Instructions: This is a 20 point exam, with equal weights for each question. The answers must be turned in no later than 25 hours after you pick up the exams, to Jim Powell (669 Evans). You may consult and cite any lecture notes and any of the references on the syllabus; you may not cite any other outside source, and under no circumstances should you discuss the exam with anyone other than the instructor before you submit your answers. Please make your answers elegant – that is, clear, concise, and, above all, correct.

Suppose a sample of N i.i.d. observations on a scalar dependent variable y_i and p-dimensional vector of (non-constant) regressors x_i satifies a linear model

$$y_i = x_i'\beta_0 + \varepsilon_i,$$

where the slope coefficients β_0 are unknown, and the unobservable error term ε_i is statistically independent of the regressors x_i, with (unknown) marginal density function $f(\varepsilon)$ that is very well behaved (i.e., having lots of continuous derivatives, with the level and derivatives of f being uniformly bounded).

A "rank regression" estimator of the slope coefficient vector β_0 is defined to minimize the sum of absolute deviations of differences in dependent variables $y_i - y_j$ and corresponding differences in regression functions $(x_i - x_j)'\beta$ across all distinct pairs of observations; that is,

$$\hat{\beta} \equiv \arg\min_{\beta \in R^p} S_n(\beta),$$

$$S_n(\beta) \equiv \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} |(y_i - y_j) - (x_i - x_j)'\beta|.$$

1. Give an argument for consistency of $\hat{\beta}$ for β_0 under the assumptions on the model given above, using analogous arguments to those for consistency of the LAD estimator of regression coefficients under a conditional median restriction.

Answer: Since the error terms $\{\varepsilon_i\}$ are i.i.d. (and independent of the regressors), the difference $u_{ij} \equiv \varepsilon_i - \varepsilon_j$ is symmetrically distributed about zero when $i \neq j$, with density $f_u(u) \equiv \int f(u+v)f(v)dv$ (by the convolution formula). Thus writing $Y_{ij} \equiv y_i - y_j$ and $X_{ij} \equiv x_i - x_j$, the conditional median of Y_{ij} given X_{ij} is $X_{ij}'\beta_0$.

First note that $\hat{\beta}$ also minimizes $\tilde{S}_n(\beta) \equiv S_n(\beta) - S_n(\beta_0)$, so, by the U-statistic theorem, $S_n(\beta)$ converges to its expectation $\bar{S}(\beta) \equiv E[S_n(\beta)]$ for each $\beta \in R^p$ as long as

$$E[(|(y_i - y_j) - (x_i - x_j)'\beta| - |(y_i - y_j) - (x_i - x_j)'\beta_0|)^2]$$

is finite. But since

$$(|(y_i - y_j) - (x_i - x_j)'\beta| - |(y_i - y_j) - (x_i - x_j)'\beta_0|)^2 \equiv (|u_{ij} - X_{ij}'(\beta - \beta_0)| - |u_{ij}|)^2$$
$$\leq \|X_{ij}\|^2 \|\beta - \beta_0\|^2 = \|x_i - x_j\|^2 \|\beta - \beta_0\|^2,$$

this will be finite as long as x_i has bounded second moments. Under this condition, $\tilde{S}_n(\beta)$ will converge pointwise to its expectation, and, since it is a convex function of β. consistency will follow if $\bar{S}(\beta)$ is uniquely minimized at the true value β_0. This will hold under two regularity conditions:

(i) The matrix $\Sigma_{xx} = V(x_i) = (1/2) \cdot E[(x_i - x_j)(x_i - x_j)'] = (1/2) \cdot E[X_{ij}X_{ij}']$ is positive definite (i.e., full rank); and

(ii) The density function of $u_{ij} = \varepsilon_i - \varepsilon_j$ at zero, $f_u(0) \equiv \int f(u+v)f(v)dv = \int [f(u)]^2 du \equiv \tau_0,$ is positive, i.e., $\tau_0 > 0.$

