

Causal Inference Midterm

Colin Wick

3/11/2021

```
knitr::opts_knit$set(message=F,warning=F)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3      v purrr   0.3.4
## v tibble  3.0.6      v dplyr   1.0.4
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

Question 1

```
mydf <- data.frame(matrix(c(4,3,10,8,6,8,2,1,2,4,1,1,6,2,4,9,10,3,5,8,3,3,4,6,1,9,5,2,9,1),ncol
= 3,byrow = TRUE))
names(mydf) <- c("Y1","Y0","X")
```

Using the ATE definition $E(Y^1) - E(Y^0)$

```
ATE <- mean(mydf$Y1) - mean(mydf$Y0)
ATE
```

```
## [1] -0.9
```

The ATE comes out to -0.9, meaning that exercise is associated with a -0.9 point decrease in reported health, on average.

Question 2

The Switching Equation is $Y_i = D_i Y_i^1 + (1 - D_i) Y_i^2$

```
mydf <- mydf %>%
  mutate(D = ifelse(Y1 > Y0,1,0),
         switch = ifelse(Y1 > Y0,Y1,Y0))
mydf
```

```
##      Y1 Y0  X D switch
## 1    4  3 10 1      4
## 2    8  6  8 1      8
## 3    2  1  2 1      2
## 4    4  1  1 1      4
## 5    6  2  4 1      6
## 6    9 10  3 0     10
## 7    5  8  3 0      8
## 8    3  4  6 0      4
## 9    1  9  5 0      9
## 10   2  9  1 0      9
```

```
ATT <- mean(mydf$Y1[mydf$D == 1]) - mean(mydf$Y0[mydf$D == 1])
ATU <- mean(mydf$Y1[mydf$D == 0]) - mean(mydf$Y0[mydf$D == 0])
```

The outcome of the switching equation is stored in our dataframe and yields $ATT = 2.2$ and $ATU = -4$.

$ATT \neq ATU$ but this is common and expected.

Question 3

```
SDO <- mean(mydf$Y1[mydf$D == 1]) - mean(mydf$Y0[mydf$D == 0])
pi <- sum(mydf$D)/10
s_bias <- mean(mydf$Y0[mydf$D == 1]) - mean(mydf$Y0[mydf$D == 0])

decomp <- data.frame(t(c(SDO,ATE,s_bias,(1-pi)*(ATT-ATU))))
names(decomp) <- c("SDO","ATE","Selection","Heterog. TE Bias")

