Field Examination: Econometrics

Department of Economics
University of California, Berkeley

August 2021

Instructions: Answer THREE of the following four questions.

By submitting your answers to this exam, you acknowledge that, on your honor, you have neither given nor received any assistance in taking this exam.

1. Consider a partially linear IV model

$$Y^* = D\theta_0 + g_0(X) + U, \quad \mathrm{E}[U \mid X, Z] = 0 \tag{1}$$
$$Z = m_0(X) + V, \quad \mathrm{E}[V \mid X] = 0, \tag{2}$$

where Z is the instrumental variable, D is the endogenous treatment, and Y^* is the true outcome. The observed data vector is

$$W^* = (X, D, Z, \Delta, \Delta \cdot Y^*),$$

where

$$\Delta = 1[Y^* \text{ is observed }].$$

. We assume that Y^* is missing at random given X, and the probability of treatment assignment (i.e., the propensity score) is

$$\pi(X) = \Pr(\Delta = 1|X)$$

A. (Warm-up) Suppose Y^* is always observed (i.e., $\pi(X) = 1$ a.s.). Write down any valid, unconditional moment equation for θ_0.

B. Write down a moment equation for θ_0 when Δ is partially observed, that is $\pi(X) \leq 1$.

C. In case B., write down an orthogonal moment equation for θ_0. Demonstrate the orthogonality property.

D. Propose a Double ML approach to estimate θ_0. Clearly state your assumptions, describe the method, and provide a formal central limit theorem statement for it.

2. Suppose a sample of $n = \binom{N}{2}$. observations on a scalar dependent variable y_{ij} and p-dimensional vector of regressors x_{ij} are related by a linear equation

$$y_{ij} = x'_{ij}\beta_0 + \varepsilon_{ij},$$

for $i = 1, ...N-1$ and $j = i+1, ..., N$, where the slope coefficients $\beta_0 \in \mathbb{R}^p$ are unknown, and the unobservable error term ε_{ij} is continuously distributed given x_{ij} and satisfies a conditional median restriction

$$\Pr\{\varepsilon_{ij} \leq 0 | x_{ij}\} = \frac{1}{2}.$$

The error term ε_{ij} is assumed to have (unknown) conditional density function $f(\varepsilon|x_{ij})$ that is very well behaved (i.e., having lots of continuous derivatives, with the level and derivatives of f being uniformly bounded) and is bounded away from zero for all ε in a neighborhood of zero, and the regressors have all moments finite.

A. Assuming $z_{ij} \equiv (y_{ij}, x'_{ij})'$ is independently and identically distributed across distinct (i, j) pairs of observations, describe how β_0 might be consistently estimated by an extremum estimator of the form

$$\hat{\beta} \equiv \arg\min_{\beta \in \mathbb{R}^p} S_n(\beta),$$
$$S_n(\beta) \equiv \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho(z_{ij}, \beta),$$

for an appropriate choice of criterion function $\rho(\cdot)$, and give additional conditions that ensure consistency and asymptotic normality of $\hat{\beta}$. Derive the expression for the asymptotic distribution of $\hat{\beta}$ under your conditions.

B. Now suppose that z_{ij} is not i.i.d., but instead satisfies the "network structure"

$$x_{ij} = x(A_i, A_j) = x(A_j, A_i)$$

and

$$\varepsilon_{ij} = U_i + U_j + V_{ij} = u(A_i) + u(A_j) + V_{ij},$$

where A_i is an unobservable "agent specific" random variable that is i.i.d. over $i = 1, ..., N$ and is independent of the i.i.d. "pair specific" error term V_{ij} for all i and j. Furthermore, suppose that U_i and V_{ij} are symmetrically distributed around zero given x_i and x_j (which implies the conditional median restriction on ε_{ij} above), and V_{ij} is independent of A_i and A_j and is continuously distributed with (well-behaved) density function $f_V(v)$ that is bounded above.. (Thus

$$f(\varepsilon|x_{ij}) = E[f_V(\varepsilon - U_i - U_j|x_{ij})],$$

which is symmetric about zero given x_{ij}.)

