POL SCI 231B (Spring 2022): Problem Set 1

Prof. Thad Dunning / GSI Clara Bicalho
Dept. of Political Science
University of California, Berkeley

Distributed Friday, February 4; due Sunday, Feb 13 at 11:59 PM

Please email only one copy of your solution set per group to clara.bicalho@berkeley.edu. Make sure to *include "Problem Set [GROUP NAME]" in the email subject or body field.*

Remember to work out the problems on your own, before you meet with your group to agree on solutions.

1. **From 2016 midterm.** Consider an instrumental variables analysis where the data can be characterized by the following quantities:
   - 500 subjects with $Y_i = 0$, $D_i = 0$, $Z_i = 0$.
   - 300 subjects with $Y_i = 1$, $D_i = 0$, $Z_i = 0$.
   - 60 subjects with $Y_i = 0$, $D_i = 0$, $Z_i = 1$.
   - 40 subjects with $Y_i = 1$, $D_i = 0$, $Z_i = 1$.
   - 20 subjects with $Y_i = 0$, $D_i = 1$, $Z_i = 1$.
   - 30 subjects with $Y_i = 1$, $D_i = 1$, $Z_i = 1$.

   Here, Z_i is treatment assignment, D_i is treatment receipt, and Y_i is an outcome.

   (a) Estimate the average causal effect using intention to treat (ITT) analysis.
   (b) Among the people assigned to the treatment group ($Z_i = 1$) in this sample, exactly how many are Compliers? Why can or can't you make this calculation?
   (c) Among the people assigned to the control group ($Z_i = 0$) in this sample, exactly how many are Compliers? Why can or can't you make this calculation?
   (d) Estimate the proportion of Compliers.
   (e) Estimate the Complier average causal effect.

2. **From 2016 midterm.** Iyer (2010) is interested in the effect of direct rule by the British during the colonial period on public goods provision in India, compared to indirect rule through native or "princely" states. Lord Dalhousie, Governor-General of India from 1848 to 1856, announced that

On all occasions where heirs natural shall fail, the territory should be made to lapse and adoption should not be permitted, excepting in those cases in which some strong political reason may render it expedient to depart from this general rule.

In other words, districts in which a native ruler lacking an heir died during the period of Dalhousie's rule should be annexed by the British, according to this "Doctrine of Lapse" policy. Iyer argues that the death of an heirless ruler in the period of Dalhousie's rule can be used as an instrumental variable for direct colonial rule.

(a) What two groups of units would you compare when doing an intent-to-treat analysis here? Numerically, how is intent-to-treat analysis related to instrumental-variables analysis?

(b) Define Compliers, Always-Treats, Never-Treats, and Defiers in this context.

(c) List the assumptions that are required to estimate a Complier average causal effect. In this context, which of those assumptions seem plausible (if any), and which seem suspect (if any)? Could you use any empirical methods to evaluate their plausibility?

(d) In some of her analyses, Iyer compares districts in which heirless rulers died during Dalhousie's rule to the remaining districts. Propose a design modification that could increase the plausibility of the assumptions you described in (c) and say how it does so.

3. Professor Smedley is interested in the effect of turnout on budgetary transfers to municipalities in Japan; in particular, whether legislators reward high municipal turnout with greater transfers. However, she argues that turnout might reflect past budgetary transfers, which also influence current budgetary transfers; or that omitted variables might influence both turnout and transfers. Thus, she suggests that regressing current transfers on turnout would lead to misleading inferences about the impact of turnout.

She therefore considers two alternate approaches for studying this question:

(i) First, Smedley proposes to use election-day rainfall as an instrumental variable, in a regression of budgetary transfers on turnout.[1] Thus, she fits the regression model

$$Y_i = \alpha + \beta X_i + \epsilon_i, \tag{1}$$

where Y_i represents budgetary transfers to municipality i and X_i measures voter turnout in municipality i in the previous election. Here, α and β are parameters of the model, and ϵ_i is a random variable. The instrument is Z_i, rainfall in municipality i on the day of the previous election.

(a) Make a list of the assumptions needed for IV analysis, and interpret them (say what they mean) for this study.

(b) Do you have any potential concerns about any of the assumptions you listed in part i? Which ones, and why?

(c) Does Smedley have any tools at her disposal for investigating the plausibility of these assumptions? For example, what kind of data analysis might be helpful? Could she use qualitative methods and "shoe-leather" research to assess the validity of any of the assumptions?

(ii) Second, Smedley knows that campaign consultants have some tools at their disposal— such as in-person get-out-the-vote contacting—that influence turnout. She therefore

---

[1]This example is based loosely on Yusaku Horiuchi and Jun Saito, 2012, "Rain, Elections, and Money: The Impact of Voter Turnout on Distributive Policy Outcomes in Japan," Asia Pacific Economic Paper, No. 379. Available at SSRN: http://ssrn.com/abstract=1906951.

proposes a randomized controlled experiment, in which some municipalities will be selected at random for get-out-the-vote efforts and subsequent budgetary transfers to municipalities will be studied.

(a) In this study, what is "intent-to-treat" analysis? What is instrumental-variables analysis? What is the instrument, and what is the endogenous regressor?

(b) Make a list of the assumptions needed for IV analysis, and interpret them (say what they mean) for this study. Do you have any potential concerns about any of these assumptions? Which ones, and why?