knitr::kable(decomp,caption = "Decomposition") %>% kableExtra::kable_classic_2()
```

Decomposition

SDO	ATE	Selection	Heterog. TE Bias
-3.2	-0.9	-5.4	3.1

We have $SDO = ATE + E(Y^0|D = 1) - E(Y^1|D = 0) + (1 - \pi)(ATT - ATU)$ as the decomposition and each value is listed above.

Question 4

a

The two key assumptions necessary for 2SLS to work (along with standard Gauss-Markov assumptions, of course) are:

$$Cov(Z, \epsilon) = 0 \text{ and } \delta \neq 0$$

The standard 2SLS estimator for β is $\frac{Cov(Y,Z)}{Cov(X,Z)} = \hat{\beta}_{IV}$. In this case (two dummy variables), the estimator is called a Wald estimator, which can be written as $\frac{E(Y|Z=1)-E(Y|Z=0)}{E(X|Z=1)-E(X|Z=0)} = \beta$

b

The Weak Instrument problem is one where the relationship between the exogenous instrument and the endogenous covariate are not strongly correlated enough to generate unbiased and/or significant estimates. Having a weak instrument implies the first-stage estimator is insignificant or extremely small, which will throw off the second-stage estimate. The most common test for weak instruments is an F-test, where an F-statistic less than 10 should be considered too weak to use in a regression.

c

$$\begin{aligned}\delta &= \frac{\frac{1}{n} \sum (Z_i - \bar{Z})(Y_i - \bar{Y})}{\frac{1}{n} \sum (Z_i - \bar{Z})(X_i - \bar{X})} \\ &\quad \text{substituting in the formula for Y} \\ &= \frac{\frac{1}{n} \sum (Z_i - \bar{Z})(\beta X_i + \epsilon_i - \beta \bar{X} - \bar{\epsilon})}{\frac{1}{n} \sum (Z_i - \bar{Z})(X_i - \bar{X})} \\ &\quad \text{rearranging to group X and e} \\ &= \frac{\frac{1}{n} \sum (Z_i - \bar{Z})(\beta(X_i - \bar{X}) + (\epsilon_i - \bar{\epsilon}))}{\frac{1}{n} \sum (Z_i - \bar{Z})(X_i - \bar{X})} \\ &\quad \text{split the terms} \\ &= \frac{\frac{1}{n} \sum \beta(Z_i - \bar{Z})(X_i - \bar{X})}{\frac{1}{n} \sum (Z_i - \bar{Z})(X_i - \bar{X})} + \frac{\frac{1}{n} \sum (Z_i - \bar{Z})(\epsilon_i - \bar{\epsilon})}{\frac{1}{n} \sum (Z_i - \bar{Z})(X_i - \bar{X})} \\ &\quad \text{cancel and simplify notation} \\ &= \beta + \frac{Cov(Z, \epsilon)}{Cov(Z, X)}\end{aligned}$$

The hanging term at the end is the small sample bias, which will tend to 0 as n increases because, by definition, $Cov(Z, \epsilon) = 0$ but for small samples this may not hold immediately, while $Cov(Z, X)$ must increase with increasing N.

d

Using observational data, there are two classical cases. First is unobserved/unobservable covariates, like (in the case of labor); grit, tenacity, aptitude. Things that are correlated with unobservables can serve as a way to estimate the impact without having to directly observe.

The other classical example is simultaneity, where both supply and demand are determined by each other, so neither can be used to estimate the other. An IV can cut through, in this case, and isolate one of these. For example, rainfall which might affect supply directly but would not influence the shopper's ability to go to the market 6 months later.

Question 5

These two methods accomplish the same outcome at different stages of the analysis process. A fixed-effects model for time-invariant heterogeneity creates a unique intercept for each panel group $\alpha_i + \beta_0$. Time demeaned variables do the same thing, but on the dependent-variable side rather than the intercept side, scaling the outcome by α_i before the regression is run. In both cases, the same α_0 is removed from the line to align the relationship between the panel group and dependent variable.

Question 6

a

It would be inappropriate to compare formerly-incarcerated children to the general population of non-incarcerated people because there are confounding factors that may obscure the effect of interest (human capital formation). For example, those who are more likely become incarcerated as children may have lower human capital & educational attainment rates than the general population regardless of whether they are actually incarcerated.

The most obvious reason why this is not good research design is social determinism and heterogenous application of law enforcement in the United States.

b

The researchers exploit the randomized judge assignment & corresponding variance in incarceration rates among judges as a source of exogenous but correlated way with term length & rate. By using this 2SLS estimator, the researchers are able to look at outcomes on the margin of incarceration without confounders or selection bias in the population.

c

The exclusion restriction is the assumption that the instrument is not correlated with unobservables in the original model. This is another way of saying that the instrument can't be a covariate, that it would be just as useful as another term in a simple OLS model. In this model, it can't be the case that judge assignment affects human capital *directly*.

The independence assumption states that the instrument can't be influencing the distribution of treatments. The instrument has to be randomly assigned to outcomes and can't be correlated with one outcome or another. In this paper, variance in judge assignment is the source of exogenous variance. The judge assignment itself must be random with respect to human capital of the individuals. That is, there can't be a process by which administrators are assigning judges based on human capital of individuals.

