Econometrics Field Exam

Department of Economics, University of California - Berkeley

August 2020

Instructions:

- Answer all of the following questions.
- Question 1 receives **100%** weight.
- Questions 2-4 each receive **0%** weight; the main purpose of these questions is to help you practice and review some of the material covered in class.
- No books, notes, tables, or calculating devices are permitted.
- You have **180** minutes to answer all questions.
- Please make your answers elegant, that is, clear, concise, and, above all, correct.

[**Question 1; 100% weight**] What is your favorite soccer team? (If left blank, it will be assumed that your answer is "River Plate" and you will still get 100 % of the points).

[**Question 2; 0% weight**] Consider the following IV non-parametric regression model

$$Y = \theta_0(Z) + U, \qquad \text{where } E[U|X] = 0$$

for some $\theta_0 \in L^2 \equiv L^2(\mathbb{Z})$, where X is the "instrument" and Z is the "endogenous" variable.

Assume that the conditional expectation operator — which maps $h \in L^2(\mathbb{Z})$ into $E[h(Z)|X = .] = \int h(z)P(dz|X = .) \in L^2(\mathbb{X})$ —, and $x \mapsto r(x) \equiv E[Y|X = x]$ are *known* to you (the applied researcher). Finally, let $||\theta||_w^2 \equiv E\left[(E[\theta(Z)|X])^2\right]$ be the so-called "weak norm".

(a.i) Show that $\theta_0 \in \arg\min_{\theta \in L^2} ||\theta_0 - \theta||_w^2$, but it may not be unique.

(a.ii) What conditions over the conditional expectation operator are needed to ensure point identification?

(a.iii) What do (a.i) and (a.ii) tell you about the relationship between the norms $||.||_w$ and $||.||_{L^2}$?

For any $k \in \mathbb{N}$, let

$$\theta_k \equiv \arg\min_{\theta \in L^2} ||\theta_0 - \theta||_w^2 + \lambda_k ||\theta||_{L^2}^2, \qquad ||\theta||_{L^2}^2 = \int |\theta(z)|^2 P(dz),$$

where $\lambda_k > 0$.

(b) For any $k \in \mathbb{N}$, show that θ_k exists (in the sense that the "arg min" is non-empty) and that it is unique.

(c) For any $k \in \mathbb{N}$, show that $||\theta_k - \theta_0||_w^2 + \lambda_k ||\theta_k||_{L^2}^2 = O(\lambda_k)$.

For any $k \in \mathbb{N}$, consider the estimator

$$\hat{\theta}_k \equiv \arg\min_{\theta \in L^2} n^{-1} \sum_{i=1}^{n} (r(X_i) - E[\theta(Z)|X_i])^2 + \lambda_k ||\theta||_{L^2}^2.$$

(d) Show that

$$||\hat{\theta}_k - \theta_k||_w^2 + \lambda_k ||\hat{\theta}_k - \theta_k||_{L^2}^2 = O_P(\delta_n),$$

where $(\delta_n)_n$ is such that

$$\sup_{\theta \in L^2} |n^{-1} \sum_{i=1}^{n} (r(X_i) - E[\theta(Z)|X_i])^2 - E[(r(X) - E[\theta(Z)|X])^2]| = O_P(\delta_n).$$

Hint: At one point, you may want to use the fact θ_k satisfies a first order condition and that $||\theta_0 - .||_w^2 + \lambda_k || \cdot ||_{L^2}^2$ is quadratic.

(e.i) What type of restrictions on $(\lambda_k)_k$ and $(\delta_n)_n$ are needed to ensure consistency of $\hat{\theta}_k$ to θ_0 *under* $||.||_w$? What is the rate of convergence?

(e.ii) Answer (e.i), but using $||.||_{L^2}$ instead of $||.||_w$.

(e.iii) Discuss the differences between the convergence rates in (e.i) and (e.ii). Why do you think they are different?

[Question 3; 0% weight] Suppose $\{y_t : 1 \leq t \leq T\}$ is an observed time series generated by the model

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

where $u_0 = 0$ and $\varepsilon_t \sim i.i.d.\ \mathcal{N}(0,1)$, while $\mu \in \mathbb{R}$ and $\rho \in (-1,1)$ are (possibly) unknown parameters.

