Problem Set #2

ECONOMICS 241A
SPRING 2010

Due April 27

1. 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.

2. Suppose y_i satisfies a Box-Cox transformation model with right-censoring: that is,

$$y_i = min\{h(\mathbf{x}_i'\beta_0 + \varepsilon_i; \lambda_0), c\},$$

where the constant c is known and the "inverse Box-Cox tranform" $h(\cdot)$ is defined as

$$h(y; \lambda) \equiv (1 + \lambda y)^{1/\lambda} \qquad \text{if } \lambda \neq 0, \qquad h(y; 0) \equiv exp\{y\}.$$

A. Assume ε_i is independent of $\mathbf{x}_i$ with c.d.f. and density functions $F(\cdot; \tau_0)$ and $f(\cdot; \tau_0)$, respectively; also, assume that $\Pr\{\mathbf{x}_i'\beta_0 + \varepsilon_i > 0\} = 1$ for all possible values of β_0 (so that y_i is well-defined with probability one). Under these assumptions, derive the average log-likelihood for an i.i.d. sample of N observations on y_i and $\mathbf{x}_i$.

B. Now suppose that there is no right-censoring (i.e., $c = \infty$), but that the parametric form of the density for ε_i is unknown, and that ε_i is known only to satisfy the conditional moment restriction

$$E[\varepsilon_i | x_i] = 0$$

(so that ε_i is no longer required to be independent of x_i). What is the smallest asymptotic covariance matrix of a regular $\sqrt{n}$-consistent estimator of β_0 and λ_0 that uses only this restriction? What is the form of the "optimal instrument" vector for this conditional moment restriction?

C. Assuming now that $c < \infty$ and that ε_i is continuously distributed with conditional median zero,

$$E[sgn\{\varepsilon_i\} | x_i] = 0,$$

propose a $\sqrt{n}$-consistent estimator of the parameters that only exploits this restriction, and derive the form of its asymptotic distribution.

3. (This was a 40 point midterm exam, with equal weights for each question. The answers were due no later than 25 hours after the exam was picked up; students were permitted to consult and cite any lecture notes and any of the references on the syllabus, but could not cite any other outside source, nor discuss the exam with anyone other than the instructor before submitting the answers.)

Suppose a sample of N i.i.d. observations on a random vector z_i (with an r-dimensional subvector vector of regressors x_i) satifies a *conditional moment restriction*

$$0 = E[u_i | x_i] \equiv E[u(z_i, \theta_0) | x_i] \equiv \rho(x_i, \theta_0),$$

where $u(z_i, \theta)$ is some q-dimensional vector of known functions of the (i.i.d.) random vector z_i and $\theta \in \Theta \subset R^p$. A recent *Econometrica* article by Ai and Chen (2004) proposed an estimator of θ_0 based upon a nonparametric estimator

$$\hat{\rho}(x_i, \theta) \equiv \hat{E}[u(z_i, \theta_0) | x_i]$$

of the conditional moment function $\mu(\cdot)$; their estimator minimizes the sample average (over x_i) of a quadratic form in $\hat{\mu}(x_i, \theta)$, with "weight matrix" possibly estimated and depending upon x.

Consider a simpler special case, where u_i, x_i, and θ are all scalar (i.e., $p = q = r = 1$), and, further, that $u(z_i, \theta)$ is bounded (with probability one) and x_i has a marginal $Uniform(0, 1)$ distribution. A kernel estimator of the function $\mu(x, \theta)$ for this special case would be

$$\hat{\rho}(x, \theta) \equiv \frac{1}{Nh} \sum_{j=1}^{N} K\left(\frac{x - x_j}{h}\right) \cdot u(z_j, \theta),$$

where $K(u)$ is a smooth, bounded, symmetric kernel function that integrates to one and $h = h_n$ is a positive bandwidth, declining to zero with N. (Since the density function $f(x)$ of x_i is identically one on its support, the usual denominator for the Nadaraya-Watson kernel regression estimator is not needed here.) With this estimator of $\mu(\cdot)$, an estimator $\hat{\theta}$ of θ_0 might be defined as a solution to the estimating equations

$$\begin{aligned} 0 &= \frac{1}{N} \sum_{i=1}^{N} w(x_i) \cdot \hat{\mu}(x_i, \hat{\theta}) \\ &\equiv \hat{M}_N(\hat{\theta}), \end{aligned}$$

where $w(x_i)$ is some bounded, scalar weighting function (which would undoubtedly need to be estimated in practice).

