LECTURE: QUANTILE REGRESSION

Motivation

A correctly specified linear regression model describes the conditional mean of the dependent variable Y_i,

$$E(Y_i|X_i) = X_i'\beta. \tag{1}$$

and, therefore, captures the effect of X_i on the conditional mean of Y_i. The effect of X_i can be quite different if one focuses on other parts of the conditional distribution of Y_i (besides the mean). Consider the following normal model with heteroskedastic errors:

$$\begin{aligned} Y_i &= X_i'\beta + U_i, \\ U_i \mid X_i &\sim N(0, \sigma^2(X_i)). \end{aligned}$$

When the conditional distribution of Y_i given X_i is continuous, for $\tau \in (0,1)$ the τ-th conditional quantile of the distribution of Y_i given X_i is defined as $q_\tau(X_i)$ that satisfies:

$$P\left(Y_i \le q_\tau(X_i) \mid X_i\right) = \tau,$$

or

$$F(q_\tau(X_i) \mid X_i) = \tau,$$

where $F(\cdot|X_i)$ denotes the conditional CDF of Y_i given X_i.

Let z_τ denote the τ-th quantile of the standard normal distribution. Since

$$\frac{Y_i - X_i'\beta}{\sigma(X_i)} \mid X_i \sim N(0,1),$$

we have

$$\begin{aligned} \tau &= P\left(\frac{Y_i - X_i'\beta}{\sigma(X_i)} \le z_\tau \mid X_i\right) \\ &= P\left(Y_i \le X_i'\beta + \sigma(X_i) z_\tau \mid X_i\right), \end{aligned}$$

which implies that the τ-th conditional quantile of the distribution of Y_i is given by

$$q_\tau(X_i) = X_i'\beta + \sigma(X_i) z_\tau.$$

The marginal effect of X_i on the τ-th quantile of Y_i is therefore given by:

$$\frac{\partial q_\tau(X_i)}{\partial X_i} = \beta + \frac{\partial \sigma(X_i)}{\partial X_i} z_\tau.$$

Thus, if the errors are homoskedastic ($\sigma(X_i) = \sigma$ a.s.) the effect of X_i is the same for all $\tau \in (0,1)$ and coincides with the effect on the conditional mean of Y_i (which corresponds to $\tau = 0.5$). However, in the heteroskedastic model, the effect can be quite different for different τ. Moreover, since $z_\tau < 0$ for $\tau < 0.5$ and $z_\tau > 0$ for $\tau > 0.5$, the contribution of $\partial\sigma(X_i)/\partial X_i$ has opposite effects on upper and lower quantiles.

Hence, even in the simple normal heteroskedastic regression model, the effect of X_i is heterogeneous over different parts of the distribution. The idea of the quantile regression model is to capture the heterogeneity of the effects across quantiles without specifying *parametrically* the conditional distribution $F(\cdot \mid X_i)$

Check function and estimation of quantiles

Recall that the mean of the distribution of Y_i (unconditional) can be defined as the minimizer of the criterion function depending on the expected squared distance of Y_i from $y \in \mathbb{R}$:

$$EY_i = \arg\min_{y\in\mathbb{R}} E\left(Y_i - y\right)^2.$$

This definition motivates least squares estimation methods. It turns out that quantiles can be defined in a similar manner as minimizers of certain distance function.

For $\tau \in (0,1)$, define the so-called *check function*

$$\rho_\tau(u) = u \cdot (\tau - 1(u < 0)).$$

If the distribution of Y_i is continuous, one can show that the τ-th quantile of the distribution of Y_i, say q_τ, minimizes the distance between Y_i and $y \in \mathbb{R}$, where the distance is defined using the check function:

$$q_\tau = \arg\min_{y\in\mathbb{R}} E\rho_\tau(Y_i - y). \tag{2}$$

This can be shown by establishing that the first-order condition for the above problem is given by $\tau - F(q_\tau) = 0$, where $F(\cdot)$ is the CDF of Y_i and q_τ is the τ-th quantile of Y_i. Hence, the estimator of q_τ can be constructed as

$$\hat{q}_{\tau,n} = \arg\min_{y\in\mathbb{R}} n^{-1}\sum_{i=1}^{n}\rho_\tau(Y_i - y).$$

Quantile regression

We assume that data $\{(Y_i, X_i')' : i = 1,\ldots,n\}$ are iid. Suppose that the conditional distribution of Y_i given X_i is continuous, and let $F(\cdot \mid X_i)$ denote the conditional CDF of Y_i:

$$F(y \mid X_i) = P(Y_i < y \mid X_i).$$

Note that we leave the functional form of the CDF $F(\cdot \mid X_i)$ unspecified (unknown).

Similarly to the linear mean regression model in (1), the quantile regression model assumes that the τ-th conditional quantile of Y_i given X_i is a parametric function of X_i:

$$q_\tau(X_i) = X_i'\beta_\tau, \tag{3}$$

where $\beta_\tau \in \mathbb{B} \subset \mathbb{R}^k$, i.e.

$$F(X_i'\beta_\tau \mid X_i) = \tau.$$

Note that β_τ depends on τ and, therefore, the effect of X_i is allowed to be heterogeneous across quantiles of the conditional distribution of Y_i.

