Consider iid data $\{(Y_i, X_i')' : i=1,\dots,n\}$ and suppose the conditional distribution of Y_i given X_i is continous. Recall that $F(y \mid X_i) = P(Y_i < y \mid X_i)$.

We will assume that the conditional quantile of $Y_i \mid X_i$ is a parametric function of X_i:

$$q_\tau(X_i) = X_i'\beta_\tau \quad , \quad \beta_\tau \in \mathbb{R}^k$$

Then define

$$Q(b) = E\, \rho_\tau(Y_i - X_i'b)$$

$$Q_n(b) = \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(Y_i - X_i'b) \xrightarrow[p]{} E\, \rho_\tau(Y_i - X_i'b)$$

provided $E\left[(Y_i - X_i'b)\,(\tau - \mathbb{1}\{Y_i - X_i'b\})\right] = O(1)$

$\leq \tau E\, |Y_i - X_i'b|$

$\leq \tau E\left[|Y_i| + \|X_i\|\, b\right] < \infty$

if $E|Y_i| < \infty$ and $E\|X_i\| < \infty$.

where $$\frac{\partial \rho_\tau(u)}{\partial u} = \tau - \mathbb{1}\{u<0\}.$$

The FOC is:

$$\frac{\partial Q_n(b)}{\partial b} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \rho_\tau(Y_i - X_i'b)}{\partial b} = -\frac{1}{n}\sum_{i=1}^{n}(\tau - \mathbb{1}\{Y_i - X_i'b < 0\})\, X_i \qquad (1)$$

Another way to see it

Consider the following model

$$Y_i = X_i'\beta_\tau + u_i$$
$$P(u_i \leq 0 \mid X_i) = \tau$$

We have a conditional moment restriction given by

$$P(u_i \leq 0 \mid X_i) = P(Y_i \leq X_i'\beta_\tau \mid X_i) = \tau$$

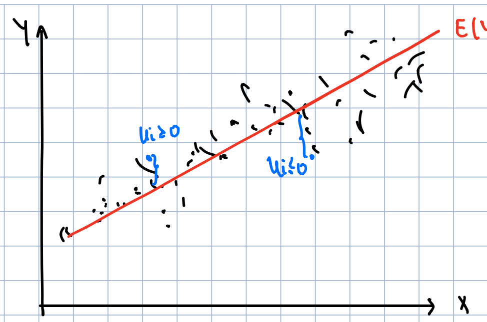
$$\Rightarrow \quad 0 = E\left[\tau - \mathbb{1}\{Y_i - X_i'\beta_\tau \leq 0\} \mid X_i\right]$$

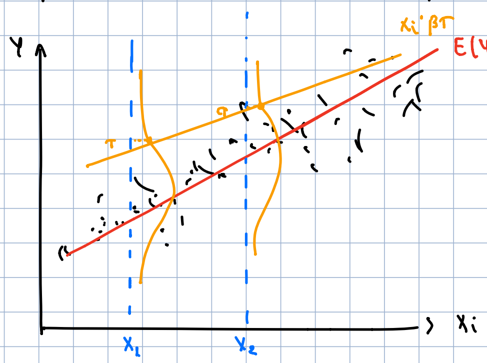
This can be also rephrased as

$$E\left[(\tau - \mathbb{1}\{Y_i - X_i'\beta_\tau\})\, g(X_i)\right] = 0$$

for any measurable function $g(\cdot)$.

So we are actually using $g(X_i) = X_i$, based on equation (1).





1. **(Quantile Regression with Instrumental Variables)** Consider the following model:

$$Y_i = X_i'\beta_\tau + U_i, \tag{1}$$

$$P(U_i \leq 0 | Z_i) = \tau, \tag{2}$$

where Y_i is the dependent variable, X_i is the k-vector of potentially endogenous regressors, $\beta_\tau \in \mathbb{R}^k$ is the vector of unknown coefficients, U_i is the unobserved error, Z_i is the l-vector of IVs, and $\tau \in (0,1)$ is known. Since X_i is potentially endogenous, we use the IVs Z_i to setup the model.