Under these two conditions, uniqueness of β_0 as the minimizer of $\bar{S}(\beta)$ and the consistency of $\hat{\beta}$ both follow from the same arguments as given for LAD consistency in the "Notes on Quantile Regression."■

2. The approximate first-order condition for the minimization problem defining $\hat{\beta}$ is

$$\hat{\Psi}_N(\hat{\beta}) \equiv \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} sgn\left\{(y_i - y_j) - (x_i - x_j)'\hat{\beta}\right\} \cdot (x_i - x_j)$$
$$\equiv \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} sgn\{\hat{\varepsilon}_i - \hat{\varepsilon}_j\} \cdot (x_i - x_j)$$
$$= o_p\left(\frac{1}{\sqrt{N}}\right),$$

where, as usual,

$$sgn\{u\} \equiv 1\{u \geq 0\} - 1\{u \leq 0\}$$

and

$$\hat{\varepsilon}_i \equiv y_i - x_i'\hat{\beta}.$$

Show that this condition is equivalent to a sample moment condition which sets the sample covariance of the regressors and the ranks of the residuals to zero (approximately); that is, show that

$$\frac{1}{N} \sum_{i=1}^{N} \left(\frac{\hat{R}_i}{N+1} - \frac{1}{2} \right) \cdot x_i = o_p\left(\frac{1}{\sqrt{N}}\right),$$

where

$$\hat{R}_i \equiv \sum_{j=1}^{N} 1\{\hat{\varepsilon}_j \leq \hat{\varepsilon}_i\}$$

is the rank of the i^{th} residual in the sample, with

$$\sum_{i=1}^{N} \hat{R}_i = \binom{N+1}{2} = \frac{N(N+1)}{2}$$

(ignoring possible ties in the residuals). [You should convert the U-statistic $\hat{\Psi}_N$ into the corresponding V-statistic before starting on the algebra.]

Answer: The U-statistic $\hat{\Psi}_N$ can be rewritten as

$$\begin{aligned}
\tilde{\Psi}_N(\hat{\beta}) &\equiv \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} sgn\left\{(y_i - y_j) - (x_i - x_j)'\hat{\beta}\right\} \cdot (x_i - x_j) \\
&\equiv \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} [1\{\hat{\varepsilon}_i \geq \hat{\varepsilon}_j\} - 1\{\hat{\varepsilon}_j \geq \hat{\varepsilon}_i\}] \cdot (x_i - x_j) \\
&= \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} 1\{\hat{\varepsilon}_i \geq \hat{\varepsilon}_j\} x_i + \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} 1\{\hat{\varepsilon}_j \geq \hat{\varepsilon}_j\} x_j \\
&\quad - \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} 1\{\hat{\varepsilon}_i \geq \hat{\varepsilon}_j\} x_j - \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} 1\{\hat{\varepsilon}_j \geq \hat{\varepsilon}_i\} x_i.
\end{aligned}$$

The first two terms are equal, and can be rewritten as

$$\begin{aligned}
\frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} 1\{\hat{\varepsilon}_i \geq \hat{\varepsilon}_j\} x_i &= \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} 1\{\hat{\varepsilon}_j \geq \hat{\varepsilon}_j\} x_j \\
&= \frac{(N+1)}{N(N-1)} \sum_{i=1}^{N} \frac{\hat{R}_i}{(N+1)} \cdot x_i.
\end{aligned}$$