Proposition 1. *Suppose that the conditional distribution of Y_i given X_i is continuous,* (3) *holds, and EX_iX_i' is positive definite. Then,*

$$\beta_\tau = \arg\min_{b\in\mathbb{R}^k} E\rho_\tau(Y_i - X_i'b).$$

Proof. $E\rho_\tau(Y_i - X_i'b) = EE\left[\rho_\tau(Y_i - X_i'b) \mid X_i\right] \geq EE\left[\rho_\tau(Y_i - X_i'\beta_\tau) \mid X_i\right] = E\rho_\tau(Y_i - X_i'\beta_\tau)$, where the inequality holds by (2) and (3). Moreover, since EX_iX_i' is positive definite, the inequality is strict when $b \neq \beta_\tau$. □

The check-function based estimator of β_τ is an extremum estimator:

$$\hat{\beta}_{\tau,n} = \arg\min_{b\in\mathbb{B}} Q_n(b), \tag{4}$$
$$Q_n(b) = n^{-1}\sum_{i=1}^{n} \rho_\tau(Y_i - X_i'b),$$

where we assume that $\mathbb{B}$ is compact.

By the WLLN and when $E|Y_i| < \infty$ and $E\|X_i\| < \infty$,

$$Q_n(b) = n^{-1}\sum_{i=1}^{n} \rho_\tau(Y_i - X_i'b) \to_p E\rho_\tau(Y_i - X_i'b) = Q(b)$$

for all $b \in \mathbb{B}$. Moreover, β_τ is the unique minimizer of $Q(b)$. Hence, to establish consistency of $\hat{\beta}_{\tau,n}$, we need to show that convergence of $Q_n(b)$ to $Q(b)$ holds uniformly over $b \in \mathbb{B}$. The latter can be established using the following property of the check function.

Proposition 2. $|\rho_\tau(u_1) - \rho_\tau(u_2)| \le 2 \cdot |u_1 - u_2|$.

Proof. Write

$$\begin{aligned}
|\rho_\tau(u_1) - \rho_\tau(u_2)| &= |u_1(\tau - 1(u_1 < 0)) - u_2(\tau - 1(u_2 < 0))| \\
&\le \tau|u_1 - u_2| + |\Delta|, \text{ where} \\
\Delta &= u_2 \cdot 1(u_2 < 0) - u_1 \cdot 1(u_1 < 0).
\end{aligned}$$

We have four possible cases:

1. Suppose that $u_2 < 0$ and $u_1 < 0$:
$$\Delta = u_2 - u_1 \quad \Longrightarrow |\Delta| \le |u_2 - u_1|.$$
2. Suppose that $u_2 > 0$ and $u_1 > 0$:
$$\Delta = 0 \quad \Longrightarrow |\Delta| \le |u_2 - u_1|.$$
3. Suppose that $u_2 > 0$ and $u_1 < 0$:
$$0 < \Delta = -u_1 \le u_2 - u_1 \quad \Longrightarrow |\Delta| \le |u_2 - u_1|.$$
4. Suppose that $u_2 < 0$ and $u_1 > 0$:
$$0 > \Delta = u_2 > u_2 - u_1 \quad \Longrightarrow |\Delta| \le |u_2 - u_1|.$$

Hence,

$$\begin{aligned}
|\rho_\tau(u_1) - \rho_\tau(u_2)| &\le (1+\tau) \cdot |u_2 - u_1| \\
&\le 2 \cdot |u_2 - u_1|.
\end{aligned}$$

□

Proposition 2 shows that the check function is Lipschitz. The result can be used to show stochastic equicontinuity of

$$H_n(b) = Q_n(b) - Q(b) = n^{-1}\sum_{i=1}^{n} \left(\rho_\tau(Y_i - X_i'b) - E\rho_\tau(Y_i - X_i'b)\right), \tag{5}$$

which in turn can be used to show uniform convergence of $Q_n(b)$ and consistency of $\hat{\beta}_{\tau,n}$.

Proposition 3. *Suppose that $E\|X_i\| < \infty$. Then, $\{H_n(b) : n \geq 1\}$ is stochastically equicontinuous: for every $\epsilon > 0$ there is $\delta > 0$ such that*

$$\limsup_{n\to\infty} P\left(\sup_{b_1\in\mathbb{R}^k} \sup_{b_2:\|b_1-b_2\|<\delta} |H_n(b_1) - H_n(b_2)| > \epsilon\right) < \epsilon.$$

Proof. We have:

$$\begin{aligned}
|H_n(b_1) - H_n(b_2)| &\leq n^{-1}\sum_{i=1}^{n} |\rho_\tau(Y_i - X_i'b_1) - \rho_\tau(Y_i - X_i'b_2)| + E\,|\rho_\tau(Y_i - X_i'b_1) - \rho_\tau(Y_i - X_i'b_2)| \\
&\leq 2\left\{n^{-1}\sum_{i=1}^{n} |X_i'(b_1 - b_2)| + E\,|X_i'(b_1 - b_2)|\right\} \\
&\leq 2\|b_1 - b_2\|\left\{n^{-1}\sum_{i=1}^{n} \|X_i\| + E\|X_i\|\right\},
\end{aligned}$$

where the result in the first line holds by the triangle inequality, the result in the second line holds by Proposition 2, and the result in the last line holds by the Cauchy-Schwartz inequality. Thus,

$$\begin{aligned}
P\left(\sup_{b_1\in\mathbb{R}^k} \sup_{b_2:\|b_1-b_2\|<\delta} |H_n(b_1) - H_n(b_2)| > \epsilon\right) &\leq P\left(2\delta\left\{n^{-1}\sum_{i=1}^{n} \|X_i\| + E\|X_i\|\right\} > \epsilon\right) \\
&\leq \frac{4\delta E\|X_i\|}{\epsilon},
\end{aligned}$$

where the result in the first line holds by Markov's inequality. To complete the proof, choose

$$\delta < \frac{\epsilon^2}{4E\|X_i\|}.$$

□

Asymptotic normality of quantile regression estimators

The first order condition for the optimization problem in (4) is given by

$$\begin{aligned}
o_p\left(\frac{1}{n^{1/2}}\right) &= \frac{\partial Q_n(\hat\beta_{\tau,n})}{\partial b} \\
&= \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \rho_\tau(Y_i - X_i'\hat\beta_{\tau,n})}{\partial b} \\
&= \frac{1}{n}\sum_{i=1}^{n} \left(\tau - 1(Y_i < X_i'\hat\beta_{\tau,n})\right) X_i \qquad (6)
\end{aligned}$$

where the o_p-term in the first line is to allow for approximate optimization by numerical methods. The equality in the third line holds because the conditional distribution of Y_i is continuous: while the derivative of $\rho_\tau(u)$ is undefined at $u = 0$, $P(Y_i = X_i'\hat\beta_{\tau,n}) = 0$, and for $u \neq 0$,

