

# POL SCI 231b (Spring 2017):

## Problem Set 5 Solution Set

Prof. Thad Dunning/GSI Natalia Garbiras-Díaz

Dept. of Political Science

University of California, Berkeley

### Suggested Solutions

1. 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:

- (a) First, Smedley proposes to use election-day rainfall as an instrumental variable, in a regression of budgetary transfers on turnout.<sup>1</sup> 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,  $\alpha$

---

<sup>1</sup>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>.

and  $\beta$  are parameters of the model, and  $\epsilon_i$  is a random variable. The instrument is  $Z_i$ , rainfall in municipality  $i$  on the day of the previous election.

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

In order for rainfall to be a good instrument for turnout, given the specified regression model,

- A. There has to be a non-zero correlation between rainfall and turnout at the municipality level (a strong “first stage” relationship);
- B. Rainfall itself cannot serve as an independent cause of budgetary transfers, working through channels other than its influence on turnout (“exclusion restriction”);
- C.  $Z_i \perp \epsilon_i$ : rainfall is exogenous.

Perhaps most importantly, the model itself must be valid, which may raise some issues in this context (discussed below). This is encompassed in point B, but the concerns are broader. For example, the regression model presumes a coefficient  $\beta$  that is constant regardless of the source of variation in turnout.

- ii. **Do you have any potential concerns about any of the assumptions you listed in part A? Which ones, and why?**

(A) is easy enough to check. Smedley can regress turnout on rainfall and verify that there is a non-zero relationship between the two.

(B) presents some worries, although we might be able to explore testable implications of these concerns. For example, suppose wet regions are more prone to flooding and so receive higher levels of transfers to allow them to cope with adverse effects. Here, rainfall has an independent effect on transfers, not working through turnout. We could look to see if, on average, wet regions receive higher transfers than dry regions. If this does not seem to be the case, our worries might abate. Of course, if rainfall depresses turnout, and if turnout boosts transfers, then the two effects may counteract each other: rainfall may have a positive direct effect on transfers (violating the exclusion restriction) but negative indirect effect (working through turnout). It may be hard to tell from the data, and thus the exclusion restriction is not completely testable. A priori reasoning and evidence should be brought to bear on this topic to the extent possible, however.

(C) can never be verified, just as the independence of unobservables from the design matrix cannot be verified. However, we could raise some questions. For example, is a wetter region more likely to have rainfall on election day than a drier region? If wet regions were more likely to have rain on election day, then we might want to change the design of the study to look at within-municipality turnout changes (instrumented by rainfall) on budgetary transfers.

As for the constant coefficient  $\beta$ , it is possible that variation in turnout induced by election-day rainfall has different effects on transfers than variation in turnout due to other factors. For example, perhaps legislators discount depressed turnout that is due to election-day deluges (“we won’t punish voters in that municipality because it wasn’t their fault they could not turn out to vote for us”). The point is that variation in the endogenous variable due to the instrument may or may not have the same effect as variation in the endogenous variable not due to the turnout. For further discussion of this point, see e.g. Dunning (2012: 269-277 as well as Appendix 9.1); for discussion of this example, see Exercise 7.2 (pp. 230-31).

- iii. **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?**

There are observable implications of the claim that variation in the endogenous regressor induced by the instrument may not have the same effect as variation unrelated to the instrument. For instance, one could interview legislators to ask if they know why particular municipalities did or did not exhibit high turnout, or what election-day rainfall was like in each municipality. To condition punishment for low turnout on whether the low turnout was weather-induced or not, legislators have to be aware of what election-day rainfall was like in particular municipalities. Thus, such interviews might provide qualitative information (causal-process observations) that are useful for evaluating whether legislators do or do not treat rainfall-induced turnout shocks differently from other sources of depressed turnout.

- (b) **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 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.**

- i. **In this study, what is “intent-to-treat” analysis? What is instrumental-variables analysis? What is the instrument?**

Intent-to-treat analysis compares the groups created by the randomization—namely, municipalities selected at random for get-out-the-vote efforts (the assigned-to-treatment group) and those assigned to the control condition. Thus, we compare budgetary transfers to these municipalities without accounting for differential turnout rates: we simply compare budgetary transfers in the two groups created by the randomization.

We here also use instrumental-variables to answer the question, “what is the effect of higher turnout on budgetary transfers, among municipalities whose turnout is influenced by get-out-the-vote campaigns?” This is an “encour-

agement design." It involves an extension of our model for non-compliance (because municipal turnout is not a 0-1 variable) but involves a similar principal. Get-out-the-vote campaigns may be effective in some municipalities but not in others. Instrumental-variables analysis will estimate the effect of turnout in the former set of municipalities. The instrument is assignment to get-out-the-vote campaigns; turnout is the endogenous independent variable.

