Econometrics Field Exam

Department of Economics, UC Berkeley

August 2023

Instructions:

- Answer all of the following questions.
- 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] Let $(Y_0, Y_1) \in \mathbb{Y} \times \mathbb{Y} = [0, \infty) \times [0, \infty)$ be a random draw from a population of "paired" durations; Y_0 corresponds to the duration outcome of a control unit, while Y_1 to that of a treated unit. The conditional hazard functions for Y_0 and Y_1 are given by

$$\begin{aligned} \lambda\left(y_0 \mid a; \theta\right) =& \lambda_0\left(y_0; \alpha\right) e^{A}. \\ \lambda\left(y_1 \mid a; \theta\right) =& \lambda_0\left(y_1; \alpha\right) e^{\beta + A}, \end{aligned} \tag{1}$$

with A unobserved, pair-specific-heterogeneity and $\lambda_0\left(y_t; \alpha\right)$, for $t = 0, 1$, a known parametric baseline hazard function indexed by the unknown parameter α. Your mission, should you choose to accept it, is to compute a consistent estimate β. You may assume that $\theta = (\alpha, \beta)'$ equals its population value (i.e., $\theta = \theta_0$) in what follows unless noted otherwise. Let $Z = (Y_0, Y_1)'$. We assume random sampling of pairs such that $(Z_1, A_1), \ldots, (Z_N, A_N)$ are iid.

Let $\Lambda_0\left(y; \alpha\right) = \int_0^y \lambda_0\left(t; \alpha\right) \mathrm{d}t$ denote the integrated baseline hazard. Recall that the conditional survival functions for the two durations can be written as

$$\begin{aligned} \Pr\left(Y_0 > y_0 \mid A = a; \theta\right) &= S\left(y_0 \mid A = a; \theta\right) = \exp\left(-\Lambda_0\left(y_0; \alpha\right) e^{a}\right) \\ \Pr\left(Y_1 > y_1 \mid A = a; \theta\right) &= S\left(y_1 \mid A = a; \theta\right) = \exp\left(-\Lambda_0\left(y_1; \alpha\right) e^{\beta + a}\right), \end{aligned}$$

while the two duration densities can be written in terms of the hazard and survival functions as

$$\begin{aligned} f\left(y_0 \mid A = a; \theta\right) &= \lambda\left(y_0 \mid a; \theta\right) S\left(y_0 \mid A = a; \theta\right) \\ f\left(y_1 \mid A = a; \theta\right) &= \lambda\left(y_1 \mid a; \theta\right) S\left(y_1 \mid A = a; \theta\right). \end{aligned}$$

(a) Consider the special case of a constant baseline hazard: $\lambda_0\left(y; \alpha\right) = \alpha$. Normalize $\alpha = 1$ (so that the mean of e^{A} is unrestricted) and show that the mean conditional duration for control units is

$$\mathbb{E}\left[Y_0 \mid A = a\right] = e^{-a}$$

while that for treated units is

$$\mathbb{E}\left[Y_1 \mid A = a\right] = e^{-\beta - a}.$$

You may find it helpful to recall that the density of an exponential random variable, Z, with a rate parameter of λ at $Z = z$ is $\lambda \exp(-z\lambda)$ with mean $1/\lambda$ and variance $1/\lambda^2$ (this hint is also useful below). From the above we get a conditional average treatment effect (CATE) equal to

$$\gamma(a) = \mathbb{E}\left[Y_1 - Y_0 \mid A = a\right] = e^{-a}\left(e^{-\beta} - 1\right) \tag{2}$$

and hence an average treatment effect (ATE) of

$$\gamma = \mathbb{E}\left[Y_1 - Y_0\right] = \mathbb{E}\left[e^{-A}\right]\left(e^{-\beta} - 1\right).$$

Discuss any restrictions on treatment response imposed by (2). Why is the average treatment effect declining in β?

(b) Show, conditional on $A = a$, that Y_0 and $Y_1 e^{\beta}$ are independent exponential random variables with identical rate parameters of e^{a}.

(c) Using the result in (b) show that

$$\mathbb{E}\left[Y_1 e^{\beta} - Y_0 \mid A = a\right] = 0 \tag{3}$$

for all $a \in \mathbb{A}$ and hence that $\mathbb{E}\left[Y_1 e^{\beta} - Y_0\right] = 0$ for *any* heterogeneity distribution.

(d) An omniscient econometrician is able to observe A for all units. Show that the optimal estimating equation based solely on restriction (3) is (use what you know about efficiency bounds for conditional moment problems)

$$\psi(Z, A; \beta) = \frac{1}{2} e^{A}\left(Y_1 e^{\beta} - Y_0\right). \tag{4}$$

(e) You are not omniscient, but you *a priori* believe that $e^{A} \sim \text{Gamma}(\eta, \lambda)$. Because you are student of conjugate priors you also know that, according to your beliefs, *a posteriori*,

$$\mathbb{E}\left[e^{A} \mid Y_0, Y_1\right] = \frac{\eta + 2}{Y_0 + Y_1 e^{\beta} + \lambda}.$$

Since you don't know A you replace it in (4) with your best guess given what you believe *a posteriori* to be true about the world. This yields a moment function of

$$\psi(Z; \beta, \eta, \lambda) = \frac{\eta + 2}{2}\left(\frac{Y_1 e^{\beta} - Y_0}{Y_0 + Y_1 e^{\beta} + \lambda}\right). \tag{5}$$

Show that this moment function is mean zero at $\beta = \beta_0$ irrespective whether "your subjective truth" is the "true objective truth" or "objectively false subjective truth". That is show that

$$\mathbb{E}\left[\psi(Z; \beta_0, \eta, \lambda)\right] = 0$$

for any η, λ and hence heterogeneity distribution for A.