(a) **(5 points)** Show that the usual quantile regression model with exogenous regressors is a special case of the model described in (1)-(2).

(b) **(5 points)** Show that the model implies that

$$E(\tau - 1(Y_i \leq X_i'\beta_\tau))h(Z_i) = 0, \tag{3}$$

for any measurable function $h(\cdot)$ of Z_i.

solution:

We just derived that above.

(c) **(10 points)** Let $W_i = (Y_i, X_i', Z_i')'$, and assume that iid data $\{W_i : i = 1, \ldots, n\}$ were generated from the model described in (1)-(2). Using some given function $h : \mathbb{R}^l \to \mathbb{R}^k$, the econometrician proposes to estimate β_τ by solving the sample analogue of equation (3). In other words, the estimator $\hat\beta_{\tau,n}$ is defined so that

$$\frac{1}{n}\sum_{i=1}^{n}(\tau - 1(Y_i \le X_i'\hat\beta_{\tau,n}))h(Z_i) = o_p(n^{-1/2}). \tag{4}$$

Suppose that consistency of $\hat\beta_{\tau,n}$ has been established, i.e. $\hat\beta_{\tau,n} \to_p \beta_\tau$. Let

$$\begin{aligned} m(b) &= E(\tau - 1(Y_i \le X_i'b))h(Z_i), \\ H_n(b) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{(\tau - 1(Y_i \le X_i'b))h(Z_i) - m(b)\right\}, \end{aligned}$$

and assume that $H_n(b)$ is stochastically equicontinuous. Show that

$$H_n(\hat\beta_{\tau,n}) - H_n(\beta_\tau) \to_p 0.$$

solution:

$$\sqrt{n}\, o_p\left(\frac{1}{\sqrt{n}}\right) = \sqrt{n}\,\frac{1}{n}\sum_{i=1}^{n}(\tau - 1\{Y_i \le X_i'\hat\beta_{\tau,n}\})h(Z_i) \quad \pm \sqrt{n}\, E(\tau - 1\{Y_i \le X_i'\hat\beta_{\tau,n}\})h(Z_i)$$

$$o_p(1) = H_n(\hat\beta_{\tau,n}) + \sqrt{n}\, m(\hat\beta_{\tau,n})$$

We want to show $H_n(\hat\beta_{\tau,n}) = H_n(\beta_\tau) + o_p(1)$, i.e. for all $\varepsilon > 0$ $\lim_{n\to\infty} P(\|H_n(\hat\beta_{\tau,n}) - H_n(\beta_\tau)\| > \varepsilon) = 0$.

$$\begin{aligned} P(\|H_n(\hat\beta_{\tau,n}) - H_n(\beta_\tau)\| > \varepsilon) &= P(\|H_n(\hat\beta_{\tau,n}) - H_n(\beta_\tau)\| > \varepsilon, \|\hat\beta_{\tau,n} - \beta_\tau\| < \delta) \\ &\quad + P(\|H_n(\hat\beta_{\tau,n}) - H_n(\beta_\tau)\| > \varepsilon, \|\hat\beta_{\tau,n} - \beta_\tau\| \ge \delta) \\ &\le P(\|H_n(\hat\beta_{\tau,n}) - H_n(\beta_\tau)\| > \varepsilon, \|\hat\beta_{\tau,n} - \beta_\tau\| < \delta) \\ &\quad + P(\|\hat\beta_{\tau,n} - \beta_\tau\| \ge \delta) \\ &\le P(\|H_n(\hat\beta_{\tau,n}) - H_n(\beta_\tau)\| > \varepsilon, \|\hat\beta_{\tau,n} - \beta_\tau\| < \delta) \\ &\quad + o(1) \\ &\le \underbrace{P\Big(\sup_{b_1 \in B}\ \sup_{b_2 \in B(b_1,\delta)} \|H_n(b_1) - H_n(b_2)\| > \varepsilon\Big)}_{<\varepsilon \text{ when taking limit by SE}} + \underbrace{o(1)}_{=0 \text{ when taking limit.}} \end{aligned}$$

