Final Midterm Exam
(with Answers)

ECONOMICS 241A
SPRING 2005

May 9, 2005

Instructions: This is a 35 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) Consider the linear model

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

and suppose that the unobservable error term ε_i satisfies *both* a conditional mean restriction

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

and a conditional median restriction

$$E[sgn(\varepsilon_i)|x_i] = 0.$$

Assuming that ε_i is continuously distributed conditional on x_i, with a conditional density $f_{\varepsilon|x}(\varepsilon|x_i)$ that has lots of derivatives and moments, derive an (infeasible) efficient estimator of β_0 under these two restrictions, and give an expression for the form of its asymptotic covariance matrix. (Assume the relevant stochastic equicontinuity condition holds, so that the order of differentiation and expectation can be interchanged if necessary.)

Answer: Stacking up the two conditional moment restrictions into a 2×1 vector $u(z_i, \beta)$ of conditional moment functions

$$u(z_i,\beta) \equiv \begin{bmatrix} y_i - x_i'\beta \\ sgn\{y_i - x_i'\beta\} \end{bmatrix},$$

the efficient estimator of β_0 under the conditional moment restriction $E[u(z_i, \beta_0)|x_i] = 0$ is the optimal instrumental variables estimator $\hat{\beta}_{OIV}$ which solves

$$0 = \frac{1}{n}\sum_{\iota=1}^{n} h^*(x_i)u(z_i, \hat{\beta}_{OIV}),$$

where $h^*(x_i)$ is the optimal instrumental variables vector, defined (after interchanging derivatives and expectations) as

$$\begin{aligned} h^*(x_i) &= \left[\frac{\partial E[u(z_i,\theta_0)|x_i]}{\partial \theta'}\right]' \cdot [Var(u(z_i,\beta_0)|x_i]^{-1} \\ &\equiv D(x_i)' \cdot [\Sigma(x_i)]^{-1}. \end{aligned}$$

For this problem, $D(x_i)$ is a $2 \times p$ matrix of the form

$$D(x_i) = \begin{bmatrix} -x_i' \\ -2f_{\varepsilon|x}(0|x_i)x_i' \end{bmatrix},$$

while

$$\begin{aligned} \Sigma(x_i) &= \begin{bmatrix} E[\varepsilon_i^2|x_i] & E[\varepsilon_i sgn\{\varepsilon_i\}|x_i] \\ E[\varepsilon_i sgn\{\varepsilon_i\}|x_i] & E[(sgn\{\varepsilon_i\})^2|x_i] \end{bmatrix} \\ &\equiv \begin{bmatrix} \sigma^2(x_i) & \tau(x_i) \\ \tau(x_i) & 1 \end{bmatrix} \qquad \Longrightarrow \\ [\Sigma(x_i)]^{-1} &= [\sigma^2(x_i) - [\tau(x_i)]^2]^{-1} \begin{bmatrix} 1 & -\tau(x_i) \\ -\tau(x_i) & \sigma^2(x_i) \end{bmatrix}, \end{aligned}$$

and

$$h^*(x_i) = \begin{bmatrix} \frac{2f_{\varepsilon|x}(0|x_i)\tau(x_i)-1}{\sigma^2(x_i)-[\tau(x_i)]^2} & \frac{2f_{\varepsilon|x}(0|x_i)\sigma^2(x_i)-\tau(x_i)}{\sigma^2(x_i)-[\tau(x_i)]^2}) \end{bmatrix} \otimes x_i.$$

The corresponding asymptotic covariance matrix of the optimal IV estimator is

$$\left[E\left(D(x_i)'[\Sigma(x_i)]^{-1}D(x_i)\right)\right]^{-1} = \left[E\left(\left[\frac{1 - 4f_{\varepsilon|x}(0|x_i)\tau(x_i) + \left[2f_{\varepsilon|x}(0|x_i)\right]^2\sigma^2(x_i)}{\sigma^2(x_i) - [\tau(x_i)]^2}\right] x_i x_i'\right)\right]^{-1}.$$