For the same estimator of β_0 you defined above, what changes in the argument for consistency of $\hat{\beta}$? Decomposing

$$\rho(z_{ij}, \beta) = q(A_i, A_j, \beta) + [\rho(z_{ij}, \beta) - q(A_i, A_j, \beta)],$$

where

$$q(A_i, A_j, \beta) \equiv E[\rho(z_{ij}, \beta)|A_i, A_j]$$

show that the second term is uncorrelated across distinct (i, j) pairs and sketch a proof of consistency of $\hat{\beta}$.

C. Assume that an approximate first-order condition for the minimization problem defining $\hat{\beta}$ holds:

$$\begin{aligned} \hat{\Psi}_N(\hat{\beta}) &\equiv \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \psi(z_{ij}, \hat{\beta}) \\ &= o_p\left(\frac{1}{\sqrt{N}}\right), \end{aligned}$$

where, as usual,

$$\psi(z_{ij}, \beta) = \frac{\partial \rho(z_{ij}, \beta)}{\partial \beta}$$

for all values of β for which the derivative is well-defined (and is a convex combination of the left- and right-derivatives otherwise).

Also, using a similar decomposition to part B. above, namely,

$$\psi(z_{ij}, \beta) = p(A_i, A_j, \beta) + [\psi(z_{ij}, \beta) - p(A_i, A_j, \beta)],$$

with

$$p(A_i, A_j, \beta) \equiv E[\psi(z_{ij}, \beta) | A_i, A_j],$$

suppose the following "stochastic equicontinuity" condition holds:

$$\left\| \left[\psi(z_{ij}, \hat{\beta}) - p(A_i, A_j, \hat{\beta})\right] - [\psi(z_{ij}, \beta_0) - p(A_i, A_j, \beta_0)] \right\| = o_p\left(\frac{1}{\sqrt{N}}\right).$$

Under these conditions, derive an expression for the asymptotic distribution of $\hat{\beta}$.

3. Consider the classic "Local Average Treatment Effects" (LATE) set-up of Angrist, Imbens and Rubin (1996, JASA). Let $X \in \{0,1\}$ be a binary treatment or policy, $Z \in \{0,1\}$ a binary instrument, $Y \in \mathbb{Y}$ the outcome of interest and $W \in \mathbb{W}$ a vector of pre-treatment control variables. Let $X(z)$ denotes a unit's potential treatment status when assigned encouragement $Z = z$. We rule out the existence of *defiers* $(X(1) < X(0))$.

Consider a subpopulation of units homogenous in $W = w$. Within such a subpopulation the average effect of the policy $X = 1$ versus $X = 0$ on *compliers* $(X(1) > X(0))$, those units induced to either take-up the treatment or not according to whether $Z = 1$ or $Z = 0$, can be shown to equal

$$\beta(w) = \frac{\mathbb{E}[Y|W = w, Z = 1] - \mathbb{E}[Y|W = w, Z = 0]}{\mathbb{E}[X|W = w, Z = 1] - \mathbb{E}[X|W = w, Z = 0]}.$$

A. To recover the unconditional LATE we need to average $\beta(W)$ over the distribution of W within the subpopulation of compliers:

$$\beta^{\text{LATE}} = \mathbb{E}[\beta(W)|X(1) > X(0)]. \tag{3}$$

Use Bayes' Rule to show that

$$\beta^{\text{LATE}} = \frac{\mathbb{E}[q_1(W) - q_0(W)]}{\mathbb{E}[p_1(W) - p_0(W)]}$$

with $q_z(w) = \mathbb{E}[Y|W = w, Z = z]$ and $p_z(w) = \mathbb{E}[X|W = w, Z = z]$ for $z = 0, 1$.

