GV481: Quantitative Analysis for Political Science

Felix Wortmann Callejón, May 26, 2023

1 Potential Outcomes Framework

Causal effect is a change in some feature of the world Y that would result from a change to some other feature of the world D. For each i, two **potential outcomes**. Y_{1i} is potential outcome for i if $D_i = 1$ and Y_{0i} is potential outcome for i if $D_i = 0$. **Observed outcomes** Y_i are realised as

$$\begin{aligned} Y_i &= D_i \cdot Y_{1i} + (1 - D_i) \cdot Y_{0i} \\ &= \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases} \end{aligned}$$

On the individual level, treatment effect is

$$\tau_i = Y_{1i} - Y_{0i}$$

Because $\tau_i = Y_{1i} - Y_{0i}$ are unobservable, interest in different estimand, the average treatment effect (ATE)

$$\delta = E[\tau_i] = E[Y_{1i} - Y_{0i}]$$

Other types of average treatment effects:

$$\begin{aligned} \tau_{ATT} &= \overbrace{E[Y_{1i} - Y_{0i} | D_i = 1]}^{\text{Average treatment effect on the treated}} \\ \tau_{ATC} &= \overbrace{E[Y_{1i} - Y_{0i} | D_i = 0]}^{\text{Average treatment effect on the controls}} \\ \tau_{CATE} &= \overbrace{E[Y_{1i} - Y_{0i} | X_i = x]}^{\text{Conditional average treatment effect}} \end{aligned}$$

Crucial assumption underlying all causal inference is **SUTVA**. SUTVA is assumption of causal effect stability: no multiple versions of treatment; potential outcomes of i unaffected by changes in the treatment exposures of $j \; \forall \; j \neq i$. Simplified, for $Y_{d_i d_j i}$ the SUTVA means:

$$Y_{11i} = Y_{10i} \text{ and } Y_{01i} = Y_{00i}$$

SUTVA is likely to be violated when units are in close contact in social and physical space and when dilution/concentration effects arise that one can assume would result from changes in the prevalence of the treatment.

2 Measuring Uncertainty

2.1 Bias and precision

$$\text{Estimator} = \text{Estimand} + \underbrace{\text{Bias}}_{\text{Systematic error}} + \underbrace{\text{Noise}}_{\text{Sampling error}}$$

2.2 Quantifying precision and Hypothesis Testing

For a refresher on NHST and statistical inference see section 1 of my MY452 cheat sheet.

2.3 OLS Regression

For a refresher on OLS regression see section 2 of my MY452 cheat sheet.

2.4 Robust Standard Errors

To account for variability in sample realisation and realisation of potential outcomes, use **robust standard errors**, the closed form solution of which is

$$SE_{\widehat{ATE}} = \sqrt{Var\left[\widehat{ATE}\right]}$$

with

$$\begin{aligned} Var\left[\widehat{ATE}\right] = &\left(\frac{N - N_1}{N - 1}\right)\frac{Var[Y_{1i}]}{N_1} \\ &+ \left(\frac{N - N_0}{N - 1}\right)\frac{Var[Y_{0i}]}{N_0} \\ &+ \left(\frac{1}{N - 1}\right)2Cov[Y_{1i}, Y_{0i}] \end{aligned}$$

with N units, N_1 assigned to treatment and $N_0 = N - N_1$. Since $Cov[Y_{1i}, Y_{0i}]$ is unobservable, default to a conservative estimator of SE, by plugging in maximum value of $Cov[Y_{1i}, Y_{0i}] = 1$. This conservative estimator is formally defined by

$$\widehat{SE}_{\widehat{ATE}} = \sqrt{\frac{\widehat{Var[Y_{1i}]}}{N_1} + \frac{\widehat{Var[Y_{0i}]}}{N_0}}$$

with estimators of the sample variances given by

$$\begin{aligned} \widehat{Var[Y_{1i}]} &= \frac{1}{N_1 - 1}\sum_{i|D_i=1}^{N} Y_{1i} - \left(\frac{\sum_{i|D_i=1}^{N} Y_{1i}}{N_1}\right)^2 \\ &= \hat{\sigma}^2_{Y|D_i=1} \end{aligned}$$

and

$$\widehat{Var}\left[Y_{0i}\right] = \frac{1}{N_0 - 1}\sum_{i|D_i=0}^{N} Y_{0i} - \left(\frac{\sum_{i|D_i=0}^{N} Y_{0i}}{N_0}\right)^2$$
$$= \hat{\sigma}^2_{Y|D_i=0}$$

