Final Midterm Exam
(With Sketch of Answers)

Economics 241A
Spring 2004

May 10, 2004

Instructions: This is a 30 point exam, with weights given for each question; all subsections of each question have equal weight. 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.

1. (10 points) Suppose a scalar dependent variable y_{ij} for n_j individuals in J groups ($i = 1, ..., n_j$ and $j = 1, ..., J$) is assumed to satisfy a linear model

$$y_{ij} = x_j'\beta_0 + \varepsilon_{ij}$$

for some group-specific regressors x_j with error terms ε_{ij} that are independent across i and j and satisfy a conditional quantile restriction

$$\Pr\{\varepsilon_{ij} < 0 | x_j\} = \pi \qquad ((*))$$

for some π between zero and one. The ε_{ij} are assumed to be continuously distributed conditional on x_j, with conditional densities that are strictly positive everywhere (with probability one).

Define the π^{th} sample quantile $\hat{q}_j$ of y_{ij} for the j^{th} group as

$$\hat{q}_j = \arg\min_c \sum_{i=1}^{n_j} |\pi - 1\{y_{ij} < c\}| \cdot |y_{ij} - c|, \qquad j = 1, ..., J.$$

Under the assumption that $N = \sum_j n_j \to \infty$ with $\lim(n_j/N) \equiv p_j > 0$ for all j, find the form of the optimal weights for a weighted least-squares regression of $\hat{q}_j$ on x_j. These weights should be "optimal" in the sense that they minimize the asymptotic covariance matrix of the resulting estimator, which you should derive explicitly using the well-known form of the asymptotic distribution of the sample quantile $\hat{q}_j$. You should also show that this estimator achieves the relevant efficiency bound for the quantile restriction defined by (*).

In addition, propose a "feasible" version of this efficient estimator (using consistent estimators of the optimal weights). Finally, calculate the probability limit of the weighted least-squares estimator of β_0 when the linear regression function is misspecified – i.e., when

$$y_{ij} = g(x_j) + \varepsilon_{ij}$$

with $g(x)$ being nonlinear in x – and discuss the asymptotic behavior of the feasible estimator under this misspecification.

Answer: This weighted least squares estimator is actually a "minimum distance" estimator,

$$\hat{\beta} = \arg\min_{b}(\hat{q} - Xb)'W_0(\hat{q} - Xb),$$

where

$$X \equiv \begin{bmatrix} x_1' \\ ... \\ x_J' \end{bmatrix}, \qquad \hat{q} \equiv \begin{bmatrix} \hat{q}_1 \\ ... \\ \hat{q}_J \end{bmatrix}, \qquad W_0 \equiv diag[w_j],$$

for $w_j \equiv w(x_j)$ the weights assigned to group j. Since there are only a finite number J of regressors x_j, we will derive the asymptotic distribution of $\hat{\beta}$ conditional on X, i.e., treating the regressors x_{ij} for each observation as having a degenerate distribution around the cell-specific value x_j, with X assumed to be a fixed (full rank) matrix, as with the classical linear model.

This estimator will behave just like a GMM estimator, replacing the average moment function $\bar{m}(b)$ with $\hat{q} - Xb$. Using the same arguments as to derive the asymptotic distribution of GMM, this minimum distance estimator will have the asymptotic distribution

$$\sqrt{N}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(0, [M_0'W_0M_0]^{-1}M_0'W_0V_0W_0M_0[M_0'W_0M_0]^{-1}),$$

where now

$$\begin{aligned} M_0 &\equiv E\left[\frac{\partial(q - Xb)}{\partial b'}\right]_{b=\beta_0} \\ &= X' \end{aligned}$$

and V_0 is the asymptotic covariance matrix of the J sample quantiles,

$$\sqrt{N}(\hat{q} - X\beta_0) \xrightarrow{d} \mathcal{N}(0, V_0).$$

By the independence of the observations across samples and groups, the matrix V_0 is diagonal, with

$$V_0 = diag\left[\frac{\pi(1-\pi)}{p_j[f_j]^2}\right],$$

for $f_j \equiv f(0|x_j)$ the conditional density of ε_{ij} in group j, evaluated at the π^{th} quantile (assumed to be zero), and $p_j \equiv \lim(n_j/N)$ as above.