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

Since

$$E\left(\tau - 1(Y_i < X_i'\beta_\tau) \mid X_i\right) = \tau - F(X_i'\beta_\tau \mid X_i) = 0,$$

we can interpret the term

$$\tau - 1(Y_i < X_i'\hat{\beta}_{\tau,n})$$

as the residual for the quantile regression. Hence, the condition in (6) can be interpreted as the normal equation for the quantile regression: the estimator $\hat{\beta}_{\tau,n}$ is chosen so that the residuals would be approximately orthogonal to the regressors in the sample.

A standard approach for establishing asymptotic normality of extremum estimators would involve the mean-value expansion of $\partial Q_n(\hat{\beta}_{\tau,n})/\partial\beta$ around $\partial Q_n(\beta_\tau)/\partial\beta$. Unfortunately, this approach cannot be used with quantile regression, because the second derivative of $Q_n(\hat{\beta}_{\tau,n})$ is zero with probability one: since $P(Y_i = X_i'\hat{\beta}_{\tau,n}) = 0$,

$$\frac{\partial\left(\tau - 1(Y_i < X_i'\hat{\beta}_{\tau,n})\right)}{\partial b} = 0 \quad \text{a.s.}$$

This is because the function $\partial Q_n(b)/\partial b$ changes with b only *in jumps*, and thus suffers from non-differentiability. Hence, an alternative approach is needed.

Below, we will describe the approach of Andrews (1994, Section 3.2). The approach can be used when estimators are defined using non-differentiable functions and heavily relies on stochastic equicontinuity.

While the sample function

$$\frac{1}{n}\sum_{i=1}^{n}\left(\tau - 1(Y_i < X_i'b)\right)X_i$$

is discontinuous (as a function of b), its population counterpart is continuous and differentiable:

$$m(b) = E\left[\left(\tau - 1(Y_i < X_i'b)\right)X_i\right] = E\left[\left(\tau - F(X_i'b \mid X_i)\right)X_i\right].$$

Hence, $m(\hat{\beta}_{\tau,n})$ can be mean-value expanded around $m(\beta_\tau) = 0$.

Define an *empirical process*

$$\nu_n(b) = \frac{1}{n^{1/2}}\sum_{i=1}^{n}\left\{\left(\tau - 1(Y_i < X_i'b)\right)X_i - m(b)\right\}, \tag{7}$$

and note that $E\nu_n(b) = 0$ for every $b \in \mathbb{B}$. The family of functions $\nu_n(\cdot)$, $\{\nu_n(\cdot) : n \geq 1\}$, is stochastically equicontinuous if for all $\epsilon > 0$ there is $\delta > 0$ such that

$$\limsup_{n\to\infty} P\left(\sup_{b_1\in\mathbb{B}}\ \sup_{b_2:\|b_1-b_2\|<\delta}\|\nu_n(b_1) - \nu_n(b_2)\| > \epsilon\right) < \epsilon.$$

Note that $\nu_n(\cdot) \in \mathbb{R}^k$, i.e. $\nu_n(\cdot)$ is vector-valued, and $\|\cdot\|$ denotes the Euclidean norm.

Let $f(\cdot \mid X_i)$ be the PDF corresponding to $F(\cdot \mid X_i)$.

Proposition 4. *Suppose that $\{\nu_n(\cdot) : n \geq 1\}$ is stochastically equicontinuous. Let*

$$B(b) = E\left[f(X_i'b \mid X_i)X_iX_i'\right],$$
$$\Omega_\tau = \tau(1-\tau)EX_iX_i',$$

Suppose that $f(\cdot \mid \cdot)$ is bounded from above, $f(\cdot \mid X_i)$ is continuous a.s., $E\|X_i\|^2 < \infty$, and EX_iX_i' and $B_\tau = B(\beta_\tau)$ are positive definite. Then,

$$n^{1/2}(\hat{\beta}_{\tau,n} - \beta_\tau) \to_d N\left(0, B_\tau^{-1}\Omega_\tau B_\tau^{-1}\right).$$

Proof. Using the definition of $\nu_n(b)$, re-write the first-order condition in (6) as

$$
\begin{aligned}
o_p(1) &= n^{1/2}\frac{\partial Q_n(\hat{\beta}_{\tau,n})}{\partial b} \\
&= \frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(\tau - 1(Y_i < X_i'\hat{\beta}_{\tau,n})\right)X_i \\
&= \frac{1}{n^{1/2}}\sum_{i=1}^{n}\left\{\left(\tau - 1(Y_i < X_i'\hat{\beta}_{\tau,n})\right)X_i - m(\hat{\beta}_{\tau,n})\right\} + n^{1/2}m(\hat{\beta}_{\tau,n}) \\
&= \nu_n(\hat{\beta}_{\tau,n}) + n^{1/2}m(\hat{\beta}_{\tau,n}). \qquad (8)
\end{aligned}
$$

Because $\hat{\beta}_{\tau,n} \to_p \beta_\tau$ and by stochastic equicontinuity of $\{\nu_n(\cdot) : n \geq 1\}$,

$$\nu_n(\hat{\beta}_{\tau,n}) - \nu_n(\beta_\tau) = o_p(1). \qquad (9)$$

To show (9), first note that, by consistency of $\hat{\beta}_{\tau,n}$, for all $\delta > 0$,

$$\lim_{n\to\infty} P\left(\|\hat{\beta}_{\tau,n} - \beta_\tau\| \geq \delta\right) = 0.$$