- ii. **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?**

We will have to make similar assumptions as in part (a)—for example, that the exclusion restriction holds (there is no other channel through which get-out-the-vote campaigns affect transfers, other than turnout). Here, the assumption that the instrument is independent of unobservables that affect transfers is more plausible—because the instrument (assignment to a get-out-the-vote campaign) is truly randomly assigned.

- iii. **What are the potential costs and benefits of this second research design, relative to the first?**

One might point to various costs and benefits. The benefit and strength to this course of action is that there would be an exogenous source of variation to study, provided that get-out-the-vote does increase turnout in a meaningful way. With a sound model of causation, causal claims might be valid. The simplest approach to this analysis would be to study the average causal effect of treatment assignment, i.e. the intention-to-treat parameter. Estimating this effect would involve comparing average outcomes across municipalities assigned to treatment with those assigned to control. (For comparison's sake, it might be more relevant to study the average transfer per person in the municipality, as some municipalities might be larger and receive more money as a result.)

The costs of an experiment in terms of actual dollars spent will be much higher to the researcher. Is the question compelling enough to warrant such expenditure? There is always the potential that the effect sizes would be too small to observe, if the get-out-the-vote induced turnout is not big. In this case it might not be worth the money. Funding agencies and advisors will have to help Smedley decide.

One weakness is that the question is somewhat altered by the new design. Whereas originally the question was akin to "Do places with higher levels of participation get rewarded by the central government?" the new question is "Do places where get-out-the-vote campaigns are run get higher levels of government transfers?". This latter question seems narrower.

2. **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  $\epsilon_i$  may not be independent. He thinks he has an instrument  $Z_i$  but is not sure if  $Z_i$  is independent of  $\epsilon_i$ . 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.**

This is a bad idea. In the background, we have a system of two equations:

$$Y_i = \alpha + \beta X_i + \gamma Z_i + \delta_i \text{ and} \quad (2)$$

$$X_i = \psi + \phi Z_i + \nu_i, \quad (3)$$

where  $\alpha$  and  $\psi$  are intercepts,  $\beta$  and  $\gamma$  are regression coefficients, and  $\delta_i$  and  $\nu_i$  are unobserved error terms. The specification test considers the null hypothesis that  $\gamma = 0$ . The difficulty is that unobserved causes of  $X_i$ —captured by  $\nu_i$ —are likely to be correlated with unobserved causes of  $Y_i$ , that is,  $\delta_i$ . (Indeed, the researcher may be using instrumental variables out of concern for omitted-variables bias). If  $\nu_i$  and  $\delta_i$  are correlated,  $X_i$  is endogenous in equation (2), so the estimator of  $\beta$  is biased. Moreover,  $X_i$  and  $Z_i$  are correlated (as long as there is a “first-stage” relationship between the instrument  $Z_i$  and the endogenous regressor  $X_i$ ). Thus, bias “propagates”: the estimators of  $\beta$  and  $\gamma$  are both biased. Unfortunately, the direction of the bias (which depends on the sign of the correlation between  $\nu_i$  and  $\delta_i$ ) and its magnitude are unknown. The  $t$ -statistic for the coefficient  $\hat{\gamma}$  is therefore uninformative, and so is the specification test.

This is the classic bias from mediation analyses in disguise. Unfortunately, parsing the relative influence of  $X_i$  and  $Z_i$  through such a strategy is not generally illuminating.

3. **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?**

Here, we have pre- and post-fine observations without a comparison group never subjected to the reduced fine. To make their case, the authors would have to argue that the election in 2002 is a relevant counterfactual for 2010 by arguing that nothing major changed between the two elections that would have also lowered turnout in the next election, independently of the reduced fines. But many things besides the fine probably changed between 2002 and 2010. Note also that in the description here, the treatment was not the same across districts (some districts faced a much larger reduction in absolute and percentage terms), so deciding what the treatment is in this study is not cut and dry.

- (b) 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.<sup>2</sup> 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.