B. Let $e(w) = \Pr(Z = 1|W = w)$ be the probability of encouragement given pre-treatment covariates or the propensity score. The efficient influence function for $\beta \overset{def}{\equiv} \beta^{\text{LATE}}$ can be show to equal

$$\begin{aligned}\psi(R, \beta, e, h) =& \frac{Z}{e(W)}(Y - \beta X) - \frac{1 - Z}{1 - e(W)}(Y - \beta X) \\ & - \left\{\frac{q_1(W) - \beta p_1(W)}{e(W)} + \frac{q_0(W) - \beta p_0(W)}{1 - e(W)}\right\}(Z - e(W))\end{aligned} \tag{4}$$

for $R = (W', X, Y, Z)'$ and $h = (q_0, q_1, p_0, p_1)$. Let $e_* \neq e_0$ be some function of W which is not equal to the propensity score. Show that

$$\mathbb{E}[\psi(R, \beta_0, e_*, h_0)] = 0. \tag{5}$$

Similarly show that, for $h_* \neq h_0$

$$\mathbb{E}[\psi(R, \beta_0, e_0, h_*)] = 0. \tag{6}$$

Comment on any implications of (5) and (6) for estimation.

C. Show that

$$\mathbb{E}\left[\frac{Z - e(W)}{e(W)[1 - e(W)]}(Y - \beta X)\right] = 0. \tag{7}$$

D. Let $e_0(W) = e(W; \eta_0)$ be a (correct) parametric model of the propensity score. Let $\hat{\eta}$ be the maximum likelihood estimate (MLE) of η_0. Consider the estimate which solves

$$\frac{1}{N}\sum_{i=1}^{N} \frac{Z_i - e(W_i; \hat{\eta})}{e(W_i; \hat{\eta})[1 - e(W_i; \hat{\eta})]}\left(Y_i - \hat{\beta} X_i\right) = 0.$$

Is this estimate efficient? Why or why not?

E. Show that the analog estimate based upon

$$\mathbb{E}\left[\begin{pmatrix} W \\ \frac{Z - e(W)}{e(W)[1 - e(W)]}\end{pmatrix}\left(Y - W'\gamma - \beta X\right)\right] = 0$$

also consistently estimates LATE. Further show that $\mathbb{E}^* \left[Y - X\beta \right| W] = W'\gamma$.

F. Provide a condition under which the analog estimate of β in part E. is locally semiparametric efficient. Discuss?

G. Can you provide a set of conditions under which the linear instrumental variables fit of Y onto X and W with Z being the excluded instrument for X provides a consistent estimate of β?

4. Suppose $\{y_t : 1 \le t \le T\}$ is an observed time series generated by the model

$$y_t = \delta t + u_t, \qquad u_t = \rho u_{t-1} + \varepsilon_t, \qquad t = 1, \ldots, T,$$

where $u_0 = 0$ and $\varepsilon_t \sim i.i.d.\ \mathcal{N}(0,1)$, while $\rho \in (-1,1)$ is a parameter of interest and $\delta \in \mathbb{R}$ is a (possibly) unknown nuisance parameter.

A. Find the log likelihood function $\mathcal{L}(\rho, \delta)$ and, for $d \in \mathbb{R}$, derive $\hat{\rho}(d) = \arg\max_\rho \mathcal{L}(\rho, d)$, the maximum likelihood estimator of ρ when δ is assumed to equal d.

B. Find the limiting distribution (after appropriate centering and rescaling) of the "oracle" estimator $\hat{\rho}(\delta)$.

C. Give conditions on $\hat{\delta}$ under which $\hat{\rho}(\hat{\delta})$ asymptotically equivalent to $\hat{\rho}(\delta)$.

D. Let

$$\hat{\delta}_{OLS} = \frac{\sum_{t=1}^T d_t y_t}{\sum_{t=1}^T d_t^2}.$$

Does $\hat{\delta}_{OLS}$ satisfy the condition derived in C.? If not, determine whether $\hat{\rho}(\hat{\delta}_{OLS})$ is asymptotically equivalent to $\hat{\rho}(\delta)$.