3 Randomised Experiments

In experimental setting, know that units were randomly assigned into control and treatment groups. Hence estimate ATE simply by comparing average observed outcomes.

$$\tau_{ATE} = E\left[Y_{1i} - Y_{0i}\right] = E\left[Y_{1i}\right] - E\left[Y_{0i}\right]$$
$$= \underbrace{E\left[Y_i|D_i = 1\right] - E\left[Y_i|D_i = 0\right]}_{\text{Difference in means}}$$

3.1 Blocking

When randomly assigning treatments to participants, one might get unlucky and have unbalanced groups along key confounders. One can balance characteristics manually by blocking. Separate sample of N into J groups **before** randomisation. Then *within* each block, randomly assign units to the treatment and control groups.

$$\tau_{ATE} = \sum_{j=1}^{J} \frac{N_j}{N}\tau_j$$

an unbiased estimator for this quantity is

$$\hat{\tau}_{ATE} = \sum_{j=1}^{J} \frac{N_j}{N}\hat{\tau}_j$$

When analyzing a blocked randomised experiment with regression, block dummies should be included when analysing results

$$Y_i = \tau D_i + \sum_{j=1}^{J} \beta_j B_{ij} + \epsilon_i$$

where B_j is a dummy for j-th block. However, when probability of receiving treatment $p_{ij} = P(D_{ij} = 1)$ is not equal across blocks, possibly need to weight treated units by inverse of probability of being assigned to treatment $\frac{1}{p_{ij}}$ and control units by inverse of probability of being assigned to control group $\frac{1}{1-p_{ij}}$ so that

$$w_{ij} = \frac{1}{p_{ij}}D_i + \frac{1}{1-p_{ij}}(1 - D_i)$$

Weights have to be used when using regression to estimate the ATE. Because randomisation within each block are independent, variance of blocking estimator is simply

$$Var(\hat{\tau}_{ATE}) = Var\sum_{j=1}^{J}\left(\frac{N_j}{N}\right)^2 Var(\hat{\tau}_j)$$

4 Selection on observables

Researchers must consider why some units received one treatment rather than another. Assignment mechanism is crucial for research design. Using **selection on observables** (SOO) design, researchers make assumption that assignment to treatment was based on observable characteristics of units. If researchers could observe all the covariates used to assign units to treatment, they could identify causal effects.

4.1 Identification Assumptions

To use SOO research design two assumptions need to hold.

$$(Y_1, Y_0) \perp\!\!\!\perp D|X$$
$$0 < Pr(D = 1|X) < 1$$

If assumptions are satisfied, ATE is then given by

$$\tau_{ATE} = \int E\left[Y_{i1} - Y_{i0}|X\right] dP(X)$$

Using the SOO assumptions, the following equation*s are therefore true:

$$\overbrace{E\left[Y_{i0}|X, D = 1\right]}^{\text{We don't observe this}} = \overbrace{E\left[Y_{i0}|X, D = 0\right]}^{\text{We do observe this}}$$
$$\underbrace{E\left[Y_{i1}|X, D = 0\right]}_{\text{We don't observe this}} = \underbrace{E\left[Y_{i1}|X, D = 1\right]}_{\text{We do observe this}}$$

Therefore, given SOO holds, can simplify the estimand to:

$$\tau_{ATE} = \int \left(E\left[Y|X, D = 1\right] - E\left[Y|X, D = 0\right] dP(X)\right)$$

If we only interested in an ATT/ATC can even relax our assumptions to $Y_0 \perp\!\!\!\perp D|X$ and $Pr(D = 1|X) < 1/Y_1 \perp\!\!\!\perp D|X$ and $0 < Pr(D = 1|X)$.