- i. **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.**

We can estimate the intention-to-treat parameter by taking the average outcome in the assigned-to-treatment group  $Y^T$  and subtracting from it the average outcome in the assigned-to-control group  $Y^C$ .

$$\widehat{ITT} = Y^T - Y^C = 75.25 - 75 = 0.25$$

A conservative estimator for the standard error of  $\widehat{ITT}$  is given by

$$\widehat{SE}(\widehat{ATE}) = \sqrt{\frac{\widehat{\sigma}_T^2}{m} + \frac{\widehat{\sigma}_C^2}{N - m}} \quad (4)$$

We can estimate  $\widehat{\sigma}_T^2$  and  $\widehat{\sigma}_C^2$  using the formula for the variance of a 0-1 box and using the observed fraction of voters under each condition as our best guess for the fraction in the box.

$$\widehat{\sigma}_T^2 = (0.7525) * (0.2475) = 0.1862$$

$$\widehat{\sigma}_C^2 = (0.75) * (0.25) = 0.1875$$

$$\widehat{SE}(\widehat{ATE}) = \sqrt{\frac{0.1862}{2000} + \frac{0.1875}{2000}} = 0.01367 \quad (5)$$

---

<sup>2</sup>This question is based loosely on a paper by Gianmarco León (2012).

The t-statistic is  $\frac{0.0025}{0.01367} = 0.18$ . The t-statistic is very small, thus we would fail to reject the hypothesis that the treatment had no effect on voting. Note moreover that with 4,000 respondents in the study group, we should have plenty of *statistical power* to reject the null hypothesis, but at 0.25% the substantive effect is very small.

- ii. **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?**

Always-Treats learn about the lowered fines no matter which treatment group they are assigned to (e.g., they pay attention to the news). Compliers only learn about the lowered fines if they were assigned to the treatment group. Never-Treats never learn about the lowered fines no matter which group they are assigned to. In Table 1, the respondents in the treatment group marked “Yes” learned about the lowered fine from the survey; they are the Compliers. Those marked “No,” on the other hand, already knew about the change in the fine, so they are the Always-Treats. In this study, there are no Never-Treats: those who were contacted were told about the fine. (We are finessing the issue of non-response here: assume all 2000 people assigned to the treatment group answered the door, the issue here is whether they already knew about the lowered fine).

Note: designs like this one are sometimes called “encouragement designs”: often, subjects are given some information which may shape their behavior. For example, the effect of watching television on political interest may be assessed by encouraging some people at random to watch television and comparing their subsequent political interest to a randomly assigned control group. With encouragement designs, as we see next, instrumental-variables analysis can come into the picture: the encouragement to watch television (the treatment assignment) is the instrument; watching television is treatment receipt; and political interest is the outcome.

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

An IV analysis requires estimating the average causal effect of treatment assignment (the ITT parameter), and the proportion of Compliers in the study group. We estimate the first by taking the difference-of-means across treatment assignment, where  $Y^T$  is the average outcome in the assigned to treatment group and  $Y^C$  the average outcome in the assigned to control group. We then divide this by fraction of people in the treatment group who “accept” the treatment,  $X^T$ , minus the fraction of people in the control group who

accept the treatment (i.e. they cross-over into the treatment group).

$$\widehat{\text{CACE}} = \frac{Y^T - Y^C}{X^T - X^C} \quad (6)$$

In our study,  $500/2000=0.25$  percent of people in the treatment group learned about the fine from the survey. Thus, we can estimate the fraction of Compliers in the control group as 0.25. Because the assigned-to-treatment group is randomly sampled from the study group of 4,000 respondents, this is an unbiased estimator for the proportion of Compliers in the study group; because the assigned-to-control group is also randomly sampled from that study group, it is our best guess for the proportion of Compliers in the study group.

Thus  $\widehat{\text{CACE}} = \frac{0.25}{0.25} = 1$ . Compliers are moved by 1 percentage point by the treatment.

Note also that we can estimate the turnout (percentage voting) among Compliers in the control group. In the treatment group, the turnout among Never-Treats (those who already knew about the fine) is 73%. Thus, for the estimated 1,500 Never-Treats in the control group (that is, 75% of the control group), the estimated turnout is also 73%. This sets up an algebraic equation that we can use to solve for the estimated turnout among Compliers in the control group:

$$73\% * (0.75) + x\% * (0.25) = 75\%. \quad (7)$$

Thus,

$$x\% = \frac{75\% - 73\% * (0.75)}{0.25} = 81\%. \quad (8)$$

Notice that if we now compare turnout among Compliers in the assigned-to-treatment and assigned-to-control groups, we arrive at the same estimate of the CACE:  $82\% - 81\% = 1\%$ .

- iv. **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?**

Let's consider the assumptions one by one:

- A. **Potential Outcomes:** Outcomes under each treatment assignment are fixed attributes of each unit. This is a basic modeling assumption here. Maybe ok.
- B. **Compliance Types:** In general, the study group may be comprised of up to three types: Compliers, Never-Treats, and Always-Treats. Here, we ruled out the Never-Treats, since people assigned to the treatment group who didn't know about the lowered fine learned about it through the survey. Of course, we are finessing the point: maybe they didn't really learn, weren't paying attention, etc. We take their self-reports of whether



they learned about the lowered fine through the survey at face value.