Next,

$$
\begin{aligned}
&P\left(\|\nu_n(\hat{\beta}_{\tau,n}) - \nu_n(\beta_\tau)\| > \epsilon\right) \\
&= P\left(\|\nu_n(\hat{\beta}_{\tau,n}) - \nu_n(\beta_\tau)\| > \epsilon, \|\hat{\beta}_{\tau,n} - \beta_\tau\| < \delta\right) + P\left(\|\nu_n(\hat{\beta}_{\tau,n}) - \nu_n(\beta_\tau)\| > \epsilon, \|\hat{\beta}_{\tau,n} - \beta_\tau\| \geq \delta\right) \\
&\leq P\left(\|\nu_n(\hat{\beta}_{\tau,n}) - \nu_n(\beta_\tau)\| > \epsilon, \|\hat{\beta}_{\tau,n} - \beta_\tau\| < \delta\right) + P\left(\|\hat{\beta}_{\tau,n} - \beta_\tau\| \geq \delta\right) \\
&= P\left(\|\nu_n(\hat{\beta}_{\tau,n}) - \nu_n(\beta_\tau)\| > \epsilon, \|\hat{\beta}_{\tau,n} - \beta_\tau\| < \delta\right) + o(1) \\
&\leq P\left(\sup_{b_1\in\mathbb{B}}\ \sup_{b_2:\|b_1-b_2\|<\delta} \|\nu_n(b_1) - \nu_n(b_2)\| > \epsilon\right) + o(1). \qquad (10)
\end{aligned}
$$

In view of the stochastic equicontinuity of $\{\nu_n(\cdot) : n \geq 1\}$ and (10),

$$\limsup_{n\to\infty} P\left(\|\nu_n(\hat{\beta}_{\tau,n}) - \nu_n(\beta_\tau)\| > \epsilon\right) < \epsilon \quad \text{for all } \epsilon > 0, \qquad (11)$$

which implies that[1]

$$\lim_{n\to\infty} P\left(\|\nu_n(\hat{\beta}_{\tau,n}) - \nu_n(\beta_\tau)\| > \epsilon\right) = 0 \quad \text{for all } \epsilon > 0. \qquad (12)$$

Using the result in (9), we can re-write (8) as

$$
\begin{aligned}
o_p(1) &= (\nu_n(\beta_\tau) + o_p(1)) + n^{1/2}m(\hat{\beta}_{\tau,n}) \\
&= \nu_n(\beta_\tau) + n^{1/2}m(\beta_\tau) + \frac{\partial m(\beta_n^*)}{\partial b'}n^{1/2}\left(\hat{\beta}_{\tau,n} - \beta_\tau\right),
\end{aligned}
$$

where the result in the last line holds by absorbing the o_p-term on the right-hand side with that on the left-hand side, and by applying a mean-value expansion to $m(\hat{\beta}_{\tau,n})$. Here, β_n^* denotes the mean value that satisfies

$$\|\beta_n^* - \beta_\tau\| \leq \|\hat{\beta}_{\tau,n} - \beta_\tau\|,$$

[1] Let $\Delta_n = \|\nu_n(\hat{\beta}_{\tau,n}) - \nu_n(\beta_\tau)\|$. Suppose (12) is not true. Then there is $\epsilon' > 0$ and $0 < \eta < \epsilon'$ such that $\limsup_{n\to\infty} P(\|\Delta_n\| > \epsilon') \geq \eta$. However, since $\eta < \epsilon'$, $\eta \leq \limsup_{n\to\infty} P(\|\Delta_n\| > \epsilon') \leq \limsup_{n\to\infty} P(\|\Delta_n\| > \eta)$, which contradicts (11). Hence, $\limsup_{n\to\infty} P(\|\Delta_n\| > \epsilon) = 0$ for all $\epsilon > 0$. Moreover, since probabilities are bounded by zero from below, lim sup can be replaced by lim.

which implies that

$$\beta_n^* \to_p \beta_\tau.$$

Moreover,

$$\begin{aligned}\frac{\partial m\left(\beta_n^*\right)}{\partial b'} &= -E\left[f(X_i'\beta_n^* \mid X_i)X_iX_i'\right] \\ &= -B(\beta_n^*) \\ &\to_p -B_\tau,\end{aligned}$$

where the result in the last line holds by continuity of $f(\cdot \mid X_i)$ and the dominated convergence theorem. Note that the matrices $B(b)$ are finite since $f(\cdot \mid X_i)$ is bounded and $E\|X_i\|^2 < \infty$.

Since $m(\beta_\tau) = 0$, we have

$$o_p(1) = \nu_n(\beta_\tau) - (B_\tau + o_p(1))\, n^{1/2}\left(\hat{\beta}_{\tau,n} - \beta_\tau\right),$$

or

$$\begin{aligned}n^{1/2}\left(\hat{\beta}_{\tau,n} - \beta_\tau\right) &= (B_\tau + o_p(1))^{-1}\left(\nu_n(\beta_\tau) + o_p(1)\right) \\ &= \left(B_\tau^{-1} + o_p(1)\right)\left(\nu_n(\beta_\tau) + o_p(1)\right), \qquad (13)\end{aligned}$$

where the equality in the second line holds by Slutsky's lemma since B_τ is positive definite. Next,

$$\begin{aligned}\nu_n(\beta_\tau) &= \frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(\tau - 1(Y_i < X_i'\beta_\tau)\right)X_i \\ &\to_d N(0, \Omega_\tau), \qquad (14)\end{aligned}$$

which holds by the CLT since Ω_τ is finite and positive definite, and

$$\begin{aligned}E\left[(\tau - 1(Y_i < X_i'\beta_\tau))^2 \mid X_i\right] &= \tau^2 - 2\tau F(X_i'\beta_\tau \mid X_i) + F(X_i'\beta_\tau \mid X_i) \\ &= \tau - \tau^2 \\ &= \tau(1-\tau),\end{aligned}$$

and therefore

$$Var\left(\left(\tau - 1(Y_i < X_i'\beta_\tau)\right)X_i\right) = \Omega_\tau.$$