To get at the third and fourth terms, the identity

$$1\{\hat{\varepsilon}_i \geq \hat{\varepsilon}_j\} \equiv 1 - 1\{\hat{\varepsilon}_j \geq \hat{\varepsilon}_j\} + 1\{\hat{\varepsilon}_i = \hat{\varepsilon}_j\}$$

is useful; with it, those two terms can be written as

$$\frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1}^{N}1\{\hat{\varepsilon}_i \geq \hat{\varepsilon}_j\}x_j = \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1}^{N}1\{\hat{\varepsilon}_j \geq \hat{\varepsilon}_i\}x_i$$

$$= \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1}^{N}x_i - \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1}^{N}1\{\hat{\varepsilon}_i \geq \hat{\varepsilon}_j\}x_i$$

$$+ \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1}^{N}1\{\hat{\varepsilon}_j = \hat{\varepsilon}_i\}x_i$$

$$= \frac{2}{N-1}\sum_{i=1}^{N}\left(\frac{1}{2}\right)\cdot x_i - \frac{(N+1)}{N(N-1)}\sum_{i=1}^{N}\frac{\hat{R}_i}{(N+1)}\cdot x_i + o_p\left(\frac{\sqrt{N}}{N(N-1)}\right)$$

where that last term follows from the usual LAD argument that exploits continuity of the $\{y_i\}$,

$$\left|\sum_{i=1}^{N}\sum_{j=1}^{N}1\{\hat{\varepsilon}_j = \hat{\varepsilon}_i\}x_i\right| \leq \max_i\{\|x_i\|\}\cdot\sum_{i=1}^{N}\sum_{j=1}^{N}1\left\{y_i - x_i'\hat{\beta} = y_j - x_j'\hat{\beta}\right\}$$

$$= \max_i\{\|x_i\|\}\cdot p \qquad (w.p.1)$$

$$= o_p(\sqrt{N}).$$

Putting all of this stuff together,

$$\tilde{\Psi}_N(\hat{\beta}) \equiv \frac{4(N+1)}{N(N-1)}\sum_{i=1}^{N}\frac{\hat{R}_i}{(N+1)}\cdot x_i - \frac{2N}{N(N-1)}\sum_{i=1}^{N}\left(\frac{1}{2}\right)\cdot x_i + +o_p\left(\frac{\sqrt{N}}{N(N-1)}\right)$$

$$= 4\left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{R}_i}{N+1} - \frac{1}{2}\right)\cdot x_i\right) + o_p\left(\frac{1}{\sqrt{N}}\right)$$

$$= o_p\left(\frac{1}{\sqrt{N}}\right),$$

which establishes the result.■

3. Rewriting the U-process $\hat{\Psi}_N(\beta)$ characterizing the first-order condition as

$$\hat{\Psi}_N(\beta) \equiv \binom{N}{2}^{-1}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho(z_i, z_j; \beta)$$

(with $z_i \equiv (y_i, x_i')'$), assume (without proof) that the estimator $\hat{\beta}$ also solves the approximate moment condition

$$\tilde{\Psi}_N(\hat{\beta}) = o_p\left(\frac{1}{\sqrt{N}}\right),$$

where $\tilde{\Psi}_N(\beta)$ is the projection of the U-process,

$$\tilde{\Psi}_N(\beta) \equiv \frac{1}{N}\sum_{i=1}^{N} \psi(z_i;\beta),$$

$$\psi(z_i;\beta) \equiv E[\rho(z_i,z_j;\beta)|z_i].$$

Calculate the form of the function ψ, and use this expression to derive the limiting normal distribution of $\sqrt{N}\,(\hat{\beta}-\beta_0)$ under the given assumptions. Your expression for the asymptotic covariance matrix should involve the nuisance parameter

$$\tau_0 \equiv E[f(\varepsilon_i)] \;=\int [f(u)]^2 du.$$

[Hint: you will need to use the fact that, if $F(u)$ is the c.d.f. of ε_i, then $F(\varepsilon_i) \equiv u_i$ is uniformly distributed on $(0,1)$.]