■

2. (10 points) Consider the nonparametric regression model

$$y_t = g(x_t) + \varepsilon_t, \qquad t = 1, ..., T,$$

where x_t and y_t are scalar, jointly-continuous random variables with finite variances, joint density function $f_{x,y}(x, y)$, marginals $f_x(x)$ and $f_y(y)$, and with $E[\varepsilon_t|x_t] \equiv 0$ (that is, $g(x_t) \equiv E[y_t|x_t]$). An estimator for the value of $g(x)$ at a fixed value $x = x_0$ is the uniform kernel regression estimator

$$\widehat{g}(x_0) \equiv \left[\frac{1}{T}\sum_{t=1}^{T} w_{tT} \cdot y_t\right] \cdot \left[\frac{1}{T}\sum_{t=1}^{T} w_{tT}\right]^{-1},$$

where the "local weight" w_{tT} takes the form

$$w_{tT} \equiv \frac{1}{h_T} \cdot 1\{|x_t - x_0| \leq \frac{h_T}{2}\}$$

and $\{h_T\}$ is a nonrandom sequence of bandwidths. Assume that

1. **i.** the functions $g(x)$ and the marginal density $f_x(x)$ of x_t have lots of continuous derivatives at $x = x_0$ (as many as needed);

   ii. ε_t and x_s are statistically independent for all t and s;

   iii. x_t is an i.i.d. sequence with $f_x(x_0) > 0$; and

   iv. ε_t is a (weakly) stationary process with autocovariance sequence $\gamma_\varepsilon(s)$ that is absolutely summable, i.e.

$$\sum_{s=0}^{\infty} |\gamma_\varepsilon(s)| < \infty.$$

(a) Consider the numerator of $\widehat{g}(x_0)$,

$$\hat{n}(x_0) \equiv \widehat{g}(x_0) \cdot \widehat{f}_x(x_0) \equiv \frac{1}{T}\sum_{t=1}^{T} w_{tT} \cdot y_t,$$

where$\widehat{f}_x(x_0)$ is the kernel density estimator of $f_x(x_0)$,

$$\widehat{f}_x(x_0) \equiv \frac{1}{T}\sum_{t=1}^{T} w_{tT}.$$

Give an expression for the variance of the numerator term $\widehat{n}(x_0)$, and show that, as $h_T \to 0$, the leading (largest) term in the expansion of the variance in powers of h_T does not depend upon the autocovariances $\gamma_\varepsilon(s)$ for $s \neq 0$.

Answer: The variance of the numerator term $\hat{n}(x_0)$ can be written as

$$\begin{aligned} Var[\hat{n}(x_0)] &= Var\left[\frac{1}{T}\sum_{t=1}^{T} w_{tT} \cdot y_t\right] \\ &= \frac{1}{T^2}\sum_{t=1}^{T} Var\left[w_{tT} \cdot y_t\right] + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s\neq i, s=1}^{T} Cov\left[w_{tT} \cdot y_t, w_{sT} \cdot y_s\right]. \end{aligned}$$

The first term is the usual expression for the variance of the numerator of the kernel regression estimator,

$$\begin{aligned} \frac{1}{T^2}\sum_{t=1}^{T} Var\left[w_{tT} \cdot y_t\right] &= \frac{1}{T} Var\left[w_{tT} \cdot y_t\right] \\ &= \frac{[\sigma^2 + g(x_0)^2] f(x)}{Th} \int \left[K\left(u\right)\right]^2 du + o\left(\frac{1}{Th}\right) \\ &= O\left(\frac{1}{Th}\right), \end{aligned}$$

for $\sigma^2 \equiv \gamma_\varepsilon(0) \equiv Var[y_t | x_t = x_0]$.And since w_{tT} and w_{sT} are independent (by serial independence of the $\{x_t\}$) with $E[w_{tT}] = \int_{-1/2}^{1/2} f(x - hu)du = O(1)$, the second term satisfies

$$\begin{aligned} \left|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s\neq i, s=1}^{T} Cov\left[w_{tT} \cdot y_t, w_{sT} \cdot y_s\right]\right| &\leq \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} E\left[w_{tT}\right] \cdot E[w_{sT}] \cdot \left|Cov[\varepsilon_t, \varepsilon_s]\right| \\ &= O(1) \cdot \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} |\gamma(t-s)| \\ &\leq O(1) \cdot \frac{1}{T}\sum_{s=-\infty}^{\infty} |\gamma(s)| \\ &= O\left(\frac{1}{T}\right) \\ &= o\left(\frac{1}{Th}\right). \end{aligned}$$

So the variance of the numerator term $\hat{n}(x_0)$ has the same form as if the error terms ε_t were serially uncorrelated.■

(b) Find conditions on the bandwidth sequence h_T under which $\hat{g}(x_0)$ is weakly consistent. Try to make your assumptions as weak (general) as possible.

