -I_n & X \\ -I_n & I_n & -X \\ -I_n & 0 & 0 \\ 0 & -I_n & 0 \end{pmatrix} \begin{pmatrix} u \\ v \\ \beta \end{pmatrix} \leq \begin{pmatrix} Y \\ Y \\ 0\iota_n \\ 0\iota_n \end{pmatrix}$$

where ι_n is a $n \times 1$ vector of ones. The Matlab command for solving linear programs is `linprog`, and the Octave command is `glpk`.

(1) Estimate a quantile regression of log quantity on log price. Do so for each decile, $\tau = 0.1, 0.2, ...0.9$.

(2) Estimate the standard error for your quantile regression coefficients. There are many ways to do this, one way is the following:

```
% estimate variance
e = y - x*b;
% bandwidth set similar to Koenker 2005 p81
c = numel(y)^(-1/3)*min(tau,1-tau);
h = std(e)*(norminv(tau+c)-norminv(tau-c))
% epanechnikov kernel
K = @(u) 0.75*(abs(u)<1).*(u.^2-1);
J = inv(((K(e/h)/h*ones(1,size(x,2))).*x)' *x/n);
xte = x.*((tau-1*(e>0))*ones(1,size(x,2)));
Om = xte'*xte/n;
V = J*Om*J/n;
```

Another option would be to use bootstrap. However you estimate the standard errors, say whether your method is robust to misspecification of the conditional quantile function.

(3) Graph your estimates of the coefficient of log price as a function of the quantile. Include 90% confidence bands. State whether your confidence bands are pointwise or uniform in τ.

Part 3. Suppose that market demand is

$$y^d = q(p, u)$$

where q is increasing in u, and $u \sim U(0, 1)$ is an unobserved variable capturing the state of demand. Suppose supply is

$$y^s = s(p, z, \epsilon)$$

for some observed variable z and unobserved variable ϵ. Then the equilibrium price solves

$$q(p, u) = s(p, z, \epsilon),$$

so

$$p = \delta(z, v)$$

for some function δ with $v = (u, \epsilon)$.

Assume that $q(p, u) = A(u)p^{\beta(u)}$

(1) Does the quantile regression in part 2 estimate $\beta(u)$ consistently?
(2) What does 2SLS from part 1 estimate in terms of $A(u)$ and $\beta(u)$?
(3) Under what assumptions can control function quantile regression be used to log $A(u)$ and $\beta(u)$?
(4) Under what assumptions can IV quantile regression be used to estimate $\log A(u)$ and $\beta(u)$?

Part 4. In this part, you will estimate $\beta(u)$ and $\log A(u)$ for $u = 0.1, 0.2, ..., 0.9$ using the control function approach.

(1) Regress log price on instruments wind speed, stormy, and mixed. Form the estimated residuals $\hat{v}$.
(2) Perform a quantile regression of log quantity on log price, a constant, and a series of $\hat{v}$. For example,

```
Pi = regress(logp,z);
vhat = logp - z*Pi;
v = vhat-min(vhat);
v = v/max(v)*pi;
xcf = [logp ones(n,1) cos(v*(1:k))];
b = qregress(logq,xcf,tau(j))
```

where `regress` and `qregress` are functions that perform regression and quantile regression (you have to create these).
(3) Plot the estimated demand curves.

Part 5. In this part, you will estimate $\beta(u)$ and $\log A(u)$ for $u = 0.1, 0.2, ..., 0.9$ using IV quantile regression.

(1) For a fine grid of possible values for $\beta(u)$, quantile regress log quantity minus $\beta(u)$ times log price on a constant and instruments, stormy and mixed. Calculate and save the norm of the coefficients of the instruments. Plot the norm of the coefficients of the instruments as a function of $\beta(u)$. Set $\hat{\beta}(u)$ to the value that minimizes the norm of the coefficients of the instruments.
(2) Repeat the previous steps using stormy, mixed, and wind speed as instruments. Comment on the results.
(3) Plot $\hat{\beta}(u)$ as function of u, along with the control function estimates from part 4 and the plain quantile regression estimates from part 2