Answer: From the usual expansion of first-order conditions for smooth M-estimators, the asymptotic distribution of $\hat{\beta}$ will be

$$\sqrt{N}\,(\hat{\beta}-\beta_0) \overset{d}{\to} N(0, D_0^{-1}V_0[D_0']^{-1}),$$

where

$$D_0 = -E\left[\frac{\partial\psi(z_i,\beta_0)}{\partial\beta'}\right]$$

and

$$V_0 = Var[\psi(z_i,\beta_0)].$$

For this problem, the projection function $\psi(z_i;\beta)$ is

$$\begin{aligned}\psi(z_i;\beta) &\equiv E[(1\{\varepsilon_i + (x_i - x_j)\delta \geq \varepsilon_j\} - 1\{\varepsilon_i + (x_i - x_j)\delta \leq \varepsilon_j\}) \cdot (x_i - x_j)|\varepsilon_i, x_i] \\ &= E[[2F(\varepsilon_i + (x_i - x_j)\delta) - 1] \cdot (x_i - x_j)|\varepsilon_i, x_i],\end{aligned}$$

with $\delta \equiv \beta - \beta_0$. So

$$\begin{aligned}\psi(z_i;\beta_0) &= [2F(\varepsilon_i) - 1] \cdot (x_i - \mu_x) \\ &\equiv [2U_i - 1] \cdot (x_i - \mu_x),\end{aligned}$$

where $\mu_x \equiv E[x_j]$ and $U_i \equiv F(\varepsilon_i)$ has a $Uniform[0,1]$ distribution with mean 1/2 and variance 1/12, which means

$$
\begin{aligned}
V_0 &= Var(2U_i) \cdot Var(x_i) \\
&= \frac{1}{3}\Sigma_{xx}.
\end{aligned}
$$

Also,

$$
\begin{aligned}
D_0 &= -E\left[\frac{\partial \psi(z_i, \beta_0)}{\partial \beta'} \right] \\
&= -E[[2f(\varepsilon_i)] \cdot (x_i - x_j)(x_i - x_j)'] \\
&= -[2\tau_0] \cdot 2\Sigma_{xx},
\end{aligned}
$$

leading to the explicit expression for the asymptotic distribution of $\hat{\beta}$,

$$
\sqrt{N}\,(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, [48\tau_0^2]^{-1}\Sigma_{xx}^{-1}).
$$

■

4. If $\hat{f}(\varepsilon)$ is the kernel density estimator of $f(\varepsilon)$ using the residuals $\hat{\varepsilon}_i$, i.e.,

$$
\hat{f}(\varepsilon) = \frac{1}{Nh_N} \sum_{j=1}^{N} K\left(\frac{\varepsilon - \hat{\varepsilon}_j}{h_N} \right),
$$

with K a smooth, symmetric, nonnegative kernel with bounded derivatives, an estimator of the nuisance parameter τ_0 is

$$
\hat{\tau} \equiv \frac{1}{N} \sum_{i=1}^{N} \hat{f}(\hat{\varepsilon}_i).
$$

Show that if $h_N \to 0$ and $h_N^2 \cdot \sqrt{N} \to \infty$ as $N \to \infty$, then $\hat{\tau}$ is (weakly) consistent for τ_0. [First use a mean-value expansion to show that the residuals $\hat{\varepsilon}_i$ can be replaced by the true error terms ε_i, then use analogous bias-variance calculations to those for the kernel density estimator $\hat{f}(\varepsilon)$ itself.]

Answer: To show the residuals can be replace by the true error terms, use the expansion

$$\begin{aligned}
\hat{\tau} &\equiv \frac{1}{N^2 h_N} \sum_{i=1}^{N} \sum_{j=1}^{N} K\left(\frac{\hat{\varepsilon}_i - \hat{\varepsilon}_j}{h_N} \right) \\
&= \frac{1}{N^2 h_N} \sum_{i=1}^{N} \sum_{j=1}^{N} K\left(\frac{\varepsilon_i - \varepsilon_j}{h_N} \right) + \frac{1}{N^2 h_N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} K'\left(\frac{\tilde{\varepsilon}_i - \tilde{\varepsilon}_j}{h_N} \right) \cdot ((\hat{\varepsilon}_i - \hat{\varepsilon}_j) - (\varepsilon_i - \varepsilon_j)) \\
&= \frac{1}{N^2 h_N} \sum_{i=1}^{N} \sum_{j=1}^{N} K\left(\frac{\varepsilon_i - \varepsilon_j}{h_N} \right) + o_p(1) \\
&\equiv \tilde{\tau} + o_p(1),
\end{aligned}$$