- C. **No Defiers:** No subjects do the opposite of what they are told (the opposite of treatment assignment). This seems plausible—it’s hard to imagine subjects learning about the fine if they are assigned to the control group but refusing to learn about it if assigned to treatment.
  - D. **Random Sampling:** Units are sampled at random from the study group and assigned to treatment or control. This is secured by random assignment of the 4,000 units to treatment or control.
  - E. **Non-Interference/SUTVA:** Each unit’s treatment receipt and outcome depends only on its own treatment assignment (and not on the assignment of other units). Here we’ve got 4,000 survey respondents spread across different Peruvian districts. How they are selected for the survey may influence the possibility for interference. If they do not interact, interference seems unlikely. But, if clusters of neighborhoods are selected for the survey (e.g., through a multistage cluster sample), interference could arise. For instance, people assigned to the control group might learn about the lowered fine from their neighbors assigned to the treatment group, which might affect their voting.
  - F. **Exclusion restriction:** Treatment assignment only affects outcomes through treatment receipt. This seems more debatable. For instance, it might be that assigned to treatment has affects on voting, other than through learning about the lowered fine. Perhaps people who are told about the lowered fine think their voting is being monitored—which might cause them to vote at higher rates. (This would make the treatment group voting mean appear larger than it would were it the case that treatment assignment affected voting only through learning about the fine).
4. (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.; we will discuss that later in the course).
- (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.

It is not clear what evidence in Table II supports this assertion. For the (sample of) the adult Pakistani population in the districts where the researchers worked, the s.d. of (log) monthly expenditures is 0.0641; for the full sample with which the researchers worked, it is 0.783. So there is greater variation among survey respondents than in the sample of the adult Pakistani population, which appears inconsistent with truncation at the extremes of the socioeconomic distribution. Some additional information about the distribution of income or expenditures (for instance, deciles of the Pakistani population and survey respondents) could be helpful.

- (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?**

```
setwd("~/Dropbox/Academic/UC_Berkeley/GSI/PS_231B/Problem_sets/PS 5/Hajj Distri
library(foreign)

hajj <- read.dta("hajj_public.dta")

index_vars <- c("x_s14aq1","x_s14aq3","x_s14aq4","x_s14aq5",
               "x_s14aq6","x_s14bq2","x_s14aq8","x_s14aq9",
               "x_s14aq12","x_s14aq13")

questions <- c("How often do you pray Namaz?",
               "How often do you do tasbih after namaz?",
               "How often do you pray Namaz in congregation in the mosque?",
               "How often did you pray Namaz in congregation in the mosque
               last Sunday?",
               "Do you pray "Tahajjud Namaz"?",
               "Are/were you able to read the Qu'ran?",
               "How frequently do you recite the Qu'ran?",
               "When you get together with your friends, would you say you
               discuss religious matters frequently, occasionally or never?",
               "Which of the following best applies to your experience of the
               most recent Ramadan?",
               "How often did you fast outside of Ramadan during the past year?")

for (i in 1:10){

  cat(questions[i])

  outcome <- hajj[, which(names(hajj)==index_vars[i])]
```

```

    print(t.test(outcome ~ hajj$hajj2006))
}

## How often do you pray Namaz?
## Welch Two Sample t-test
##
## data: outcome by hajj$hajj2006
## t = -5.6679, df = 1172.6, p-value = 1.819e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.15188284 -0.07377051
## sample estimates:
## mean in group 0 mean in group 1
## 0.7599388 0.8727655
##
## How often do you do tasbih after namaz?
## Welch Two Sample t-test
##
## data: outcome by hajj$hajj2006
## t = -2.7315, df = 1386.7, p-value = 0.006384
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.11818881 -0.01938728
## sample estimates:
## mean in group 0 mean in group 1
## 0.5305810 0.5993691
##
## How often do you pray Namaz in congregation in the mosque?
## Welch Two Sample t-test
##
## data: outcome by hajj$hajj2006
## t = -2.0548, df = 1463.8, p-value = 0.04008
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.084383176 -0.001957614
## sample estimates:
## mean in group 0 mean in group 1
## 0.2018349 0.2450053
##
## How often did you pray Namaz in congregation in the mosque
## last Sunday?
## Welch Two Sample t-test
##
## data: outcome by hajj$hajj2006