Answer: This is too easy – since the expression for the bias of the numerator $\hat{n}(x_0)$ does not involve the dependence (correlation) structure of the data, and since the variance of $\hat{n}(x_0)$ has the same form as if the errors ε_t were serially uncorrelated, it converges to $n(x_0)$ in probability under the usual conditions $h \to 0$, $Th \to \infty$ as $T \to \infty$, which also suffice for $\hat{f}(x_0) \xrightarrow{p} f(x_0)$ and thus $\hat{g}(x_0) \xrightarrow{p} g(x_0)$.■

3. (15 points) Suppose that economic theory suggests that a latent dependent variable y_i^* satisfies a classical linear model

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

but that you do not observe y_i^* over its entire range. Instead, you observe a random sample of size n of y_i and x_i, where

$$\begin{aligned} y_i &\equiv \tau_i(y_i^*) \\ &= 0 \qquad if \qquad y_i^* \leq 0, \\ &= y_i^* \qquad if \qquad 0 < y_i^* \leq L_i, \\ &= L_i \qquad if \qquad L_i < y_i^* \leq U_i, \text{ and} \\ &= y_i^* - (U_i - L_i) \qquad if \qquad U_i < y_i^*. \end{aligned}$$

That is, the latent variable y_i^* is observed unless it is less than zero or in the interval (L_i, U_i), where the threshold variables L_i and $U_i > L_i > 0$ are assumed known for all i.

A. Assuming that ε_i is normally distributed with zero mean and unknown variance σ_0^2, and is independent of x_i, derive the form of the average log-likelihood function for the unknown parameters of this problem and the form of the asymptotic distribution of the corresponding maximum likelihood estimator.

Answer: The usual procedure for constructing the likelihood for a limited dependent variable problem is to replace the density of the dependent variable with an integral when the dependent variable is censored, where the range of integration is the region in which the censored latent variable resides. Thus the average log-likelihood for the unknown parameters is

$$\mathcal{L}_n(\beta, \sigma^2) = \frac{1}{n}\sum_{i=1}^{n}\left[\begin{array}{c} 1\{y_i = 0\} \cdot \log\left\{\Phi\left(\frac{-x_i'\beta}{\sigma}\right)\right\} + 1\{0 < y_i < L_i\} \cdot \log\left\{\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right\} \\ +1\{y_i = L_i\} \cdot \log\left\{\Phi\left(\frac{U_i - x_i'\beta}{\sigma}\right) - \Phi\left(\frac{L_i - x_i'\beta}{\sigma}\right)\right\} + 1\{y_i > U_i\} \cdot \log\left\{\frac{1}{\sigma}\phi\left(\frac{y_i - (U_i - L_i) - x_i'\beta}{\sigma}\right)\right\} \end{array}\right].$$

A simple expression for the asymptotic distribution of the MLE $\hat{\theta} = (\hat{\beta}', \hat{\sigma}^2)'$ is the usual

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, \mathcal{I}(\theta_0)^{-1}\right),$$

where

$$\mathcal{I}(\theta_0) = -E\left[\frac{\partial^2 \mathcal{L}_n(\beta_0, \sigma_0^2)}{\partial\theta\partial\theta'}\right].$$

The expression for this is quite messy, due to the presence of σ^{-1} (and not σ^2) all over the place; any progress on these calculations is laudable (and will be rewarded!).■

B. Suppose that the parametric form of the error distribution is unknown. Find a $\sqrt{n}$-consistent estimator of β_0, imposing a suitable stochastic restriction on the conditional distribution of ε_i given x_i, and *without* imposing a scale normalization on β_0. If possible, give an expression for the asymptotic distribution of your estimator.