Obviously the best choice of W_0 is

$$W_0^* = V_0^{-1},$$

which yields the asymptotic distribution

$$\sqrt{N}(\hat{\beta}^* - \beta_0) \xrightarrow{d} \mathcal{N}(0, [M_0'V_0^{-1}M_0]^{-1}) = \mathcal{N}(0, [X'V_0^{-1}X]^{-1})$$

with the smallest asymptotic covariance matrix. The best weights are thus proportional to $p_j[f_j]^2$, i.e., weight each cell by the relative fraction of observations in the cell and the square of the cell density at the p^{th} quantile.

A feasible version of this weighted least-squares estimator would use weights $\hat{p}_j[\hat{f}_j]^2$, where $\hat{p}_j \equiv n_j/N$ (obviously consistent for p_j) and $\hat{f}_j$ is some nonparametric (e.g., kernel) estimator of the density of y_{ij} in cell j, evaluated at the relevant quantile $\hat{q}_j$. As long as the density estimator is consistent for each cell, the feasible version of $\hat{\beta}^*$ will have the same asymptotic distribution as its infeasible counterpart, for the same reason that only the probability limit of the weight matrix matters for GMM estimation.

The efficiency bound stuff is a little tricky. Rewriting the quantile restriction (*) in the usual conditional moment restriction form

$$E[(\pi - 1\{y_{ij} - x'_{ij}\beta_0\})|x_{ij}] \equiv E[u(y_{ij}, x_{ij}, \beta_0)|x_{ij}] = 0,$$

the semiparametric efficiency bound for the asymptotic covariance matrix of a regular estimator of β_0 under (*) is

$$\begin{aligned} B &\equiv \left[\lim \frac{1}{N} \sum_{i,j} \left[E\left(D(x_{ij})'[\Sigma(x_{ij})]^{-1} D(x_{ij})\right)\right]\right]^{-1} \\ &= \left[\sum_j p_j \left(D(x_j)'[\Sigma(x_j)]^{-1} D(x_j)\right)\right]^{-1} \end{aligned}$$

under the assumption that X is nonrandom, with

$$\begin{aligned} D(x_j) &\equiv \frac{\partial}{\partial b'} E[u(y_{ij}, x_{ij}, b)|x_{ij} = x_j]_{b=\beta_0} \\ &= \frac{\partial}{\partial b'}[\pi - \Pr\{y_{ij} \leq x'_{ij} b|x_{ij} = x_j\}]_{b=\beta_0} \\ &= -f(0|x_j)x'_j \end{aligned}$$

and

$$\begin{aligned} \Sigma(x_j) &\equiv V[u((y_{ij}, x_j, b)|x_j] \\ &= V[(\pi - 1\{y_{ij} - x'_{ij}\beta_0\})|x_{ij} = x_j] \\ &= \pi(1-\pi). \end{aligned}$$

So

$$\begin{aligned} B &= \left[\sum_j p_j(-f_j x'_j)'[\pi(1-\pi)]^{-1}(-f_j x'_j)\right]^{-1} \\ &= \left[\sum_j [\frac{p_j[f_j]^2}{\pi(1-\pi)} \pi(1-\pi)]^{-1} x_j x'_j\right]^{-1} \\ &= [X'V_0^{-1}X]^{-1}, \end{aligned}$$

the asymptotic covariance matrix of β^*, which is thus efficient under the moment restriction $(*)$. [Whew!]

To make matters worse, the question also asks about the distribution of the feasible estimator

$$\begin{aligned} \tilde{\beta} &\equiv \arg\min_b (\hat{q} - Xb)'\hat{V}^{-1}(\hat{q} - Xb) \\ &= \left[\sum_j \hat{p}_j[\hat{f}_j]^2 x_j x'_j\right]^{-1} \sum_j \hat{p}_j[\hat{f}_j]^2 x_j y_j \end{aligned}$$

under misspecification of the linear form of the quantiles, i.e., when $p\lim \hat{q} \equiv g \neq X\beta_0$ for any β_0. In this case, clearly

$$\tilde{\beta} \xrightarrow{p} \delta_0 \equiv \left[\sum_j p_j[f_j]^2 x_j x'_j\right]^{-1} \sum_j p_j[f_j]^2 x_j g(x_j),$$

a weighted linear projection of the cell quantiles $g(x_j)$ on the x_j.