```

```

## t = -2.9827, df = 1216.9, p-value = 0.002914
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.12937219 -0.02670775
## sample estimates:
## mean in group 0 mean in group 1
##      0.3010949      0.3791349
##
## Do you pray "Tahajjud Namaz"?
## Welch Two Sample t-test
##
## data: outcome by hajj$hajj2006
## t = -6.6805, df = 1492.6, p-value = 3.351e-11
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.2050086 -0.1119441
## sample estimates:
## mean in group 0 mean in group 1
##      0.2737003      0.4321767
##
## Are/were you able to read the Qu'ran?
## Welch Two Sample t-test
##
## data: outcome by hajj$hajj2006
## t = -1.3718, df = 1378.9, p-value = 0.1703
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.08027866 0.01420491
## sample estimates:
## mean in group 0 mean in group 1
##      0.6462481      0.6792850
##
## How frequently do you recite the Qu'ran?
## Welch Two Sample t-test
##
## data: outcome by hajj$hajj2006
## t = -0.53688, df = 1237.1, p-value = 0.5914
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.06527777 0.03722667
## sample estimates:
## mean in group 0 mean in group 1
##      0.6155172      0.6295428
##

```

```

## When you get together with your friends, would you say you
##           discuss religious matters frequently, occasionally or never?
## Welch Two Sample t-test
##
## data:  outcome by hajj$hajj2006
## t = -3.2219, df = 1149.5, p-value = 0.001309
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.08361396 -0.02032074
## sample estimates:
## mean in group 0 mean in group 1
##      0.8684628      0.9204301
##
## Which of the following best applies to your experience of the
##           most recent Ramadan?
## Welch Two Sample t-test
##
## data:  outcome by hajj$hajj2006
## t = -1.1809, df = 1322.5, p-value = 0.2378
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.04622333  0.01148443
## sample estimates:
## mean in group 0 mean in group 1
##      0.9006116      0.9179811
##
## How often did you fast outside of Ramadan during the past year?)
## Welch Two Sample t-test
##
## data:  outcome by hajj$hajj2006
## t = -2.0223, df = 1544.6, p-value = 0.04332
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.0488028244 -0.0007442482
## sample estimates:
## mean in group 0 mean in group 1
##      0.05198777      0.07676130

```

A majority of the components of the index shows that the pilgrimage had an effect, although there is some variation. For example, there seems to be no effect of the pilgrimage on the ability to read the Qu'ran or the frequency with which individuals recite it.

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

The ten variables in the index are (per the note to Table IV): “How frequently do you: pray, do tasbih after prayer, pray in the mosque? Did you pray in the mosque last Sunday? Do you pray optional night prayers? Can you read the Qu’ran? How frequently do you: read the Qu’ran? discuss religious matters? keep fast during Ramadan? keep fast outside Ramadan?” These seem to be general indicators of religiosity, and they might be contrasted from the elements in the “local Islamic practice” index. But it is not clear that they all tap “global” Islamic practice, and they may involve a mix of religious belief and participation.

- (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  $\beta^k$ ."**<sup>3</sup> **True or false? Explain your answer.**

False. The IVLS estimator is *consistent* under the assumptions of the model; it suffers however from small-sample bias (due to the fact that the IVLS estimator is a ratio of random variables).

- (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 components of several other indices, are all binary variables (see Table IV and codebook). Suppose  $\tau_{\text{unstand}}^j$  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  $\tau_{\text{stand}}^j$  as a function of  $\tau_{\text{unstand}}^j$ . How does  $\tau_{\text{stand}}^j$  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.**

From  $\tau_{\text{stand}}^j = \frac{\tau_{\text{unstand}}^j}{\sigma_{\text{control}}^j}$ , we can see that the absolute value of  $\tau_{\text{stand}}^j$  is minimized when  $\sigma_{\text{control}}^j$  is as large as possible. Now, the variance of a binary variable is

$$p(1 - p),$$

where  $p$  is the proportion of 1s, so a binary variable has maximum variance when  $p = 1/2$ . Since  $\sigma_{\text{control}}^j$  is the standard deviation (the square root of the variance) of the potential outcomes under control for variable  $j$ , the absolute value of  $\tau_{\text{stand}}^j$  is minimized when the proportion of units for whom  $Y_i^j(0) = 1$  is  $1/2$ . As this proportion becomes greater than or less than  $1/2$  (and fixing  $\tau_{\text{unstand}}^j$ ),  $\tau_{\text{stand}}^j$  gets larger in absolute value.

One might see this as a bug, because the standardized effect size depends on whatever the variance of potential outcomes in the control group happens to be. Of course, this is true for any standardized effect, whether  $j$  is binary or not.

---

<sup>3</sup>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.