Final Exam
Empirical Analysis 1
Date: Thursday, December 9, 2021

1. The exam is closed book and closed notes with the exception of **one (two-sided) sheet of paper**.
2. No calculators are allowed.
3. There are a total of 100 possible points.
4. Answer as many questions as you can. You do not need to answer the questions in order. Try to answer the later parts of a question even if you have difficulty with earlier parts.
5. Please **clearly** write your answers in a blue book with your name written on it.
6. Please clearly label your final answers where appropriate.
7. Any students caught cheating will fail the course. The Dean of Students will be notified as well.
8. Good luck!

1. (8 points) For $0 \leq a < b < \infty$, let $X_n, n \geq 1$ be a sequence of random variables on $\mathbf{R}$ such that $P\{a \leq X_n \leq b\} = 1$ for $n \geq 1$ and let X be another random variable such that $P\{a \leq X \leq b\} = 1$. Show that $X_n \xrightarrow{d} X$ if and only if for all $k \geq 1$, $E[X_n^k] \to E[X^k]$. (Hint: Let $f : [a, b] \to \mathbf{R}$ be continuous and bounded. The Weierstrass approximation theorem states that for any $\delta > 0$ there is a (finite-order) polynomial $p : [a, b] \to \mathbf{R}$ such that $\sup_{a \leq x \leq b} |f(x) - p(x)| < \delta$.)

2. (14 points) Let $(Y_i(1), Y_i(0), D_i(1), D_i(0), X_i, Z_i), i = 1, \ldots, n$ be an i.i.d. sequence of random variables such that $D_i(1), D_i(0), X_i$, and Z_i are binary (i.e., take on only values 0 or 1). Suppose

   (i) $(Y_i(1), Y_i(0), D_i(1), D_i(0), X_i) \perp\!\!\!\perp Z_i$

   (ii) $P\{D_i(1) \neq D_i(0) | X_i = 1\} > 0$ and $P\{D_i(1) \neq D_i(0) | X_i = 0\} > 0$

   (iii) $P\{D_i(1) \geq D_i(0) | X_i = 1\} = 1$ and $P\{D_i(1) \geq D_i(0) | X_i = 0\} = 1$

   (a) (5 points) For $x \in \{0, 1\}$, provide a consistent estimator $\hat{\beta}_{n,x}$ of $\beta_x = E[Y_i(1) - Y_i(0) | D_i(1) > D_i(0), X_i = x]$. Justify your answer.

   (b) (9 points) Provide a consistent estimator $\hat{p}_n$ of $p = P\{X_i = 1 | D_i(1) > D_i(0)\}$. Justify your answer.

3. (48 points) Let $(Y_i(1), Y_i(0), X_i, D_i), i = 1, \ldots, n$ be i.i.d. where $Y_i(1) \in \mathbf{R}$ and $Y_i(0) \in \mathbf{R}$ are potential outcomes under treatment and control, respectively, $X_i \in \mathbf{R}^k$ is a vector of observed, baseline covariates, and D_i is an indicator for receipt of treatment. As usual, define the observed outcome to be

$$Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i) .$$

   Assume that

$$(Y_i(1), Y_i(0), X_i) \perp\!\!\!\perp D_i .$$

   The parameter of interest is the average treatment effect,

$$\tau = E[Y_i(1) - Y_i(0)] .$$

   (a) (8 points) A natural estimator of τ in this setting is

$$\hat{\tau}_n^{\text{diff}} = \frac{1}{n_1} \sum_{1 \leq i \leq n : D_i = 1} Y_i - \frac{1}{n_0} \sum_{1 \leq i \leq n : D_i = 0} Y_i ,$$

   where, for $d = 0, 1$, $n_d = |\{1 \leq i \leq n : D_i = d\}|$. Show that

$$\sqrt{n}(\hat{\tau}_n^{\text{diff}} - \tau) = \begin{pmatrix} \frac{n}{n_1} & -\frac{n}{n_0} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} (Y_i(1) - E[Y_i(1)]) D_i \\ \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} (Y_i(0) - E[Y_i(0)])(1 - D_i) \end{pmatrix}$$

   (b) (8 points) Use the result in the preceding question to show that

$$\sqrt{n}(\hat{\tau}_n^{\text{diff}} - \tau) \xrightarrow{d} N(0, \sigma^2_{\text{diff}})$$

   with

$$\sigma^2_{\text{diff}} = \frac{\text{Var}[Y_i(1)]}{P\{D_i = 1\}} + \frac{\text{Var}[Y_i(0)]}{P\{D_i = 0\}} .$$

   Clearly state any additional assumptions needed to justify your answer.