Furthermore, the rate of convergence of the feasible estimator $\tilde{\beta}$ to its probability limit δ_0 will be governed by the rate of convergence of the estimated densities $\{\hat{f}_j\}$ to their true values, which is slower than $\sqrt{N}$. Using suboptimal weights that don't involve the conditional density estimators (e.g., using just $\hat{p}_j$ as weights) would yield estimators with the usual $\sqrt{N}$ convergence rates under misspecification, though of course these estimators converge to different weighted linear projection. This point (and much of the setup and results of this problem) were discussed by Gary Chamberlain (1994), "Quantile Regression, Censoring and the Structure of Wages," in Sims, C., ed., *Advances in Econometrics: Proceedings from the Sixth World Congress* (Cambridge U. Press), in case you want to follow up on quantile minimum distance estimation.■

2. (20 points) For the censored regression model with a single (scalar) regressor,

$$y_i = \max\{0, x_i \cdot \beta_0 + u_i\}, \qquad i = 1, ..., N,$$

suppose that the error terms u_i are symmetrically distributed about zero conditionally, not on x_i, but on some q-dimensional vector of "instrumental variables" z_i. The regressors x_i are assumed to be related to the instruments z_i by a linear reduced form:

$$x_i = z_i'\pi_0 + v_i,$$

where the error terms u_i and v_i are jointly continuous and symmetrically distributed given z_i – more precisely, for any fixed numbers α and λ, the linear combination $\alpha u_i + \lambda v_i$ is symmetric about zero given z_i.

A. Consider the following two-stage procedure:first, estimate π_0 by least squares, then estimate β_0 by symmetrically-censored least squares (SCLS) estimation, after replacing the "endogenous" regressors x_i by their fitted values $\hat{x}_i \equiv z_i'\hat{\pi}$. Thus, the second-stage estimator $\hat{\beta}$ will be the (consistent) solution to the equation

$$\begin{aligned} 0 &= \frac{1}{N}\sum_{i=1}^{n} 1\{\hat{x}_i \cdot \hat{\beta} > 0\} \cdot \min\{y_i - \hat{x}_i \cdot \hat{\beta}, \hat{x}_i \cdot \hat{\beta}\} \cdot \hat{x}_i \\ &\equiv \frac{1}{N}\sum_{i=1}^{n} \psi(y_i, z_i, \hat{\pi}, \hat{\beta}), \end{aligned}$$

where

$$\hat{\pi} \equiv \left[\frac{1}{N}\sum_{i=1}^{n} z_i z_i'\right]^{-1} \left[\frac{1}{N}\sum_{i=1}^{n} z_i x_i'\right].$$

Assuming this estimator is consistent, and assuming i.i.d. sampling, all needed moments exist, etc., derive the asymptotic distribution of the second-stage estimator $\hat{\beta}$. (Don't check regularity conditions, stochastic equicontinuity, etc. – just do the calculations.)

Answer: The first-stage LS estimator $\hat{\pi}$ satisfies the standard asymptotic linearity relation

$$\begin{aligned} \sqrt{N}(\hat{\pi} - \pi_0) &= \frac{1}{\sqrt{N}}\sum_i v_i D^{-1} z_i + o_p(1) \\ &\xrightarrow{d} \mathcal{N}(0, D^{-1}CD^{-1}), \end{aligned}$$

for

$$\begin{aligned} D &\equiv E[z_i z_i'], \\ C &= V[v_i z_i] = E[v_i^2 z_i z_i']. \end{aligned}$$

Writing the second stage estimator $\hat{\beta}$ as the solution to

$$0 = \bar{\Psi}(\hat{\pi}, \hat{\beta}) \equiv \frac{1}{N} \sum_{i=1}^{n} \psi(y_i, z_i, \hat{\pi}, \hat{\beta}),$$

we'll just assume the stochastic equicontinuity condition

$$\sqrt{N}\left[\left(\bar{\Psi}(\hat{\pi}, \hat{\beta}) - \bar{\Psi}(\pi_0, \beta_0)\right) - E\left(\bar{\Psi}(\pi, \beta) - \bar{\Psi}(\pi_0, \beta_0)\right)_{\pi=\hat{\pi}, \beta=\hat{\beta}}\right] = o_p(1)$$

holds. Here

$$\begin{aligned} \psi(y_i, z_i, \pi, \beta) &\equiv 1\{z_i'\pi \cdot \beta > 0\} \cdot \min\{\max\{0, x_i \cdot \beta_0 + u_i\} - z_i'\pi\beta, z_i'\pi\beta\} \cdot z_i'\pi \\ &= 1\{z_i'\pi \cdot \beta > 0\} \cdot \min\{\max\{-z_i'\pi\beta, u_i + v_i'\beta_0 - z_i'(\pi\beta - \pi_0\beta_0)\}, z_i'\pi\beta\} \cdot z_i'\pi \end{aligned}$$