The result follows from (13) and (14). □

Stochastic equicontinuity of the empirical process $\nu_n(\cdot)$

Unlike $H_n(\cdot)$ in equation (5), which is scaled by n^{-1}, the empirical process $\nu_n(\cdot)$ in equation (7) is scaled by $n^{-1/2}$. Hence, the asymptotic behavior of $\nu_n(\cdot)$ is driven by the CLT instead of the LLN. As a result, establishing stochastic equicontinuity of $\{\nu_n(\cdot)\}$ is more demanding than that for $\{H_n(\cdot)\}$. While the complete formal proof of the result is beyond the scope of this course, we will outline the main steps and ideas.

First, we will show that the distance between $\nu_n(b_1)$ and $\nu_n(b_2)$ can be bounded using the distance $\|b_1 - b_2\|$.

Proposition 5. *Suppose that $f(\cdot \mid \cdot)$ is bounded from above, and $E\|X_i\|^3 < \infty$. Then for small $\|b_1 - b_2\| < 1$,*

$$E\|\nu_n(b_1) - \nu_n(b_2)\|^2 = O\left(\|b_1 - b_2\|\right).$$

Proof. First,

$$\begin{aligned}
&\|\nu_n(b_1) - \nu_n(b_2)\|^2 \\
&= \left\| \frac{1}{n^{1/2}} \sum_{i=1}^{n} \{1(Y_i < X_i'b_1)X_i - E\left[F(X_i'b_1 \mid X_i)X_i\right]\} - \frac{1}{n^{1/2}} \sum_{i=1}^{n} \{1(Y_i < X_i'b_2)X_i - E\left[F(X_i'b_2 \mid X_i)X_i\right]\} \right\|^2 \\
&= \sum_{j=1}^{k} \left|\nu_{n,j}(b_1) - \nu_{n,j}(b_2)\right|^2, \qquad (15)
\end{aligned}$$

where for $j = 1, \ldots, k$,

$$\nu_{n,j}(b) = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \{1(Y_i < X_i'b)X_{i,j} - E\left[F(X_i'b \mid X_i)X_{i,j}\right]\}. \qquad (16)$$

Note that we defined $\nu_{n,j}(b)$ without the τ term as the latter does not depend on b. Next,

$$\begin{aligned}
E\left|\nu_{n,j}(b_1) - \nu_{n,j}(b_2)\right|^2 &= Var\left(\nu_{n,j}(b_1) - \nu_{n,j}(b_2)\right) \\
&= Var\left(\nu_{n,j}(b_1)\right) + Var\left(\nu_{n,j}(b_2)\right) - 2Cov(\nu_{n,j}(b_1), \nu_{n,j}(b_2)). \qquad (17)
\end{aligned}$$

For the variance terms, we have

$$\begin{aligned}
Var\left(\nu_{n,j}(b)\right) &= E\left[1(Y_i < X_i'b)X_{i,j}^2\right] - \left(E\left[F(X_i'b \mid X_i)X_{i,j}\right]\right)^2 \\
&= E\left[F(X_i'b \mid X_i)X_{i,j}^2\right] - \left(E\left[F(X_i'b \mid X_i)X_{i,j}\right]\right)^2. \qquad (18)
\end{aligned}$$

Let $u_1 \wedge u_2$ denote the minimum between u_1 and u_2. For the covariance term, we have

$$Cov(\nu_{n,j}(b_1), \nu_{n,j}(b_2)) = E\left[F(X_i'b_1 \wedge X_i'b_2 \mid X_i)X_{i,j}^2\right] - E\left[F(X_i'b_1 \mid X_i)X_{i,j}\right] \cdot E\left[F(X_i'b_2 \mid X_i)X_{i,j}\right],$$

where

$$\begin{aligned}
&E\left[F(X_i'b_1 \wedge X_i'b_2 \mid X_i)X_{i,j}^2\right] = \\
&\qquad E\left[F(X_i'b_1 \mid X_i)X_{i,j}^2 \cdot 1(X_i'b_1 \leq X_i'b_2)\right] + E\left[F(X_i'b_2 \mid X_i)X_{i,j}^2 \cdot 1(X_i'b_1 > X_i'b_2)\right]. \qquad (19)
\end{aligned}$$

Moreover,

$$\begin{aligned}
F(X_i'b_1 \mid X_i) &= F(X_i'b_2 \mid X_i) + f(X_i'b_{1,2} \mid X_i) \cdot X_i'(b_1 - b_2) \\
&= F(X_i'b_2 \mid X_i) + \|X_i\| O(\|b_1 - b_2\|), \qquad (20)
\end{aligned}$$

where the result in the last line holds with probability one because $f(\cdot \mid \cdot)$ is bounded from above (and by the Cauchy-Schwartz inequality).

