Problem Set 1
Due Sunday, February 5 at 11:59 PM (a minute before midnight) via Canvas
ECON 31720, University of Chicago, Winter 2023

Assignments must be typeset in nicely formatted LaTeX. Programming assignments must use low-level commands, be commented clearly (not excessively), formatted nicely in 80 characters per column, etc. **You will be graded on the exposition of your written answers, the clarity of your code, and the interpretability and beauty of your tables and graphics.** The problem sets are individual assignments, but you may discuss them with your classmates. Submit the problem sets through Canvas in a single zip/tar/rar file. **Late problem sets will not be accepted under any circumstances.**

1. Suppose that $D \in \{0,1\}$ is a binary treatment variable and $Y(0)$, $Y(1)$ are potential outcomes for an outcome variable, Y. Assume that selection on observables holds, so that $Y(0), Y(1)$ is independent of D, conditional on X. Let $p(x) \equiv \mathbb{P}[D = 1|X = x]$ and $P \equiv p(X)$. Let
$$W \equiv D + (1 - D)\left(\frac{P}{1 - P}\right).$$
Consider the weighted linear regression of Y on D and a constant, with weights W. That is,
$$(\beta_0, \beta_1) = \underset{b_0, b_1}{\arg\min}\, \mathbb{E}\left[W(Y - b_0 - b_1 D)^2\right].$$
True, False, or Uncertain: β_1 is equal to the average treatment on the treated (ATT).

2. Let D be a binary treatment and let $Y(0)$ and $Y(1)$ be associated potential outcomes. Let X be another observables. Suppose that D is not independent of $(Y(0), Y(1))$ either unconditionally, or conditional on X. Let $\mu^\star = \mathbb{E}[Y(1) - Y(0)]$ be the average treatment effect. Consider the quantities:
$$\mu_1 \equiv \mathbb{E}[Y|D = 1] - \mathbb{E}[Y|D = 0] \qquad \text{(uncontrolled contrast)}$$
$$\mu_2 \equiv \mathbb{E}\left[\mathbb{E}[Y|D = 1, X] - \mathbb{E}[Y|D = 0, X]\right] \qquad \text{(controlled imputation)}$$
Construct an economic story with a numerical example that produces $|\mu_2 - \mu^\star| > |\mu_1 - \mu^\star|$ Explain the relevance to selection on observables.

3. Suppose that we have a sample of data $\{(Y_i, D_i, X_i)\}_{i=1}^n$. Let $\hat{\mu}_d(x)$ denote an estimator of $\mathbb{E}[Y_i|D_i = d, X_i = x]$. Compare two imputation estimators of the average treatment effect:
$$\hat{\alpha}_1 \equiv \frac{1}{n}\sum_{i=1}^n \hat{\mu}_1(X_i) - \hat{\mu}_0(X_i)$$
$$\hat{\alpha}_2 \equiv \frac{1}{n}\sum_{i=1}^n D_i\left(Y_i - \hat{\mu}_0(X_i)\right) + (1 - D_i)\left(\hat{\mu}_1(X_i) - Y_i\right).$$

Provide intuitive descriptions of both $\hat{\alpha}_1$ and $\hat{\alpha}_2$. When will they be equal if $\hat{\mu}_d(x)$ is computed from a linear regression(s)?

4. This question is about "Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany" by Nico Voigtländer and Hans-Joachim Voth, published in *The Quarterly Journal of Economics* in 2012. The paper, as well as the data and code used in the paper, are available on Canvas.

   (a) Critically discuss the authors' empirical strategy for estimating the effect medieval antisemitism on 20th century antisemitism. What approach do the authors take? What assumptions are needed to justify their strategy? What are the primary threats to their strategy? Do you find their results credible?
   Two paragraphs: one on approach, assumptions, threats, and one on evaluation.

   (b) Replicate column (1) of Table VI. Remember to read the table notes carefully and to consult the authors' Stata code. The `nnmatch` command in Stata is well-documented in a paper by Abadie, Drukker, Herr, and Imbens (2004, *The Stata Journal*). For the purposes of this problem you may use the bootstrap to compute standard errors (although the bootstrap is not valid.) For bonus points, try to implement the formulas given in the paper by Abadie et al.

   (c) Discuss any oddities or inconsistencies between how the authors describe their results in Table VI, column (1), and how they actually implemented them in their code.

   (d) Evaluate the sensitivity of the authors' results to their choice of control variables.

   (e) Implement propensity score matching estimators of the ATE, ATU, and ATT, using the same covariates as the authors do in panel B. I leave the specifics up to you, but you might consider nearest neighbor matching on the propensity score, and/or a blocking approach. Compare your estimates to the authors' estimates. For the purposes of this problem you may use the bootstrap to compute standard errors (although the bootstrap might not be valid, depending on your approach.)
   Note: Optimization packages are not considered high-level commands for the purpose of this class, since they are not statistical in nature. You may (and should) use one to optimize a likelihood, e.g. for a logit estimator.

5. Consider the binary treatment potential outcomes model with $D \in \{0, 1\}$ and $Y = DY(1) + (1 - D)Y(0)$. For concreteness, suppose that D is whether one enrolls in a job training course, and Y is earnings at some point afterwards. Suppose that we also have a set of predetermined covariates, X. Our data consists of these variables for both workers who enrolled in the job training course, and those who did not.

   (a) Suppose that the job training experiment accepts all applicants into the course. Using the definitions in the lecture notes, show that the selection on observables model is not falsifiable.