4.2 Estimation using SOO and Regression

Estimate τ_{OLS} using fully saturated regression. Suppose that the covariates take on a finite number of values: $x^1, x^2, ..., x^k$ then

$$Y_i = \hat{\tau}_{OLS}D_i + \sum_{k=1}^{K} Z^k\hat{\beta}_k + \epsilon_i$$

with

$$Z^k = \begin{cases} 1 & \text{if } X = x^k \\ 0 & \text{if } X \neq x^k \end{cases}$$

τ_{OLS} is conditional variance weighted average of stratum-specific causal effects, i.e. regression weights by the marginal distribution of X and the conditional variance of $Var\left[D|X\right]$ in each stratum:

$$\tau_{OLS} = \sum_{k=1}^{K} E\left[Y|D=1, X=x^k\right] - E\left[Y|D=0, X=x^k\right] \cdot w_k$$

with

$$w_k = \frac{Var\left[D|X=x^k\right] Pr(X=x^k)}{\sum_{k=1}^{K} Var\left[D|X=x\right] Pr(X=x)}$$

4.3 Estimation using SOO and Matching

To estimate τ_{ATT}, match treated units to untreated units with the same X, while match untreated units to treated units to estimate τ_{ATC}. Finally, to estimate τ_{ATE}, match in both directions. Matching estimator of τ_{ATT} can be constructed by "imputing" missing potential outcome of each treated unit using observed outcome from the "closest" untreated unit

$$\hat{\tau}_{ATT} = \frac{1}{N_1} \sum_{D_i=1} (Y_i - Y_{j(i)})$$

where $Y_{j(i)}$ is outcome of untreated observation such that $X_{j(i)}$ is closest value to X_i among untreated units. Can also use the average for M closest matches

$$\hat{\tau}_{ATT} = \frac{1}{N_1} \sum_{D_i=1} Y_i - \left(\frac{1}{M} \sum_{m=1}^{M} Y_{jm(i)}\right)$$

Works well when good matches for each treated unit available, so M is usually small (typically, $M = 1$ or $M = 2$). Can also match on more than one variable. In that case, need to define distance metric to measure "closeness". For propensity score matching, the propensity score is defined as the selection probability conditional on confounding variables

$$p(X) = P(D = 1|X)$$

5 Instrumental Variables

5.1 Binary instruments

If using IVs to get more precise τ's out of experiments, usually work with binary instrument Z_i

$$Z_i = \begin{cases} 1 \text{ if unit } i \text{ assigned to receive treatment} \\ 0 \text{ if unit } i \text{ not assigned to receive treatment} \end{cases}$$

Let D_{iz} denote potential treatment status for unit i given $Z_i = z$.

$$D_{iz} = \begin{cases} 1 \text{ if unit } i \text{ takes treatment given } Z_i = z \\ 0 \text{ if unit } i \text{ does not take treatment given } Z_i = z \end{cases}$$

Finally, let D_i denote the observed treatment, which is realised as $D_i = Z_i \cdot D_{i1} + (1 - Z_i) \cdot D_{i0}$ such that

$$D_i = \begin{cases} D_{i1} \text{ if } Z_i = 1 \\ D_{i0} \text{ if } Z_i = 0 \end{cases}$$

Given these definitions, can classify any given unit according to their potential treatment status

- **Compliers**: $D_{i1} > D_{i0}$ (i.e. $D_{i0} = 0$ and $D_{i1} = 1$).
- **Always-takers**: $D_{i1} = D_{i0} = 1$.
- **Never-takers**: $D_{i1} = D_{i0} = 0$.
- **Defiers**: $D_{i1} < D_{i0}$ (i.e. $D_{i0} = 1$ and $D_{i1} = 0$).

In the IV framework, two estimands:
The **intention-to-treat** (ITT) effect is causal effect of treatment assignment, ignoring non compliance. The **local average treatment effect** (LATE) for compliers is average causal effect for units whose treatment status is entirely determined by instrument. To estimate $LATE$s, have to make four critical assumptions

1. **Exogeneity of the Instrument**

   $$Y_i(D_i, Z_i), D_{i1}, D_{i0} \perp\!\!\!\perp Z_i$$

   This assumption yields ITT.

   $$\begin{aligned} \tau_{ITT} &= E\left[Y_i(D_{i1}, Z_i = 1) - Y_i(D_{i0}, Z_i = 0)\right] \\ &= E\left[Y_i|Z_i = 1\right] - E\left[Y_i|Z_i = 0\right] \end{aligned}$$

2. **Exclusion Restriction**
   Instrument has no direct effect on outcome, once fix value of treatment

   $$Y_i(D_i = 1, Z_i = 1) = Y_i(D_i = 1, Z_i = 0) = Y_{i1}$$

   and

   $$Y_i(D_i = 0, Z_i = 1) = Y_i(D_i = 0, Z_i = 0) = Y_{i0}$$

   Z_i only affects Y_i through D_i.