By (18) and (20),

$$\begin{aligned}
Var\left(\nu_{n,j}(b_1)\right) &= E\left[F(X_i'b_2 \mid X_i)X_{i,j}^2\right] + E\left[\|X_i\|X_{i,j}^2\right] \cdot O(\|b_1 - b_2\|) \\
&\quad - \left\{E\left[F(X_i'b_2 \mid X_i)X_{i,j}\right] + E\left[\|X_i\|X_{i,j}\right] \cdot O(\|b_1 - b_2\|)\right\}^2 \\
&= Var\left(\nu_{n,j}(b_2)\right) + O(\|b_1 - b_2\|), \qquad (21)
\end{aligned}$$

where the result in the last line holds because $\|b_1 - b_2\|^2 < \|b_1 - b_2\|$, and $E\left[\|X_i\|X_{i,j}^2\right] \leq E\|X_i\|^3$. By (19) and (20),

$$\begin{aligned}
&E\left[F(X_i'b_1 \wedge X_i'b_2 \mid X_i)X_{i,j}^2\right] \\
&= E\left[F(X_i'b_2 \mid X_i)X_{i,j}^2\right] + E\left[\|X_i\|X_{i,j}^2 \cdot 1(X_i'b_1 \leq X_i'b_2)\right] O(\|b_1 - b_2\|),
\end{aligned}$$

and therefore,

$$Cov(\nu_{n,j}(b_1), \nu_{n,j}(b_2)) = Var\left(\nu_{n,j}(b_2)\right) + O(\|b_1 - b_2\|). \qquad (22)$$

The result follows by (15), (17), (21), and (22). □

Note that to have the result of Proposition 5, it is important that $E\nu_n(\cdot) = 0$.

The result of Proposition 5 together with Markov's inequality imply that $\|\nu_n(b_1) - \nu_n(b_2)\|$ is small in probability when $\|b_1 - b_2\|$ is small. However, for stochastic equicontinuity instead of a fixed pair b_1, b_2, we need to consider $\sup_{b_1,b_2 \in \mathbb{B}: \|b_1 - b_2\| < \delta} \|\nu_n(b_1) - \nu_n(b_2)\|$. Using so-called *maximal inequalities*, one can strengthen the result from fixed b_1, b_2 to that with the supremum, provided that the size of the class of functions under the supremum is not too large in the sense defined below.

Given iid data $\{W_i : i = 1, \ldots, n\}$, let $\mathcal{G}$ be a class of real-valued measurable functions g of W_i. The empirical process $\mathbb{G}_n$ evaluated at g is defined as

$$\mathbb{G}_n g = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \left(g(W_i) - Eg(W_i)\right), \quad g \in \mathcal{G}.$$

Given two functions l and u (of W_i),[2] the bracket $[l, u]$ is the set of all functions g such that $l \leq g \leq u$:

$$[l, u] = \{g \in \mathcal{G} : l \leq g \leq u\}.$$

A collection of brackets $\{[l, u]_\alpha : \alpha \in \mathcal{A}\}$ covers $\mathcal{G}$ if $\mathcal{G} \subset \cup_{\alpha \in \mathcal{A}} [l, u]_\alpha$. Note that the functions l and u do not have to be in $\mathcal{G}$. With the L_p-norm defined as

$$\|g\|_p = \left(E|g|^p\right)^{1/p} = \left(E|g(W_i)|^p\right)^{1/p},$$

we say that a bracket $[l, u]$ is an ϵ-bracket in L_p if $\|u - l\|_p < \epsilon$. The *bracketing number* $N_{[]}(\epsilon, \mathcal{G}, L_p)$ is the smallest number of ϵ-brackets $[l, u]$ needed to cover $\mathcal{G}$. The *entropy with bracketing* is $\log N_{[]}(\epsilon, \mathcal{G}, L_p)$. The entropy measures the size of $\mathcal{G}$.

The following result is a maximal inequality based on the entropy with bracketing (see van der Vaart, 1998, Corollary 19.34).

Proposition 6. *Suppose that (i) $\|g\|_2 < \delta$ for all $g \in \mathcal{G}$, and (ii) there is an envelope function G such that $|g| \leq G$ for all $g \in \mathcal{G}$. Then for some constant $c > 0$,*[3]

$$E \sup_{g \in \mathcal{G}} |\mathbb{G}_n g| \leq c \cdot \left\{ \int_0^\delta \sqrt{\log N_{[]}(\epsilon, \mathcal{G}, L_2)} d\epsilon + n^{1/2} E\left[G \cdot 1\left(G > \delta \sqrt{\frac{n}{\log N_{[]}(\delta, \mathcal{G}, L_2)}} \right) \right] \right\}. \tag{23}$$

The result shows that a maximal inequality for the empirical process indexed by functions in $\mathcal{G}$ depends on the size of $\mathcal{G}$ as measured by the integral of the entropy with bracketing. The following result is similar to Theorem 19.5 in van der Vaart (1998).

Proposition 7. *Suppose that*

$$\int_0^1 \sqrt{\log N_{[]}(\epsilon, \mathcal{G}, L_2)} d\epsilon < \infty. \tag{24}$$

Then, $\{\mathbb{G}_n : n \geq 1\}$ is stochastically equicontinuous: i.e. for every $\varepsilon > 0$ there is $\delta > 0$ such that

$$\limsup_{n \to \infty} P\left(\sup_{g_1 \in \mathcal{G}} \sup_{g_2 \in \mathcal{G}: \|g_1 - g_2\|_2 < \delta} |\mathbb{G}_n (g_1 - g_2)| > \varepsilon \right) < \varepsilon.$$

Remark. The bracketing number $N_{[]}(\epsilon, \mathcal{G}, L_2)$ increases to infinity as $\epsilon \downarrow 0$, and therefore the integral in (24) is determined by the behavior of the bracketing numbers near zero. The upper bound of the integral is arbitrary, and one can use any constant. Moreover, $N_{[]}(\epsilon, \mathcal{G}, L_2)$ is a decreasing function of ϵ, and for all large enough ϵ's, $N_{[]}(\epsilon, \mathcal{G}, L_2) = 1$. Hence, the entropy integral condition in (24) can be replaced with

$$\int_0^\infty \sqrt{\log N_{[]}(\epsilon, \mathcal{G}, L_2)} d\epsilon < \infty. \tag{25}$$

[2] To simplify the notation, sometimes we suppress the dependence of functions on W_i.

