Econometrics Field Exam

Department of Economics, University of California - Berkeley

August 2022

Answer three of the four questions. Please start each question on a new page and label all work clearly. Good luck!

[Q1] Let Y_{it} equal log-earnings at age 40 of generation t for family i. You are interested in the following model

$$Y_{it} = \rho_0 Y_{it-1} + A_i + U_{it}, \ \ \mathbb{E}\left[U_{it} \middle| \mathcal{I}_{it}\right] = 0 \tag{1}$$

with $\mathcal{I}_{it}$ denoting the childhood information set of generation $t = 1, \ldots, T$. This information set includes A_i and $Y_{i0}^{t-1} = (Y_{i0}, \ldots, Y_{it-1})'$ as well as knowledge of the magnitude of ρ_0. Available to the econometrician is a dataset consisting of three (consecutive) generations of earnings data for a random sample of United States residents currently aged forty. To be concrete it may be helpful to think of the $t = 2$ generation as born ˜1980, the $t = 1$ generation as born ˜1955, and the $t = 0$ "initial" generation as born ˜1930.

[a] Consider the linear regression of Y_{it} onto a constant and Y_{it-1}

$$\mathbb{E}^*\left[Y_{it} \middle| Y_{it-1}\right] = a + bY_{it-1}.$$

Recall that $a = \mathbb{E}\left[Y_{it}\right] - b\mathbb{E}\left[Y_{it-1}\right]$ and hence that

$$\mathbb{E}^*\left[Y_{it} \middle| Y_{it-1}\right] = \mathbb{E}\left[Y_{it}\right] + b\left(Y_{it-1} - \mathbb{E}\left[Y_{it-1}\right]\right).$$

The slope coefficient b is a measure of income persistence across generations. Assume that $0 \leq b \leq 1$. Which society exhibits more intergenerational mobility, one where $b = 0$ or one where $b = 1$? Explain. **[2 - 4 sentences].**

[b] Assume that $\rho = 0$. Is b also zero? Explain? **[2 - 4 sentences].** Here A_i is a family-specific latent variable. List some reasons why A_i might differ across families? **[2 - 3 sentences].**

[c] You are advising the President of the United States on her "equality of opportunity initiative". She says "we've got an opportunity crisis on our hands, $b = 0.80$. If we introduce a negative income tax for individuals below the average income, $Y_{it} < \mathbb{E}\left[Y_{it}\right]$, and a positive income tax for those above, $Y_{it} > \mathbb{E}\left[Y_{it}\right]$, we can improve equality today and for the next generation as well." Assume that the actual intergenerational income process is given by (1). Discuss the president's policy proposal. **[3 - 5 sentences].** Assume that $\rho = 0$. Is her proposal likely to decrease inequality today? **[1 -2 sentences].** Is it likely to decrease inequality in the next generation? **[1 -2 sentences].**

[d] Show that (1) implies that

$$\mathbb{E}\left[U_{i2}U_{i1}|\mathcal{I}_{i1}\right]=0$$

and hence, by iterated expectations, that $\mathbb{E}\left[U_{i2}U_{i1}\right]=0$.

[e] Show that (1) also implies that $\mathbb{E}\left[U_{i2}A_i\right]=\mathbb{E}\left[U_{i1}A_i\right]=0$.

[f] Let $\triangle Y_{i2}=Y_{i2}-Y_{i1}$ and $\triangle Y_{i1}=Y_{i1}-Y_{i0}$ and consider the moment function

$$\psi_{1i}\left(\rho\right)=\left(\triangle Y_{i2}-\rho\triangle Y_{i1}\right)Y_{i0}.$$

Show that $\mathbb{E}\left[\psi_{1i}\left(\rho_0\right)\right]=0$ with ρ_0 denoting the true, or population, value of ρ. Describe how you might use this moment condition to construct a consistent estimate of ρ_0?