3. **First-stage relationship**
   Instrument must have an effect on treatment.

$$0 < P(Z_i = 1) < 1 \;\&\; P(D_{i1} = 1) \neq P(D_{i0} = 1)$$

4. **Monotonicity**
   Monotonicity says that presence of instrument never dissuades someone from taking treatment.

$$D_{i1} \geq D_{i0}$$

Under these four assumptions, we LATE for compliers is identified

$$\begin{aligned}\tau_{LATE} &= E\left[Y_{i1} - Y_{i0} | D_{i0} > D_{i1}\right] \\ &= \frac{\overbrace{E\left[Y_i|Z_i=1\right] - E\left[Y_i|Z_i=0\right]}^{\text{ITT}}}{\underbrace{E\left[D_i|Z_i=1\right] - E\left[D_i|Z_i=0\right]}_{\text{Proportion of compliers}}}\end{aligned}$$

Wald estimator:

$$\hat{\tau}_{LATE} = \frac{\bar{Y}_{Z_i=1} - \bar{Y}_{Z_i=0}}{\bar{D}_{Z_i=1} - \bar{D}_{Z_i=0}}$$

Two-stage least-squares (2SLS). First predict treatment status D_i

$$D_i = \alpha_1 + \beta_1 Z_i + \epsilon_{1i}$$

then use fitted values $\hat{D}_i$ to estimate average effect of D_i on Y_i for compliers

$$Y_i = \alpha_2 + \beta_2 \hat{D}_i + \epsilon_{2i}$$

$\hat{\beta}_2$ is an estimate of τ_{LATE}.

6 Regression Discontinuity

In cases where Regression Discontinuity Design (RDD) is viable, treatment (D) isn't randomly assigned, but determined by value of observed covariate X lying on either side of fixed threshold c. The estimand of RDD is average causal effect of treatment at the cutoff, type of $LATE$. The crucial identification assumption is that the potential outcomes are continuous in X around c, i.e. units with scores barely below the cutoff can be used as counterfactuals for units with scores barely above it. In sharp RD design, assignment to treatment D_i is completely determined by value of covariate X_i being on either side of threshold c:

$$D_i = \mathbb{1}(X_i > c) = \begin{cases} 1 & \text{if } X_i > c \\ 0 & \text{if } X_i < c \end{cases}$$

Local Average Treatment effect is then

$$\begin{aligned}\tau_{LATE} &= E\left[Y_1 - Y_0 | X = c\right] \\ &= E\left[Y_1|X=c\right] - E\left[Y_0|X=c\right]\end{aligned}$$

Since don't observe $E\left[Y_1|X=c\right]$ and $E\left[Y_0|X=c\right]$, rely on identification assumption that $E\left[Y_1|X=c\right]$ and $E\left[Y_0|X=c\right]$ are continuous in c, that is

$$\begin{aligned}E\left[Y_0|X=c, D=0\right] &= E\left[Y_0|X=c, D=1\right] \\ E\left[Y_1|X=c, D=0\right] &= E\left[Y_1|X=c, D=1\right]\end{aligned}$$

Thus, can estimate $LATE$ using

$$\begin{aligned}\tau_{LATE} &= E\left[Y_1 - Y_0 | X = c\right] \\ &= E\left[Y_1|X=c\right] - E\left[Y_0|X=c\right] \\ &= \lim_{x \downarrow c} E\left[Y|X=c\right] - \lim_{x \uparrow c} E\left[Y|X=c\right]\end{aligned}$$