[3] For general empirical processes, the supremum in (23) may not be measurable, which can be addressed by using outer probabilities and expectations (see van der Vaart, 1998, Chapter 18.2).

Proof of Proposition 7. Since the integral in (24) is finite, $N_{[]}(\epsilon, \mathcal{G}, L_2) < \infty$ for all $\epsilon > 0$. Therefore, for every $\epsilon > 0$ there is a finite collection of ϵ-brackets covering $\mathcal{G}$.

To satisfy condition (i) of Proposition 6, consider $\mathcal{G}_\delta = \{g_1 - g_2 : g_1, g_2 \in \mathcal{G}, \|g_1 - g_2\|_2 < \delta\}$. If $g_1 \in [l_1, u_1]$ and $g_2 \in [u_2, l_2]$, then

$$g_1 - g_2 \in [l_1 - u_2, u_1 - l_2].$$

Moreover, if $[l_j, u_j]$ are $\epsilon/2$-brackets, then

$$\|(u_1 - l_2) - (l_1 - u_2)\|_2 \le \|u_1 - l_1\|_2 + \|u_2 - l_2\|_2 < \epsilon.$$

Hence,

$$N_{[]}(\epsilon, \mathcal{G}_\delta, L_2) \le N_{[]}^2(\epsilon/2, \mathcal{G}, L_2),$$

where $N_{[]}^2(\epsilon/2, \mathcal{G}, L_2)$ bound is obtained by considering all possible pairs of l_i's and u_j's of the collection of $\epsilon/2$-brackets that covers $\mathcal{G}$.

The envelope function G for $\mathcal{G}_\delta$ can be constructed as follows. For ϵ large enough, $N_{[]}(\epsilon, \mathcal{G}_\delta, L_2) = 1$. Let l and u be the corresponding functions. We can take $G = |l| + |u|$. Moreover, since $\|u - l\|_2^2 = \|u\|_2^2 + \|l\|_2^2 - 2E[u \cdot l] < \infty$, $\|G\|_2^2 < \infty$. Thus, condition (ii) of of Proposition 6 is satisfied with a square-integrable envelope function.

Let

$$a(\delta) = \delta \sqrt{\frac{1}{\log N_{[]}(\delta, \mathcal{G}, L_2)}},$$

and note that for all $\delta > 0$,

$$0 < a(\delta) < \infty.$$

For the second term on the right-hand side in (23) and every fixed $\delta > 0$,

$$\begin{aligned} n^{1/2} E\left[G \cdot 1\left(G > a(\delta) n^{1/2}\right)\right] &\le \frac{1}{a(\delta)} E\left[G^2 \cdot 1\left(G > a(\delta) n^{1/2}\right)\right] \\ &\to 0 \text{ as } n \to \infty. \end{aligned}$$

The inequality in the first line holds because for the purpose of the expectation, $n^{1/2} a(\delta) < G$. The result in the second line holds by the dominated convergence theorem since $EG^2 < \infty$ and $\lim_{n\to\infty} 1(G > a(\delta) n^{1/2}) = 0$.

For the first term on the right-hand side in (23),

$$\int_0^\delta \sqrt{\log N_{[]}(\epsilon, \mathcal{G}, L_2)} d\epsilon \to 0 \text{ as } \delta \to 0,$$

which holds because $\int_0^\delta \sqrt{\log N_{[]}(\epsilon, \mathcal{G}, L_2)} d\epsilon < \infty$. □

In the case of quantile regression, the maximal inequality can be applied as follows. Note that by (15) in the proof of Proposition 5, without loss of generality it suffices to show stochastic equicontinuity of $\{\nu_{n,1}(\cdot) : n \ge 1\}$, where the empirical process $\nu_{n,1}(b)$, $b \in \mathbb{B} \subset \mathbb{R}^k$, is given by

$$\nu_{n,1}(b) = \frac{1}{n^{1/2}} \sum_{i=1}^n \{1(Y_i < X_i' b) X_{i,1} - E[F(X_i' b \mid X_i) X_{i,1}]\}.$$

Proposition 8. *Suppose that $f(\cdot \mid \cdot)$ is bounded from above, and $P(\|X_i\| \le M) = 1$ for some $M > 0$. Then, $\{\nu_{n,1}(\cdot) : n \ge 1\}$ is stochastically equicontinuous: for every $\epsilon > 0$ there is $\delta > 0$ such that*

$$\limsup_{n\to\infty} P\left(\sup_{b_1 \in \mathbb{B}} \sup_{b_2 : \|b_1 - b_2\| < \delta} |\nu_{n,1}(b_1) - \nu_{n,1}(b_2)| > \epsilon\right) < \epsilon.$$

Proof. Consider the class of functions

$$\mathcal{G} = \{g_b(y,x) = 1(y < x'b)x_1 : b \in \mathbb{B} \subset \mathbb{R}^k\},$$

where $y \in \mathbb{R}$, $x \in \mathbb{R}^k$, and x_1 denotes the first element of x. By the same arguments as in the proof of Proposition 5,

$$\begin{aligned} \|g_{b_1} - g_{b_2}\|_2^2 &= E\left[\left(1(Y_i < X_i'b_1) - 1(Y_i < X_i'b_2)\right)^2 X_{i,1}^2\right] \\ &= E\left[F(X_i'b_1 \mid X_i)X_{i,1}^2\right] + E\left[F(X_i'b_2 \mid X_i)X_{i,1}^2\right] - 2E\left[F(X_i'b_1 \wedge X_i'b_2 \mid X_i)X_{i,1}^2\right] \\ &\leq K \cdot \|b_1 - b_2\| \end{aligned} \tag{26}$$

for some constant $K > 0$ that depends on the bound for the density $f(\cdot \mid \cdot)$ and $E\|X_i\|^3$, or

$$\|g_{b_1} - g_{b_2}\|_2 \leq \sqrt{K \cdot \|b_1 - b_2\|}.$$

Hence, whenever $\|b_1 - b_2\| < \epsilon^2/K$, $\|g_{b_1} - g_{b_2}\|_2 < \epsilon$.