Monotonicity states that the instrument must be pushing all observations in the same direction. It can't be the case that getting a lottery sometimes harms and sometimes helps cases, it must work in one direction or the other. Lower incarceration rates/length must be positive for human capital, and more lenient judges must assign lower

incarceration rates/terms.

d

The LATE estimate is the local average treatment effect. In essence, this captures the effect, on the margin, of the treatment. In an IV, it measures the local difference for the cases for those who would not have been treated if not for the instrument. Its difference from a DiD effect, which measures the change in rate/intercept pre- and post-treatment or a simple OLS estimator which measures a linear relationship with a variable.

In this case, the compliers are those who might have otherwise gotten a more lenient sentence if they had a different judge AND those who would have gotten a more strict sentence with a different judge. The population of people for whom the random discretion of the assigned judge was the deciding factor in sentencing.

Question 7

a

The running variable would be personal income and the cutoff would be \$20k.

b

Sorting on the running variable means that people see a program cutoff and target their income/hours/etc in order to meet the requirements to participate. In this project we can visually assess whether there is sorting by creating a histogram and seeing whether there is obvious bunching before the \$20k point. We can formally test this with a McCrary test to ensure smoothness along the running variable around this point.

c

Barrecca, Guldi, Lindo and Waddell suggest using a “donut hole” RDD design for uniformly bunched random variables. In the case of birthweights, it is obvious that the underlying data is not sorted to meet a specific mark, rather the measurement technology or exigence around measurement does not create a smooth distribution of birthweights. To accomodate this, they employ a strategy where they are able to exclude points in the immediate vicinity of bunch points while using all the points around them in their regression.

d

The smoothness assumption states that, in the absence of treatment, there would not be a sharp gap in outcomes. In this example, we would say that the relationship between child health outcomes and income would have some continuous (likely increasing) relationship for all income levels. Some polynomial formula of income and child health could fit the data well. In having this assumption, we would then see that there is a break in outcomes across the cutoff.

e

Covariate balance tests see whether the cutoff point is systemically affecting other covariates, which would indicate that an RDD is not viable. This is not the same as L and R sides of the cutoff having different summary statistics wrt a given covariate. Rather it is focused on making sure there is no sharp discontinuity in covariates.

Secondly, placebo tests will help validate the continuity assumption. If a McCrary test holds at the cutoff point, then it should similarly hold for other arbitrary cutoffs near the desired cutoff or on any arbitrary place of the running variable.

Question 8

a

In this case, there will be a selection bias for families that opt-in to the program even once they are granted access. If it was the case that they were forced to participate then this effect would not happen. "Lottery win" will be the instrument used.

1. SUTVA - Treatment cannot affect other treated/untreated
2. Independence - Assignment of treatment cannot be correlated with outcome
3. Exclusion Restriction - Treatment cannot be a relevant second-stage covariate (correlated with error)
4. First-stage relevance - Treatment must be correlated with endogenous variable $\beta_{first_stage} \neq 0$
5. Monotonicity - Treatment must move outcomes in the same direction

1. Technically, SUTVA may be violated if the lottery size is very small and new families can be drawn from lack of participation, but for a statewide trial, this effect would be negligible. That is, opting-out of the lottery may have a non-zero effect on the odds of another family winning the lottery, but for large n (winner pool) and N (total population) this is negligible.
2. Lottery win is not influenced by child health outcomes or whether there is uptake of the program.
3. "Lottery win" has no direct relationship with child health outcomes
4. "Lottery win" is correlated with endogenous variable of program uptake
5. Winning the lottery has at least no effect on family nutrition/uptake. Uptake of the program can be assumed to improve health outcomes.

b

Compliers - those who are assigned treatment and take treatment. People who would not have taken the food unless they won the lottery but since they won they will.

Defiers - those who are assigned treatment and opt-out of the treatment. People who would refuse the program if they won, and also those who could use the food if they lost.

Never-takers - people who would never participate in the program under any circumstance.

Always-takers - people who will find a way into the food assistance program even if it's against the rules

The LATE IV parameter estimates the effect on the Compliers (those who were operating on the margin of treatment)

c

There cannot be always-takers in this program. In a sense there could be, if they forged a lottery ticket or broke the rules of the lottery/program in some way. This is either impossible or a negligible population.