where the $\{\tilde{\varepsilon}_i\}$ are intermediate values, and the third equality follows from the root-N consistency of $\hat{\beta}$ and

$$\begin{aligned}
\left| \frac{1}{N^2 h_N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} K'\left(\frac{\tilde{\varepsilon}_i - \tilde{\varepsilon}_j}{h_N} \right) \cdot ((\hat{\varepsilon}_i - \hat{\varepsilon}_j) - (\varepsilon_i - \varepsilon_j)) \right| &\leq \frac{1}{h_n^2 \sqrt{N}} \cdot \sqrt{N} \left\| \hat{\beta} - \beta_0 \right\| \cdot K_1 \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \| x_i - x_j \| \\
&= \frac{1}{h_n^2 \sqrt{N}} \cdot O_p(1),
\end{aligned}$$

where K_1 is an upper bound for $K'(\cdot)$.

The infeasible estimator $\tilde{\tau}$ is asymptotically equivalent to a U-statistic,

$$\begin{aligned}
\tilde{\tau} &\equiv \frac{1}{N^2 h_N} \sum_{i=1}^{N} \sum_{j=1}^{N} K\left(\frac{\varepsilon_i - \varepsilon_j}{h_N} \right) \\
&= \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{1}{h_N} K\left(\frac{\varepsilon_i - \varepsilon_j}{h_N} \right) + \frac{1}{N^2 h_N} \sum_{i=1}^{N} K(0) \\
&\equiv \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} p_n(\varepsilon_i, \varepsilon_j) + o(1),
\end{aligned}$$

with

$$p_n(\varepsilon_i, \varepsilon_j) \equiv \frac{1}{h_N} K\left(\frac{\varepsilon_i - \varepsilon_j}{h_N} \right)$$

being the kernel of the U-statistic. To verify consistency of $\tilde{\tau}$ (and thus of $\hat{\tau}$), we first need to show

$$E \| p_n(\varepsilon_i, \varepsilon_j) \|^2 = o(N),$$

from which it follows that $\tilde{\tau} - E[\tilde{\tau}] \xrightarrow{p} 0$ by the U-statistic theorem, and then to show the bias term $E[\tilde{\tau}] - \tau_0$ also converges to zero as $N \to \infty$. Direct calculation gives

$$
\begin{aligned}
E\|p_n(\varepsilon_i,\varepsilon_j)\|^2 &= E\left[\frac{1}{h_N^2}\left(K\left(\frac{\varepsilon_i-\varepsilon_j}{h_N}\right)\right)^2\right] \\
&= \frac{1}{h_N}\int\int\frac{1}{h_N}\left[K\left(\frac{u-v}{h}\right)\right]^2 f(v)f(u)dvdu \\
&= \frac{1}{h_N}\int\int[K(z)]^2 f(z-hu)f(u)dvdu \\
&\leq \frac{1}{h_N}K_0^2,
\end{aligned}
$$

where $|K(u)| \leq K_0$. Since it is assumed $h_N^{-1} = o(N^{1/4})$, it follows that $\tilde{\tau} - E[\tilde{\tau}] \xrightarrow{p} 0.$ To show that the bias vanishes as $N \to \infty$, the usual expectation calculations for kernel regression give

$$
\begin{aligned}
E[\tilde{\tau})] &= \int\int K(u)f(x-hu)f(x)dudx \\
&\to \int\int K(u)f(x)f(x)dudx \\
&= \int K(u)du \cdot \int[f(x)]^2 dx \\
&= \tau_0
\end{aligned}
$$

as $h \to 0$ by dominated convergence, because $\int K(u)du = 1$.■