Since

$$\begin{aligned} \bar{\Psi}(\pi_0, \beta_0) &= \frac{1}{N} \sum_{i=1}^{n} 1\{z_i'\pi_0 \cdot \beta_0 > 0\} \cdot \min\{y_i - z_i'\pi_0 \cdot \beta_0, z_i'\pi_0 \cdot \beta_0\} \cdot z_i'\pi_0 \\ &= \frac{1}{N} \sum_{i=1}^{n} 1\{z_i'\pi_0\beta_0 > 0\} \cdot \min\{\max\{z_i'\pi_0\beta_0, u_i + v_i'\beta_0\}, z_i'\pi_0\beta_0\} \cdot z_i'\pi_0, \\ &\equiv \frac{1}{N} \sum_{i=1}^{n} 1\{z_i'\pi_0\beta_0 > 0\} \cdot \min\{\max\{z_i'\pi_0\beta_0, \varepsilon_i\}, z_i'\pi_0\beta_0\} \cdot z_i'\pi_0, \end{aligned}$$

which is an odd function of

$$\varepsilon_i \equiv u_i + v_i'\beta_0.$$

Since ε_i is symmetric about zero by the joint symmetry assumption on u_i and v_i – it follows that

$$E\left[\bar{\Psi}(\pi_0, \beta_0)\right] = 0 = \bar{\Psi}(\hat{\pi}, \hat{\beta}).$$

These equalities, combined with the stochastic equiwhatever condition, yield the relation

$$\sqrt{N} E\left(\bar{\Psi}(\pi, \beta)\right)_{\pi=\hat{\pi}, \beta=\hat{\beta}} = \sqrt{N}\bar{\Psi}(\pi_0, \beta_0) + o_p(1). \qquad ((**))$$

A Taylor's series expansion of the left-hand side of this expression yields

$$\sqrt{N} E\left(\bar{\Psi}(\pi, \beta)\right)_{\pi=\hat{\pi}, \beta=\hat{\beta}} = J_0\sqrt{N}(\hat{\pi} - \pi_0) + H_0\sqrt{N}(\hat{\beta} - \beta_0) + o_p(1),$$

where the $(1 \times q)$ Jacobian matrix J_0 is defined as

$$\begin{aligned} J_0 &\equiv \frac{\partial}{\partial \pi'} E\left(\bar{\Psi}(\pi_0, \beta_0)\right) \\ &= \frac{\partial}{\partial \pi'} E\left[1\{z_i'\pi\beta > 0\} \cdot \min\{\max\{-z_i'\pi\beta, \varepsilon_i - z_i'(\pi\beta - \pi_0\beta_0)\}, z_i'\pi\beta\} \cdot z_i'\pi\right]_{\beta=\beta_0, \pi=\pi_0} \\ &= \beta_0 E\left[1\{|\varepsilon_i| < z_i'\pi_0\beta_0\}\left(z_i'\pi_0\right) z_i'\right] \end{aligned}$$

and the (1×1) Hessian "matrix" H_0 is

$$
\begin{aligned}
J_0 &\equiv \frac{\partial}{\partial \beta} E\left(\bar{\Psi}(\pi_0, \beta_0)\right) \\
&= \frac{\partial}{\partial b'} E\left[1\{z_i'\pi\beta > 0\} \cdot \min\{\max\{-z_i'\pi\beta, \varepsilon_i - z_i'(\pi\beta - \pi_0\beta_0)\}, z_i'\pi\beta\} \cdot z_i'\pi\right]_{\beta=\beta_0, \pi=\pi_0} \\
&= E\left[1\{\left|u_i + v_i'\beta_0\right| < z_i'\pi_0\beta_0\}(z_i'\pi_0)^2\right].
\end{aligned}
$$

Assuming $H_0 \neq 0$, solving out for $\sqrt{N}(\hat{\beta} - \beta_0)$ in (**) yields

$$
\sqrt{N}(\hat{\beta} - \beta_0) = \frac{-1}{H_0}\left[\sqrt{N}\bar{\Psi}(\pi_0, \beta_0) - J_0\sqrt{N}(\hat{\pi} - \pi_0)\right] + o_p(1).
$$