6.1 Estimation of τ_{LATE}

In estimation of τ_{LATE}, trim sample to reasonable window around cutoff c, so only retain values for which $c - h \leq X_i \leq c + h$, where h is some positive value that determines size of window. Then generate $\tilde{X}$, which measures distance to threshold

$$\tilde{X} = X - c$$

Finally decide on a model for $E\left[Y|X\right]$. Could use linear model with same slopes above and below the cutoff, linear model with different slopes above and below the cutoff, or non-linear model. For linear model with same slopes

$$\begin{aligned}E\left[Y|X,D\right] &= D \cdot E\left[Y_1|X\right] + (1-D) \cdot E\left[Y_0|X\right] \\ &= D \cdot (\tau + \mu + \beta X) + (1-D) \cdot (\mu + \beta X) \\ &= D \cdot \tau + D \cdot \mu + D \cdot \beta X + \mu + \beta X - D \cdot \mu - D \cdot \beta X \\ &= \mu + \tau D + \beta X - \beta c + \beta c \\ &= (\mu - \beta c) + \tau D + \beta (X - c) \\ &= \gamma + \tau D + \beta \tilde{X}\end{aligned}$$

Thus we can run a regression of the following specification

$$E\left[Y|X,D\right] = \gamma + \tau D_i + \beta \tilde{X}_i$$

If alternatively decide on linear model with different slops, will interact treatment status indicator with $\tilde{X}$ to yield

$$E\left[Y|X,D\right] = \gamma + \tau D + \beta_0 \tilde{X} + \beta_1 (\tilde{X} \cdot D)$$

can also specify a model with non-linear interaction terms. To achieve this, would add higher order polynomials to our model:

$$\begin{aligned} E[Y|X,D] = \gamma + \tau D + \beta_0 \tilde{X} \\ + \beta_1 \tilde{X}^2 + ... + \beta_{k-1} \tilde{X}^k + \beta_k (\tilde{X} \cdot D) \\ + \beta_{k+1} (\tilde{X}^2 \cdot D) + ... + \beta_{2k-1} (\tilde{X}^k \cdot D) \end{aligned}$$

6.2 Bandwith selection & Valiation

A small h decreases the bias of $\hat{\tau}_{LATE}$. But small h also increases the variance of $\hat{\tau}_{LATE}$. Use the "optimal" bandwith, determined by an algorithm.

In practice, three falsification checks should be carried out. Balance checks: Are covariates discontinuous at the threshold? Sorting: Are units able to "sort" around the threshold? Compound Treatment: Are there multiple changes at the cutoff?

7 Differences-in-difference

Differences-in-Difference (DD) design allows for comparison over time in treatment group, controlling for concurrent time trends using control group. The estimand is the average treatment effect on the treated in the post-treatment period.

First, let N units be denoted by $i \in \{1, ..., N\}$. Two time periods are denoted by $t \in \{0, 1\}$. Period $t = 0$ indicates time period before treatment, while period $t = 1$ indicates time period after treatment took place. Further, two groups of units are denoted by $G_i \in 0, 1$. Units in control group ($G_i = 0$) do not receive treatment. Units in treatment group ($G_i = 1$) receive treatment, but only in post-treatment period ($t = 1$). Whether a unit received treatment in a given period is denoted by treatment indicator $Z_{it} \in \{0, 1\}$. Each unit has two potential outcomes in each period $Y_{it}(Z)$. In each period, observed outcomes Y_{it} are realised as

$$Y_{it} = Z_{it} Y_{it}(1) + (1 - Z_{it}) Y_{it}(0)$$

The estimand is defined as

$$\begin{aligned} \tau_{ATT} =& E[Y_{i1}(1) - Y_{i1}(0)|G_i = 1] \\ =& E[Y_{i1}(1)|G_i = 1] - E[Y_{i1}(0)|G_i = 1] \end{aligned}$$

Use the difference in difference estimator:

$$\begin{aligned} \hat{\tau}_{ATT} =& \overbrace{(E[Y_{i1}(1)|G_i = 1] - E[Y_{i1}(0)|G_i = 0])}^{\Delta \text{ between } G_1 \text{ and } G_0 \text{ in } t_1} \\ & - \underbrace{(E[Y_{i0}(1)|G_i = 1] - E[Y_{i0}(0)|G_i = 0])}_{\Delta \text{ between } G_1 \text{ and } G_0 \text{ in } t_0} \end{aligned}$$

This assumes parallel trends

$$E[Y_{i1}(0) - Y_{i0}(0)|G_i = 1] = E[Y_{i1}(0) - Y_{i0}(0)|G_i = 0]$$

This assumption is untestable: we cannot observe potential outcomes under the control condition for treated units in the post-treatment period.