(c) What are the potential costs and benefits of this second research design, relative to the first?

4. A researcher is interested in finding the effect of X on Y and plans to estimate the model $Y_i = \alpha + \beta X_i + \epsilon_i$. He is concerned that X_i and ϵ_i may not be independent. He thinks he has an instrument Z_i but is not sure if Z_i is independent of ϵ_i. Therefore, he proposes the following specification test: regress Y on X and Z, perform a t-test to determine whether Z significantly predicts Y, and use IV regression only if the t-statistic proves to be insignificant. Evaluate this procedure.

5. A social scientist is interested in assessing the consequences of fines for people who do not vote on voter turnout. Peru, which has compulsory voting, passed a law in 2006 lowering the fine for not voting from about US$50 to a smaller number ranging from US$6 to US$25 (depending on poverty levels of different districts).

(a) Suppose this researcher compares voting rates in two national elections that took place in 2002 and 2010 and finds that turnout declined. Can the decline be readily attributed to the effects of the new law? Why or why not?

Now, suppose that the researcher takes advantage of the fact that information about the law was not widely provided, and many voters did not in fact know about the change in the amount of the fine.[2] Before the 2010 elections, the researcher implements a national survey in which she randomly assigns respondents to two groups: one group that receives information about the lower level of the fine, and another group where no information about the legal change is given. Using publicly available data, the researcher then tracks turnout for each respondent. After the election, enumerators also recontact respondents in the group assigned to receive information to ask if they already knew about the law or were informed of it by the survey. As shown in Table 2—which also shows overall voting rates—not all respondents in the treatment group first learned of the fine through the survey; some already knew about it.

Treatment Assignment	Learned of Fine?	N	Voting (By Contacted)	Voting (Overall)
Control	No	2000	75%	75%
Treatment	Yes	500	82%	75.25%
	No	1,500	73%	

Table 1: Results from the Peru experiment.

(b) Do an intention-to-treat analysis, using the data in table 1. Also, attach a standard error to your estimate, using the "conservative formula" for the standard error of difference of means discussed in class. Interpret your finding.

(c) Now, suppose that the researcher reasons that knowing about the law is necessary for the law to change voting behavior. Thus, he conceives of treatment receipt in this experiment as "learning about the lower level of the fine." Is there any non-compliance in this experiment? Who are the Always-Treats? What about Never-Treats and Compliers?

(d) Conduct an instrumental-variables analysis to estimate the Complier average causal effect. Also, estimate the turnout (percentage voting) among Compliers in the control group.

(e) List the assumptions needed for the instrumental-variables analysis in part (iii), as in the class lecture. Which of them do you think likely to be valid in your analysis in the previous item? Which might be violated? How does your answer inform your interpretation of what we can learn from this study?

6. (Before working this question, you should read Clingingsmith et al., 2009, "Estimating the Impact of the Hajj: Religion and Tolerance in Islam's Global Gathering," *Quarterly Journal of Economics* 124 (3): 1133–1170; and you should download the replication dataset from bCourses. Note that the "Hajj Distribution" folder includes four files: 1. A codebook defining variables; 2. A Stata .do file containing a small amount of replication code (which is largely useless); 3. A Stata .dta file, which you will need to read into R; 4. a Word doc. that Clingingsmith sent to us providing further details on the construction of the seemingly unrelated regressions reported in Tables IV-VIII (as per footnote 7). Do not worry for now about 4.).

(a) On p. 1145 of Clingingsmith et al., the authors state that while survey respondents are broadly representative of the adult Pakistani population, there is "some truncation of the extremes of the socioeconomic distribution." What evidence in Table II is consistent or inconsistent with the quoted claim? Explain your answer.

(b) Conduct an intent-to-treat analysis for each of the outcome variables included in the "global Islamic practice" index (see second line of Table IV). (You can do this with unstandardized outcome variables). What do you conclude about the likely effect of the Hajj pilgrimage on the components of this index?

(c) Conceptually, do you think all of the variables included in the "global Islamic practice" index belong there? Why or why not?

(d) Discussing equation (1) on p. 1142 of their article, Clingingsmith et al. state "As long as success in the Hajj lottery only affects outcomes by inducing applicants to undertake the Hajj, this provides unbiased estimates of β^k."[3] True or false? Explain your answer.

(e) Clingingsmith et al. create normalized versions of their outcome variables to calculate average effect sizes. However, the "global Islamic practice" variables, as with the

---

[3]By "this" in the quotation, the authors appear to mean fitting an instrumental variables regression to equation (1), using success in the lottery as an instrumental variable."

components of several other indices, are all binary variables (see Table IV and codebook). Suppose τ_{unstand}^j is the unstandardized effect size and $\tau_{\text{stand}}^j = \frac{\tau_{\text{unstand}}^j}{\sigma_{\text{control}}^j}$ is the standardized effect for binary variable j, where $\sigma_{\text{control}}^j$ is the standard deviation of potential outcomes under control. Express the minimum possible absolute value of τ_{stand}^j as a function of τ_{unstand}^j. How does τ_{stand}^j vary as a function of the proportion of units in the study group for whom $Y_i^j(0) = 1$? Is this a feature or a bug? Explain your answer.