(f) Fix η, λ at some arbitrary values. Show that the method of moments estimate of β based on (5) has a limit distribution of

$$\sqrt{N}\left(\hat{\beta}_{\lambda}-\beta\right) \rightarrow \mathcal{N}\left(0, \Lambda\left(\beta, \lambda\right)^{-1}\right) \tag{6}$$

where

$$\Lambda\left(\beta, \lambda\right)=\frac{\mathbb{E}\left[\frac{Y_1 e^{\beta}}{Y_0+Y_1 e^{\beta}+\lambda}\left(1-\frac{Y_1 e^{\beta}-Y_0}{Y_0+Y_1 e^{\beta}+\lambda}\right)\right]^2}{\mathbb{E}\left[\left(\frac{Y_1 e^{\beta}-Y_0}{Y_0+Y_1 e^{\beta}+\lambda}\right)^2\right]}. \tag{7}$$

Note that neither $\hat{\beta}_{\lambda}$, nor its limit distribution, depends on η. Suggest an approach for choosing λ in practice? Is there a best choice? Worst choice?

(g) Your friends in the CLE courteously suggest to you that your focus on estimating β is comically misguided. They note that the ATE may be directly estimated by

$$\hat{\gamma}=\frac{1}{N}\sum_{i=1}^{N}\left(Y_{i1}-Y_{i0}\right).$$

You instead suggest an estimate of

$$\hat{\gamma}_{\lambda}=\frac{1}{2N}\sum_{i=1}^{N}\left(Y_{0i}+Y_{1i}e^{\hat{\beta}_{\lambda}}\right)\left(e^{-\hat{\beta}_{\lambda}}-1\right).$$

Discuss how you would explain your alternative estimation approach to your colleague. Comment on possible advantages and disadvantages of your approach.

[Question 2] 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 $\rho \in (-1, 1)$ is a parameter of interest and $\mu \in \mathbb{R}$ is a (possibly) unknown nuisance parameter.

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

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

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

Let

$$\hat{\mu}_{OLS} = T^{-1} \sum_{t=1}^{T} y_t.$$

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

[Question 3] Consider a policy learning problem with data vector $W = (X, D, Y)$. Here, $D \in \{1, 0\}$ is a binary treatment, X is a vector of baseline (pre-treatment) covariates, $Y(1)$ is an outcome when treated and $Y(0)$ is an outcome when non-treated, respectively, and

$$Y = DY(1) + (1 - D)Y(0)$$

is realized outcome. The policy classifier $G : \mathcal{X} \to \{1, 0\}$ maps a vector of observables X to a policy prescription $\{1, 0\}$ (treat, not treat). The average welfare of the classifier G is given by

$$W(G) = \mathrm{E}Y(1)1\{X \in G\} + \mathrm{E}Y(0)1\{X \in G^c\},$$

where G^c is a complement of G, that is $G \sqcup G^c = \mathcal{X}$.

(a.i) Derive the optimal classifier

$$G^* = \arg \min_{\text{all classifiers}} W(G)$$

(a.ii) Calculate the optimal value of welfare $W(G^*)$.

(b) **Prove** that the optimal classifier G^* is indeed optimal. That is, show that any other policy classifier $G : \mathcal{X} \to \{1, 0\}$ attains a non-negative regret

$$R(G) = W(G) - W(G^*) \geq 0.$$

(c) Let $(X_i, D_i, Y_i)_{i=1}^n$ be an i.i.d sample. Sketch an Empirical Welfare Maximization (EWM) policy classifier.

(d) Let $(X_i, D_i, Y_i)_{i=1}^n$ be an i.i.d sample. Consider the regression functions

$$\mathrm{E}[Y = 1 \mid D = d, X = x] = x'\theta_d, \quad d \in \{1, 0\}.$$

where x is a high-dimensional sparse vector:

$$\dim(x) = p \gg n, \max_{d \in \{1,0\}} \|\theta_d\|_0 = s \ll n.$$

Sketch a possible **plug-in** classifier, appropriate for this setup.

(e) Given an i.i.d sample of $(X_i, D_i, Y_i)_{i=1}^n$, sketch a possible DML estimator of the optimal average welfare $W(G^*)$. Specifically,

- Write a definition of $W(G^*)$ and identify the nuisance functions that it depends on.
- Write down an *orthogonal* moment equation for $W(G^*)$. Explain why this equation is likely to be less sensitive to the biased estimation of its functions than the one defined above.
- Sketch a Double Machine Learning estimator of $W(G^*)$.