Finally, substituting in the expressions for $\bar{\Psi}(\pi_0, \beta_0)$ and $\sqrt{N}(\hat{\pi} - \pi_0)$ gives

$$
\begin{aligned}
\sqrt{N}(\hat{\beta} - \beta_0) &= \frac{1}{\sqrt{N}} \sum_{i=1}^{n} \xi_i + o_p(1) \\
&\xrightarrow{d} \mathcal{N}(0, Var(\xi_i)),
\end{aligned}
$$

for

$$
\begin{aligned}
\xi_i &\equiv \frac{\psi(y_i, z_i, \pi_0, \beta_0) - J_0 D^{-1} z_i \cdot v_i}{H_0} \\
&= \frac{[1\{z_i'\pi_0\beta_0 > 0\} \cdot \min\{\max\{-z_i'\pi_0\beta_0, \varepsilon_i\}, z_i'\pi_0\beta_0\} \cdot z_i'\pi_0] - J_0 D^{-1} z_i \cdot v_i}{H_0}.
\end{aligned}
$$

■.

B. Suppose instead that the reduced form for x_i' was substituted into the model for the dependent variable y_i, and the reduced-form parameter $\delta_0 \equiv \pi_0\beta_0$ for the resulting censored regression model for y_i and z_i was estimated using SCLS. Given the SCLS estimator $\hat{\delta}$ and the least-squares estimator $\hat{\pi}$ from the first stage, propose an efficient way to combine these two estimators to obtain an estimator of β_0, and derive its asymptotic distribution. Discuss the sense in which this estimator is efficient. (As above, don't bother listing or verifying regularity conditions.)

Answer: Using the relationship $\delta_0 \equiv \pi_0\beta_0$, we can stack the asymptotic linearity relationships for $\hat{\pi}$ and $\hat{\delta}$, the solution of

$$
\bar{\Psi}(\hat{\delta}, 1) = 0,
$$

as

$$
\begin{aligned}
\sqrt{N}\begin{pmatrix} \hat{\pi} - \pi_0 \\ \hat{\delta} - \beta_0\pi_0 \end{pmatrix} &= \frac{1}{\sqrt{N}} \sum_{i=1}^{n} \begin{pmatrix} v_i D^{-1} z_i \\ H_0^{-1}\psi(y_i, z_i, \delta_0, 1) \end{pmatrix} + o_p(1) \\
&\equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{n} \zeta_i + o_p(1) \\
&\xrightarrow{d} \mathcal{N}(0, V_0),
\end{aligned}
$$

with $V_0 = V[\zeta_i]$. Again, a minimum-distance approach to joint estimation of the $q+1$ structural parameters π_0 and β_0 using the $2q$ "reduced form" estimators $\hat{\pi}$ and $\hat{\delta}$ will produce "efficient" estimators. The efficient minimum distance estimators π^* and β^* are defined as

$$\begin{pmatrix} \pi^* \\ \beta^* \end{pmatrix} = \arg\min_{\pi,\beta} (\hat{\pi}' - \pi', \hat{\delta}' - \beta\pi') V_0^{-1} \begin{pmatrix} \hat{\pi} - \pi \\ \hat{\delta} - \beta\pi \end{pmatrix},$$

whose asymptotic distribution will have the "usual" form

$$\sqrt{N} \begin{pmatrix} \pi^* - \pi_0 \\ \beta^* - \beta_0 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, [M_0' V_0^{-1} M_0]^{-1}).$$

In this expression, the $2q \times (q+1)$ matrix M_0 will be

$$\begin{aligned} M_0 &\equiv \frac{\partial \begin{pmatrix} \hat{\pi} - \pi_0 \\ \hat{\delta} - \beta_0\pi_0 \end{pmatrix}}{\partial(\pi_0', \beta_0)} \\ &= -\begin{bmatrix} I_q & 0 \\ \beta_0 \cdot I_q & \pi_0 \end{bmatrix}. \end{aligned}$$

This estimator will yield the efficient combination of $\hat{\pi}$ and $\hat{\delta}$, but will not be globally efficient, since those reduced-form estimators need not be jointly efficient under the assumption of conditional symmetry.■