Brackets $[l, u]$ for $\mathcal{G}$ can be constructed as follows. Let $b_1, \ldots, b_J \in \mathbb{B}$ be as described below. We define

$$l_j(y,x) = 1\left(y < x'b_j - \frac{\epsilon^2}{2K} \cdot \text{sign}(x_1)\right) x_1, \tag{27}$$

$$u_j(y,x) = 1\left(y < x'b_j + \frac{\epsilon^2}{2K} \cdot \text{sign}(x_1)\right) x_1. \tag{28}$$

We have $l_j \leq u_j$, and similarly to (26),

$$\|l_j - u_j\|_2 \leq \epsilon.$$

A function $g_b \in [l_j, u_j]$ if

$$l_j(y,x) \leq g_b(y,x) \leq u_j(y,x).$$

When $x_1 > 0$, the condition holds if

$$x'b_j - \frac{\epsilon^2}{2K} \leq x'b \leq x'b_j + \frac{\epsilon^2}{2K},$$

or

$$|x'(b - b_j)| \leq \frac{\epsilon^2}{2K}.$$

Since $|x'(b - b_j)| \leq M\|b - b_j\|$, $g_b \in [l_j, u_j]$ when $\|b - b_j\| \leq \epsilon^2/(2KM)$. Hence, the number of brackets needed to cover $\mathbb{B}$ is the same as the number of balls of radius $\epsilon^2/(2KM)$ needed to cover the compact set $\mathbb{B} \subset \mathbb{R}^k$. With properly chosen $b_1, \ldots, b_J$, this number (J) is of order

$$\left(\frac{2\sup_{b\in\mathbb{B}} \max_{1\leq j\leq k} |b_j|}{\epsilon^2/(2KM)}\right)^k = \frac{\left(4KM \cdot \sup_{b\in\mathbb{B}} 2\sup_{b\in\mathbb{B}} \max_{1\leq j\leq k} |b_j|\right)^k}{\epsilon^{2k}}$$

Hence, when constructing l_j and u_j in (27)-(28), $b_1, \ldots, b_J$ can be chosen as the centers of the balls with radius $\epsilon^2/(2KM)$ that cover $\mathbb{B}$. We have that there is a constant $\Delta > 0$ such that

$$N_{[]}(\epsilon, \mathcal{G}, L_2) \leq \frac{\Delta}{\epsilon^{2k}}.$$

Lastly, we verify the entropy integral condition in (25).

$$\begin{aligned} \int_0^\infty \sqrt{\log N_{[]}(\epsilon, \mathcal{G}, L_2)} d\epsilon &\leq \int_0^\infty \sqrt{\log \frac{\Delta}{\epsilon^{2k}}} d\epsilon \\ &= \Delta^{\frac{1}{2k}} \int_0^\infty \sqrt{\log \frac{1}{\epsilon^{2k}}} d\epsilon \\ &= \Delta^{\frac{1}{2k}} (2k)^{1/2} \int_0^\infty u^{1/2} e^{-u} du, \end{aligned}$$

where the equality in the second line holds by the change of variable argument: Define $u = \log \frac{\Delta}{\epsilon^{2k}}$, so that $e^u = \frac{\Delta}{\epsilon^{2k}}$, $\epsilon^{2k} = \Delta e^{-u}$, $\epsilon = \Delta^{1/(2k)} e^{-u/(2k)}$, and $d\epsilon = -\frac{\Delta^{1/(2k)}}{2k} e^{-u/(2k)}$. Note that the bounds of integration switch:

$$\begin{aligned}
\int_0^\infty \sqrt{\log \frac{\Delta}{\epsilon^{2k}}} d\epsilon &= -\frac{\Delta^{\frac{1}{2k}}}{2k} \int_\infty^0 u^{1/2} e^{-u/2k} du \\
&= \Delta^{\frac{1}{2k}} \int_0^\infty u^{1/2} e^{-u/2k} d(u/2k) \\
&= \Delta^{\frac{1}{2k}} (2k)^{1/2} \int_0^\infty u^{1/2} e^{-u} du.
\end{aligned}$$

$$\begin{aligned}
\int_0^\infty u^{1/2} e^{-u} du &= \int_0^1 u^{1/2} e^{-u} du + \int_1^\infty u^{1/2} e^{-u} du \\
&\leq \int_0^1 e^{-u} du + \int_1^\infty u e^{-u} du \\
&= 1 - \frac{1}{e} + \int_1^\infty u e^{-u} du \\
&= 1 + \frac{1}{e},
\end{aligned}$$

where the result in the last line is by applying integration by parts to the second integral:

$$\int_1^\infty u e^{-u} du = -\int_1^\infty u e^{-u} d(-u) = -\int_1^\infty u d(e^{-u}) = -u e^{-u}\Big|_1^\infty + \int_1^\infty e^{-u} du = \frac{2}{e}.$$

$\square$

Bibliography

Andrews, D. W. K., 1994. Empirical process methods in econometrics. In: Engle, R., D.McFadden (Eds.), Handbook of Econometrics. Vol. 4. Elsevier Science, Amsterdam: North Holland, Ch. 37, pp. 2247–2294.

van der Vaart, A. W., 1998. Asymptotic Statistics. Cambridge University Press, Cambridge.