(d) **(10 points)** Find the asymptotic distribution of $H_n(\beta_\tau)$. Justify any additional assumptions you have to make.

solution:

$$H_n(\beta_\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[(\tau - \mathbb{1}\{y_i \le x_i'\beta_\tau\}) h(z_i) - \underbrace{E(\tau - \mathbb{1}\{y_i \le x_i'\beta_\tau\}) h(z_i)}_{\text{By LIE}} \Big]$$

$$E\big[(\tau - E[\mathbb{1}\{y_i \le x_i'\beta_\tau\} | z_i]) h(z_i) \big] = E[(\tau - \tau) h(z_i)] = 0.$$

$$\xrightarrow{d} N(0, E(\tau - \mathbb{1}\{u_i \le 0\})^2 h(z_i) h(z_i)')$$

$$= N(0, E[E[(\tau - \mathbb{1}\{u_i \le 0\})^2 | z_i] h(z_i) h(z_i)']) = N(0, \tau(1-\tau) E h(z_i) h(z_i)')$$

because $E((\tau - \mathbb{1}\{u_i \le 0\})^2 | z_i) = E[\tau^2 - 2\tau \mathbb{1}\{u_i \le 0\} + \mathbb{1}\{u_i \le 0\} | z_i]$

$$= \tau^2 - 2\tau^2 + \tau = \tau - \tau^2 = \tau(1-\tau).$$

And we need to assume $E h(z_i) h(z_i)'$ is finite and p.d.

(e) **(10 points)** Let $F(x|X_i, Z_i)$ denote the conditional CDF of U_i conditional on X_i and Z_i, and let $f(x|X_i, Z_i)$ denote the corresponding PDF. Assume that $f(\cdot|X_i, Z_i)$ is continuous at zero. Using the results of parts (c)-(d), derive the asymptotic distribution of $\hat\beta_{\tau,n}$. Assume that you can change the order of expectation and differentiation when needed. Justify any additional assumptions you have to make.

solution:

$$o_p(1) = H_n(\hat\beta_{\tau,n}) + \sqrt{n}\, m(\hat\beta_{\tau,n})$$

$$\underset{\text{mean value exp}}{=} H_n(\beta_\tau) + o_p(1) + \sqrt{n} \Big[\underbrace{m(\beta_\tau)}_{=0} + \frac{\partial m(\beta_\tau^*)}{\partial \beta} (\hat\beta_{\tau,n} - \beta_\tau) \Big]$$

where $$\frac{\partial m(\beta)}{\partial \beta} = \frac{\partial}{\partial \beta} \Big[E(\tau - \mathbb{1}\{y_i \le x_i'\beta\}) h(z_i) \Big] = \frac{\partial}{\partial \beta} \Big[E\{ (\tau - F(x_i'\beta | x_i, z_i)) h(z_i) \} \Big]$$

$$= -E[f(x_i'\beta | z_i, x_i) h(z_i) x_i']$$

Then $$\frac{\partial m(\beta_\tau^*)}{\partial \beta} \to -E[f(x_i'\beta_\tau | z_i, x_i) h(z_i) x_i'].$$

Finally, assuming $E[f(x_i'\beta_\tau | x_i, z_i) h(z_i) x_i']$ has full rank K, we write

$$\sqrt{n}(\hat\beta_{\tau,n} - \beta_\tau) = (E[f(x_i'\beta_\tau | z_i, x_i) h(z_i) x_i'] + o_p(1))^{-1} (H_n(\beta_\tau) + o_p(1))$$

$$\xrightarrow{d} E[f(x_i'\beta_\tau | z_i, x_i) h(z_i) x_i'] \, N(0, \tau(1-\tau) E h(z_i) h(z_i)')$$

$$= N(0, V_{\tau,h})$$

where

$$V_{\tau,h} = \tau(1-\tau)\left(E[f(x_i'\beta_\tau \mid z_i, x_i)\, h(z_i)\, x_i']\right)^{-1} E\, h(z_i) h(z_i)' \left(E[f(x_i'\beta_\tau \mid z_i, x_i)\, x_i\, h(z_i)']\right)^{-1}$$