   (c) (4 points) Empirical researchers often try to exploit X_i by defining an estimator $\hat{\tau}_n^{\text{reg}}$ as the ordinary least squares estimate of the coefficient on D_i in a regression of Y_i on a constant, D_i and X_i. While $\hat{\tau}_n^{\text{reg}}$ and $\hat{\tau}_n^{\text{diff}}$ are both consistent for τ, the former estimator need not be more precise than $\hat{\tau}_n^{\text{diff}}$. Explain briefly why $\hat{\tau}_n^{\text{reg}}$ is consistent for τ.

(d) For this reason, it is useful to consider the following estimator:

$$\hat{\tau}_n^{\text{adj}} = \frac{1}{n_1} \sum_{1 \le i \le n : D_i = 1} \left(Y_i - (X_i - \bar{X}_n)'\hat{\gamma}_{1,n} \right) - \frac{1}{n_0} \sum_{1 \le i \le n : D_i = 0} \left(Y_i - (X_i - \bar{X}_n)'\hat{\gamma}_{0,n} \right) ,$$

where $\bar{X}_n = \frac{1}{n} \sum_{1 \le i \le n} X_i$ and, for $d = 0, 1$, $\hat{\gamma}_{n,d}$ is obtained as the ordinary least squares estimate of the coefficient on X_i in a regression of Y_i on a constant and X_i using *only* observations with $D_i = d$. This estimator is provably more precise that $\hat{\tau}_n^{\text{diff}}$. To see this, complete the following exercises:

i. (10 points) Show that

$$\begin{aligned} \hat{\tau}_n - \tau = & \left(\frac{1}{n_1} \sum_{1 \le i \le n : D_i = 1} (Y_i(1) - E[Y_i(1)]) - (X_i - E[X_i])'\gamma_1 \right) \\ & + \left(\frac{1}{n_0} \sum_{1 \le i \le n : D_i = 0} (Y_i(0) - E[Y_i(0)]) - (X_i - E[X_i])'\gamma_0 \right) \\ & + (\bar{X}_n - E[X_i])'(\gamma_1 - \gamma_0) + o_P(n^{-1/2}) . \end{aligned}$$

ii. (10 points) Use the result in the preceding question to show that

$$\sqrt{n}(\hat{\tau}_n^{\text{adj}} - \tau) \xrightarrow{d} N(0, \sigma^2_{\text{adj}})$$

with

$$\sigma^2_{\text{adj}} = \frac{\text{Var}[Y_i(1) - X_i'\gamma_1]}{P\{D_i = 1\}} + \frac{\text{Var}[Y_i(0) - X_i'\gamma_0]}{P\{D_i = 0\}} + (\gamma_1 - \gamma_0)'\text{Var}[X_i](\gamma_1 - \gamma_0) ,$$

where, for $d = 0, 1$, $\gamma_d = \text{Var}[X_i]^{-1}\text{Cov}[Y_i(d), X_i]$. Clearly state any additional assumptions needed to justify your answer.

iii. (8 points) Show that

$$\sigma^2_{\text{diff}} - \sigma^2_{\text{adj}} = \Delta'\text{Var}[X_i]\Delta \ge 0 ,$$

where

$$\Delta = \sqrt{\frac{P\{D_i = 0\}}{P\{D_i = 1\}}}\gamma_1 + \sqrt{\frac{P\{D_i = 1\}}{P\{D_i = 0\}}}\gamma_0 .$$

(Hint: You may wish to start by expanding $\text{Var}[Y_i(d) - X_i'\gamma_d]$.)

4. (30 points) Let $(X_i, U_i), i = 1, \ldots, n$ be i.i.d. such that $U_i | X_i \sim N(0, 1)$. Suppose $Y_i = X_i'\beta + V_i$, where, for a *known* γ, $V_i = \exp(X_i'\gamma)U_i$ and $E[X_i V_i] = 0$. Let $\hat{\beta}_n$ be the MLE of β.

(a) (5 points) Is the OLS estimator of β necessarily the best linear unbiased estimator of β? Explain briefly.

(b) (5 points) Write the (conditional) log-likelihood function of $Y_1, \ldots, Y_n$ given $X_1, \ldots, X_n$.

(c) (7 points) Derive an expression for $\hat{\beta}_n$.

(d) (7 points) Use the Fisher Information to derive the limit in distribution of $\hat{\beta}_n$ after appropriate centering and normalization.

(e) (6 points) Describe the Wald test for the null hypothesis $\beta = 0$ versus the alternative hypothesis that $\beta \neq 0$.