A. Under what (sufficient) conditions will $\hat{\theta} \xrightarrow{p} \theta_0$ as $N \to \infty$? (Your conditions should be "high-level," i.e., involving only general conditions about Θ, $\hat{M}_N(\theta)$, and its limiting value, and you need not verify those conditions.)

B. Give conditions under which, for each value of θ, $\hat{M}_N(\theta)$ is asymptotically equivalent to a "smoothed U-statistic" of the form

$$M_N^*(\theta) \equiv \left(\begin{array}{c} N \\ 2 \end{array} \right)^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} p_N(z_i, z_j, \theta)$$

for some "kernel" $p_N(\cdot)$ that is symmetric in i and j (i.e., $p_N(z_i, z_j, \theta) = p_N(z_{j,} z_i, \theta)$). That is, find the form of $p_N(\cdot)$ and conditions on h_N so that

$$\sqrt{N}\left(\hat{M}_N(\theta) - M_N^*(\theta)\right) \xrightarrow{p} 0.$$

C. Defining

$$r_N(z_i, \theta) \equiv E[p_N(z_i, z_j, \theta)|z_i]$$

and

$$\begin{aligned} \mu_N(\theta) &\equiv E[r_N(z_i, \theta)] \\ &= E[p_N(z_i, z_j, \theta)], \end{aligned}$$

the projection of the U-statistic $M_N^*(\theta)$ is defined to be

$$\tilde{M}_N(\theta) \equiv \mu_N(\theta) + \frac{2}{N}\sum_{i=1}^{N}\left[r_N(z_i, \theta) - \mu_N(\theta)\right].$$

What conditions on h_N ensure that

$$\sqrt{N}\left(M_N^*(\theta) - \tilde{M}_N(\theta)\right) \xrightarrow{p} 0?$$

D. Assuming the functions $K(\cdot)$ and $u(\cdot)$ are sufficiently smooth, calculate

$$r(z_i, \theta) \equiv \lim_{N\to\infty} r_N(z_i, \theta)$$

and

$$\mu(\theta) \equiv \lim_{N\to\infty} E[r_N(z_i, \theta)] = E[r(z_i, \theta)].$$

E. Suppose (without proof!) it can be shown that

$$\max_{\theta\in\Theta}\left\|\sqrt{N}\left(\hat{M}_N(\theta) - M_N(\theta)\right)\right\| \xrightarrow{p} 0,$$

where $M_N(\theta)$ is the limit of the projection of the U-statistic, i.e., where

$$M_N(\theta) \equiv \mu(\theta) + \frac{2}{N}\sum_{i=1}^{N}\left[r(z_i, \theta) - \mu(\theta)\right],$$

with $r(\cdot)$ and $\mu(\cdot)$ defined in the previous problem. Under this condition, use the usual Taylor's series arguments (ignoring the negligible remainder terms) to show that the estimator $\hat{\theta}$ that solves

$$0 = \hat{M}_N(\hat{\theta})$$

is $\sqrt{N}$-consistent and asymptotically normal,

$$\sqrt{N}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N(0, V_0)$$

for some (scalar) V_0, and find the form of V_0, which should involve the random variables

$$\sigma_0^2(x_i) \equiv Var[u(z_i, \theta_0)|x_i]$$

and

$$d_0(x_i) \equiv E\left[\frac{\partial u(z_i, \theta_0)}{\partial \theta}|x_i\right].$$

F. For what choice of weight function $w(x_i)$ will the corresponding estimator $\hat{\theta}$ be asymptotically efficient? Explain.

G. For the further special case of a linear model, i.e., $z_i = (y_i, x_i)'$ with

$$y_i = \theta_0 \cdot x_i + \varepsilon_i,$$

suppose

$$u(z_i, \theta) = y_i - \theta \cdot x_i,$$

i.e., the error terms ε_i have conditional mean equal to zero given x_i. Derive the form of V_0 for this case. Under what conditions will a constant weighting function $w(x_i) = 1$ be optimal?

H. How do your answers to problem #7 change if the errors are restricted to have conditional median zero rather than conditional mean zero, i.e., if the moment function is changed to

$$u(z_i, \theta) = sgn\{y_i - \theta \cdot x_i\}?$$

[You can ignore the fact that this moment function is discontinuous in θ, and thus may violate some regularity conditions imposed earlier.]