[g] Solve (1) recursively to get

$$Y_{it}=\left(\sum_{s=0}^{t-1}\rho_0^s\right)A_i+\rho_0^t y_{i0}+\left(\sum_{s=0}^{t-1}\rho_0^s U_{it-s}\right).$$

Next show that when $|\rho_0|<1$ that

$$\mathbb{E}\left[Y_{it}|\,A_i\right]\rightarrow\frac{A_i}{1-\rho_0}$$

as $t\rightarrow\infty$. Call this limit the *steady state mean* for family/unit i.

[h] Assume that

$$Y_{i0}=\frac{A_i}{1-\rho_0}+V_{i0} \tag{2}$$

with V_{i0} orthogonal to all U_{it} for $t=1,\ldots,T$ and also orthogonal to A_i. Consider the additional moment function

$$\psi_{2i}\left(\rho\right)=\left(Y_{i2}-\rho Y_{i1}\right)\triangle Y_{i1}.$$

Show that $\mathbb{E}\left[\psi_{2i}\left(\rho_0\right)\right]=0$ with ρ_0 denoting the true, or population, value of ρ.

[i] How would you estimate ρ_0 maintaining both (1) and (2)? How could you test restriction (2), while maintaining (1)?

[j] How do you expect your methods to perform when ρ_0 is close to one? **[3 - 5 sentences].**

[Q2] All part of this question carry equal weight. Consider the following IV non-parametric regression model

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

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".

1. (i) Show that $\theta_0 \in \arg\min_{\theta \in L^2} ||\theta_0 - \theta||_w^2$, but it may not be unique. (ii) What conditions over the conditional expectation operator are needed to ensure point identification? (iii) What does (i) and (ii) tell you about the relationship between the norms $||.||_w$ and $||.||_{L^2}$?

Let, for any $k \in \mathbb{N}$,

$$\theta_k \equiv \arg\min_{\theta \in L^2} ||\theta_0 - \theta||_w^2 + \lambda_k ||\theta||_{L^2}^2$$

where $\lambda_k > 0$. [1] For any $k \in \mathbb{N}$:

2. Show that θ_k exists (in the sense that the "arg min" is non-empty) and that is unique.

3. Show that $||\theta_k - \theta_0||_w^2 + \lambda_k ||\theta_k||_{L^2}^2 = O(\lambda_k)$.

Consider the estimator for any $k \in \mathbb{N}$,

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

4. 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.

5. 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?

[Q3] Double ML for Sample Selection. Suppose we are interested in the Average Treatment Effect parameter

$$\theta = \mathbb{E}[Y(1) - Y(0)],$$

where $Y(1)$ and $Y(0)$ are potential outcomes with and without treatment, respectively. The treatment D is randomly assigned given X:

$$(Y(1), Y(0)) \perp D \mid X,$$

and the propensity score is

$$\pi(X) = \Pr(D = 1 \mid X).$$

[1] $||\theta||_{L^2}^2 = \int |\theta(z)|^2 P(dz)$.

Furthermore, the outcome $Y = (1 - D) Y(0) + DY(1)$ is partially missing, and the selection indicator $S = 1$ if and only if Y is observed. To sum up, the data vector

$$W = (D, X, S, S \cdot Y),$$

consists of the pre-treatment covariates X, the treatment indicator $D \in \{1, 0\}$, the selection indicator S, and the actual outcome Y if and only if $S = 1$. The selection is exogenous, that is,

$$S \perp Y \mid D = d, X \quad \forall d \in \{1, 0\}.$$

1. Write down any valid, unconditional moment equation for θ_0. **Note!** A moment function for θ must depend only on data vector W, the parameter θ, and, possibly, other identified functions (e.g., the propensity score $\pi(X)$).

2. Write down an orthogonal moment equation for θ_0. Demonstrate the orthogonality property.

3. 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.

[Q4] Suppose $\{y_t : 1 \leq t \leq T\}$ 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 $\delta \in \mathbb{R}$ and $\rho \in (-1, 1)$ are (possibly) unknown parameters.

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

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

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

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