(f) **(10 points)** What is the efficient choice of the function h? What is the corresponding asymptotic variance of $\hat{\beta}_{\tau,n}$?

solution:

Rewrite the conditional moment restriction as

$$0 = \tau - P(y_i \le x_i'\beta_\tau \mid z_i)$$

$$= E[\underbrace{\tau - F(x_i'\beta_\tau \mid x_i, z_i)}_{m(x_i, z_i, \beta_\tau)} \mid z_i]$$

Then

$$h^*(z_i) = \frac{1}{E[m^2(x_i, z_i, \beta_\tau) \mid z_i]} E\left[\frac{\partial m(x_i, z_i, \beta_\tau)}{\partial \beta} \mid z_i\right]$$

- $$E[(\tau - F(x_i'\beta_\tau \mid x_i, z_i))^2 \mid z_i] = E[E[(\tau - \mathbb{1}\{y_i \le x_i'\beta_\tau\})^2 \mid x_i, z_i] \mid z_i]$$
  $$= E[E[\tau^2 - 2\tau \mathbb{1}\{y_i \le x_i'\beta_\tau\} + \mathbb{1}\{y_i \le x_i'\beta_\tau\} \mid x_i, z_i] \mid z_i]$$
  $$= \tau(1-\tau)$$

- $$E\left[\frac{\partial m}{\partial \beta} \mid z_i\right] = -E[f(x_i'\beta_\tau \mid x_i, z_i)\, x_i \mid z_i]$$

Then

$$h^*(z_i) = \frac{E[f(x_i'\beta_\tau \mid x_i, z_i)\, x_i \mid z_i]}{\tau(1-\tau)}$$

which implies that

$$V_{h^*} = \tau(1-\tau)\left\{E\left(E[f(x_i'\beta_\tau \mid x_i, z_i)\, x_i \mid z_i]\, E[f(x_i'\beta_\tau \mid x_i, z_i)\, x_i' \mid z_i]\right)\right\}^{-1}$$

(g) **(5 points)** Consider the case of exogenous regressors when $X_i = Z_i$. Suppose that β_τ is estimated using the check function approach: $\tilde{\beta}_{\tau,n} = \arg\min_b \sum_{i=1}^n \rho_\tau(Y_i - X_i'b)$, where $\rho_\tau(u) = u(\tau - 1(u < 0))$. Given the results in (f), is $\tilde{\beta}_{\tau,n}$ an efficient estimator? Justify your answer.

solution:

$$h^*(X_i) = \frac{1}{E[m^2|X_i]} E\left[\frac{\partial m}{\partial \beta} \Big| X_i\right]$$

- $E[m^2|X_i] = E[(\tau - 1\{Y_i \le X_i'\beta_\tau\})^2 | X_i] = \tau(1-\tau)$
- $E[\frac{\partial m}{\partial \beta}|X_i] = E\left[\frac{\partial}{\partial \beta}(\tau - 1\{Y_i \le X_i'\beta_\tau\}) \Big| X_i\right]$

$$\underset{\substack{\text{interchange} \\ E,\ \partial/\partial\beta}}{=} \frac{\partial}{\partial \beta}[\tau - F(X_i'\beta_\tau|X_i)] = -f(X_i'\beta_\tau|X_i)X_i$$

So the efficient instrument is

$$h^*(X_i) = \frac{1}{\tau(1-\tau)} f(X_i'\beta_\tau|X_i)X_i$$

Therefore, the check function approach is not efficient in the case of quantile regression.