(a) Find the log likelihood function $\mathcal{L}(\mu, \rho)$ and, for $r \in (-1,1)$, derive $\hat{\mu}(r) = \arg\max_{\mu} \mathcal{L}(\mu, r)$, the maximum likelihood estimator of μ when ρ is assumed to equal r.

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

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

(d) Does $\hat{\rho} = 0$ satisfy the condition derived in (c)? If not, determine whether $\hat{\mu}(0)$ is asymptotically equivalent to $\hat{\mu}(\rho)$.

[Question 4; 0% weight] Let $\{(W_i, X_i, Y_i)\}_{i=1}^{N}$ be a simple random sample drawn from a population characterized by (unknown) distribution F_0; Y is a scalar outcome variable of interest, X a vector of policy or treatment variables, and W a vector of additional control variables or confounders. Assume that the conditional distribution of Y given W and X satisfies:

$$Y = X'\beta_0 + h_0(W) + U, \qquad \mathbb{E}[U|W, X] = 0. \tag{1}$$

Here $\beta_0 \subset \mathbb{R}^K$ is the finite dimensional parameter of interest and $h_0(W)$ an unknown function mapping from a subset of $W \in \mathbb{W} \subset \mathbb{R}^{\dim(W)}$ into $\mathcal{H} \subset \mathbb{R}$. You may assume that all the expressions which appear below are well-defined (i.e., don't worry about regularity conditions in what follows).

(a) Assume that the researcher knows that $h_0(W) = k(W)'\delta_0$ for some known $J \times 1$ vector of basis functions $k(W)$ (which includes a constant) and unknown parameter δ_0. Show that the coefficient on X in the least squares fit of Y onto X and $k(W)$ has an asymptotic limiting distribution of

$$\sqrt{N}\left(\hat{\beta}_{\text{OLS}} - \beta_0\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2 \mathbb{E}\left[(X - \Pi_0 k(W))(X - \Pi_0 k(W))'\right]^{-1}\right), \tag{2}$$

with $\Pi_0 = \mathbb{E}\left[Xk(W)'\right]\mathbb{E}\left[k(W)k(W)'\right]^{-1}$ the $K \times J$ matrix of projection coefficients associated with the multivariate regression of X onto $k(W)$. Note this result maintains the additional assumption of homoscedasticity. For your reference the partitioned matrix formula is:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)CA^{-1} & -A^{-1}B(D - CA^{-1}B) \\ -(D - CA^{-1}B)CA^{-1} & (D - CA^{-1}B) \end{bmatrix}.$$

(b) Let

$$e_0(w) = \mathbb{E}[X|W = w]. \tag{3}$$

be a vector of conditional expectation functions (CEFs) of the K policy variables given the confounders. Assume that $e_0(W)$ is known. Show that following unconditional moment identifies β_0:

$$\mathbb{E}[m(Z, \beta_0, e_0(W))] = 0. \tag{4}$$

with $m(Z, \beta_0, e_0(W)) = (Y - X'\beta_0)(X - e_0(W))$. Let $\hat{\beta}_E$ be the method-of-moments estimator associated with this restriction. Is this estimator more or less efficient than the OLS estimator described in part (a)?

(c) Augment the identifying "E-moment" (4) from part (b) with the auxiliary conditional moment

$$\mathbb{E}[(X - e_0(W))|W] = 0. \tag{5}$$

Argue that the efficient estimator based on (4) and (5) coincides with the just-identified moment based upon the unconditional moment restriction

$$\psi(Z, \beta_0, e_0(W), h_0(W)) = m(Z, \beta_0, e_0(W)) - \mathbb{E}^*[m(Z, \beta_0, e_0(W))|(X - e_0(W)); W].$$

Here $\mathbb{E}^* [Y|X;W]$ denotes the linear regression of Y onto X within a subpopulation homogenous in W (i.e., a conditional linear predictor). You may assume that all "unknowns" except β_0 are known. You may not assume that "unknown unknowns are known" only that "known unknowns are known".

Further show that

$$\psi (Z, \beta_0, e_0 (W), h_0 (W)) = (Y - X'\beta_0 - h_0 (W)) (X - e_0 (W)). \tag{6}$$

(d) Replace $h_0 (w)$ in (6) with some arbitrary function of w. Is (6) still mean zero? Now replace $e_0 (w)$ in (6) with some arbitrary function of w. Is (6) still mean zero? Comment on the implications of your answer for estimation.

(e) Write, in the language of your choice, a short Haiku inspired by questions (a) to (d) above.