   (b) Suppose that the job training experiment has the following structure. First, we open the program to everyone and collect a list of workers who apply to take the program. Then, we offer the program to a random subset of these applicants, but do not provide job training for any of the applicants not in this random subset. We collect data on the outcomes for workers who took the program, workers who applied to take the program but were not randomized in, and other workers who did not even apply to the program.

Explain how this structure could be used to falsify selection on observables.

6. Consider a balanced panel data setting with time periods $t = 1, \dots, T$. Suppose that $D_{it} \in \{0, 1\}$ is a binary treatment with associated potential outcomes $Y_{it}(0)$ and $Y_{it}(1)$. Assume that treatment is absorbing, so that $D_{it} = 1$ implies $D_{is} = 1$ for $s \geq t$. Let $E_i = \min\{t : D_{it} = 1\}$ denote the event time or cohort. Suppose that $\mathbb{P}[E_i = e] > 0$ for all $e = 2, \dots, T$, as well as for $e = +\infty$, so that there are always-untreated units. Assume that common trends holds, so that

$$\mathbb{E}[Y_{it}(0) - Y_{is}(0)|E_i = e] = \mathbb{E}[Y_{it}(0) - Y_{is}(0)|E_i = e']$$

for all t, s, e, and e'.

Consider the random variable $\tilde{Y}_{it}$ defined as follows:

- Regress Y_{it} on a full set of time and cohort indicators in the subset of observations (i, t) with $D_{it} = 0$.
- For all (i, t) (regardless of D_{it}), use the previous regression to construct fitted values $\dot{Y}_{it}$ based on the time period and cohort.
- Let $\tilde{Y}_{it} \equiv Y_{it} - \dot{Y}_{it}$.

Let $\text{ATT}_t(e) = \mathbb{E}[Y_{it}(1) - Y_{it}(0)|E_i = e]$.

(a) Show that

$$\mathbb{E}[\tilde{Y}_{it}|E_i = e] = \begin{cases} 0, & \text{if } t < e \\ \text{ATT}_t(e), & \text{if } t \geq e \end{cases}.$$

(b) Consider a pooled regression of $\tilde{Y}_{it}$ onto D_{it} and a constant. Let δ denote the population regression coefficient on D_{it}. Show that

$$\delta = \sum_{s=1}^{T} \sum_{e=1}^{T} \omega_s(e) \text{ATT}_s(e),$$

for some weights $\omega_s(e)$ that are non-negative and sum to one: $\sum_{s=1}^{T} \sum_{e=1}^{T} \omega_s(e) = 1$.

(c) Let $R_{it} \equiv t - E_i$ denote relative time, and let $D_{it}^j \equiv \mathbb{1}[R_{it} = j]$ be relative time indicators. Consider a pooled no-constant regression of $\tilde{Y}_{it}$ onto all post-treatment relative time indicators (so $\{D_{it}^j\}_{j=0}^{T - \min E_i}$) together with an indicator for relative time strictly less than 0. Show that the coefficients on D_{it}^j can also be written as non-negative weighted averages of $\text{ATT}_t(e)$.

(d) Discuss the content and significance of these result in the context of event studies.

7. Design a Monte Carlo experiment for an event study, and use it to illustrate the following points:
   (a) Common trends will not generally hold for both Y_{it} and $\log(Y_{it})$ simultaneously.
   (b) The static two-way fixed effects estimator will incorporate negatively-weighted treatment effects.
   (c) The dynamic two-way fixed effects estimator will incorporate negatively-weighted treatment effects unless relative time treatment effects are homogenous across cohorts. If there is such heterogeneity, then analyzing the coefficients on the leads may not be a good way to test for common trends.
   (d) In the dynamic two-way fixed effects estimator, one needs to leave out two relative time indicators. Unlike in the usual dummy variable trap, the choice of which two indicators can make a substantive difference.
   (e) Either a direct (Callaway and Sant'Anna) or imputation approach can be used to consistently estimate the cohort-weighted average treatment effect at any given relative time. Compare the finite-sample bias and standard deviations of the two estimators.

8. This paper is about "The War on Poverty's Experiment in Public Medicine: Community Health Centers and the Mortality of Older Americans," by Martha Bailey and Andrew Goodman-Bacon, which was published in *The American Economic Review* in 2015. The paper is posted on Canvas, along with the authors' replication package.
   (a) Reproduce Table 2, column (1), panel A.
   (b) On pg. 1078, the authors state:

       > *Our empirical strategy uses variation in* when *and* where *CHC programs were established to quantify their effects on mortality rates.*

       Do you agree with this statement? Why or why not?
   (c) Take the specification in Table 2, column (1), panel A, but simplify it so that it only includes year fixed effects that are not interacted with urban dummies. Use this simpler specification to repeat the estimates for Table 2, column (1), panel A. Provide a formal theoretical explanation of how interacting year fixed effects with urban dummies affects the required common trends assumption.
   (d) Create two versions of Figure 5: one that uses the specification in Table 2, column (1), panel A, and another that uses the simpler specification in the previous part.
   (e) Use the direct (Callaway and Sant'Anna) approach to estimate cohort-averaged relative time effects for the simpler no-urban specification, and produce counterparts to both Figure 5 and Table 2, column (1), panel A. Compute standard errors using a nonparametric block bootstrap.

(f) Use the imputation approach to estimate cohort-averaged relative time effects for the simpler no-urban specification, and produce counterparts to both Figure 5 and Table 2, column (1), panel A. Compute standard errors using a nonparametric block bootstrap.

(g) Now interact year fixed effects with urban dummies, as in the authors' original Table 2, column (1) specification, and attempt to use both the direct and imputation approaches. Discuss any difficulties or problems in doing so.