Answer: Since $y_i = \tau_i(y_i^*) = \tau_i(x_i'\beta_0 + \varepsilon_i)$ is nondecreasing in ε_i, the parameter of interest β_0 should be identifiable under a conditional median restriction on the errors – that is, if $med\{\varepsilon_i|x_i\} = 0$, then $med\{y_i|x_i\} = \tau_i(x_i'\beta_0)$, so (nonlinear) least absolute deviations estimation is a natural estimation approach, imposing the condition $E[sgn\{\varepsilon_i\}|x_i] = 0$. The LAD estimator $\hat{\beta}_{LAD}$ is defined as

$$\hat{\beta}_{LAD} = \arg\min_{\beta\in B} \frac{1}{n}\sum_{i=1}^{n} \left|y_i - \tau_i\left(x_i'\beta\right)\right|,$$

where B is the (compact) parameter space. The asymptotic distribution of nonlinear LAD estimation gives

$$\sqrt{n}\left(\hat{\beta}_{LAD} - \beta_0\right) \xrightarrow{d} \mathcal{N}\left(0, H^{-1}VH^{-1}\right),$$

where

$$\begin{aligned} H &\equiv 2E\left[f(0|x_i)z_i z_i'\right] \\ V &\equiv E\left[z_i z_i'\right], \end{aligned}$$

for

$$\begin{aligned} z_i &\equiv \frac{\partial \tau_i\left(x_i'\beta_0\right)}{\partial \beta} \\ &= \left[1\left\{0 < x_i'\beta_0 < L_i\right\} + 1\left\{U_i < x_i'\beta_0\right\}\right] x_i \end{aligned}$$

and $f(0|x_i)$ is the conditional density of ε_i given x_i at its median value, zero. Like the censored LAD estimator, no scale normalization is needed on the coefficient vector β_0 with this estimation approach. For local and global identification, the H matrix must indeed be nonsingular, etc.■

C. Now suppose that y_i^* is never observed, but only the range that it falls into is observed. More specifically, the dependent variable y_i is now defined as

$$\begin{aligned} y_i &\equiv t_i(y_i^*) \\ &= 0 \quad if \quad y_i^* \leq 0, \\ &= 1 \quad if \quad 0 < y_i^* \leq L_i, \\ &= 2 \quad if \quad L_i < y_i^* \leq U_i, \text{ and} \\ &= 3 \quad if \quad U_i < y_i^*. \end{aligned}$$

Describe an alternative consistent estimator of β_0 under a semiparametric restriction on the conditional distribution of the errors given the regressors. Is a scale normalization on β_0 needed, or are all the components of β_0 (including the scale) identifiable under your restriction?

Answer: Just as for part B., a conditional median restriction $E[sgn\{\varepsilon_i\}|x_i] = 0$ serves to identify the unknown coefficient vector β_0, since the latent variable transformation $y_i = t_i(y_i^*)$ is nondecreasing in y_i^*, and thus is nondecreasing in ε_i. The nonlinear LAD estimator for β_0 is

$$\hat{\beta}_{LAD} = \arg\min_{\beta\in B} \frac{1}{n}\sum_{i=1}^{n} \left|y_i - t_i\left(x_i'\beta\right)\right|$$

here; but since $\partial t_i(\mu)/\partial\mu = 0$ whenever it is well-defined, the usual asymptotic normality theory for NLLAD estimation is not applicable, and, like Manski's maximum score estimator, the estimator $\hat{\beta}$ will not be $\sqrt{n}$-consistent. Still, a scale normalization is not needed here, since $t_i\left(x_i'\beta\right) \neq t_i\left(\alpha(x_i'\beta)\right)$ unless $\alpha = 1$ or $x_i'\beta = 0$ (unlike the binary response model, where $t(\mu) = 1\left\{\mu \geq 0\right\} = 1\left\{\alpha\mu \geq 0\right\} = t\{\alpha\mu\}$ if $\alpha > 0$). An alternative estimation approach might be based upon a "single-index" restriction (or stronger restriction of independence of the errors and regressors), which would yield a $\sqrt{n}$-consistent estimator for β_0, but only up to scale. ■