Potenial Outcome Model

$$\tau = y_{1i} - y_{0i}$$

$$T \equiv \begin{cases} 1, \text{ if unit get treat} \\ 0, \text{ otherwise} \end{cases}, y_i \equiv \begin{cases} y_{1i}, \text{ if } T = 1 \\ y_{0i}, \text{ if } T = 0 \end{cases} \quad (1)$$

The fundamental problem of causal inference is that for each individual i, we can only observe one of the outcomes, either y_{0i} or y_{1i}, and we can not observe both y_{0i} and y_{1i} at the same time. Put in other words, we have the missing data problem.
ATE average of all potential outcomes for the treated minus the average of all potential outcomes for the untreated

$$ATE = E[y_{1i} - y_{0i}]$$

Selection Bias

$$\begin{aligned} &E[Y_i|T_i = 1] - E[Y_i|T_i = 0] = \\ &E[Y_{1i}|T_i = 1] - E[Y_{0i}|T_i = 0] = \\ &\underbrace{E[Y_{1i}|T_i = 1] - E[Y_{0i}|T_i = 1]}_{\text{ATT}} + \underbrace{E[Y_{0i}|T_i = 1] - E[Y_{0i}|T_i = 0]}_{\text{Bias}} \end{aligned}$$

average treatment effect on the treated (ATT) treatment effect for those who received treatment:

$$ATT = E[y_{1i} - y_{0i}|T = 1]$$

The Treatment Assignment Mechanism

- Pure random assignment, $y_{0i}, y_{1i} \perp T_i$ (participation in the treatment program does not depend on the outcome).

$$\widehat{ATE} = \frac{1}{N_{T=1}} \sum y_{1i} - \frac{1}{N_{T=0}} \sum y_{0i}$$

which will converge to

$$\begin{aligned} &E[Y_{1i}|T_i = 1] - E[Y_{0i}|T_i = 0] \\ &= E[Y_{1i}] - E[Y_{0i}] \\ &= E[Y_{1i} - Y_{0i}](ATE) \end{aligned}$$

- Random assignment conditional on observables: The other names for this condition are: ignorability of treatment, unconfoundedness, and selection on observables. $y_{0i}, y_{1i} \perp T_i|x_i$ (participation in the treatment program does not depend on the outcome). Here we will use the propensity model.

Regression Analysis of Experiments

$$\begin{aligned} y_i &= T_i y_{1i} + (1 - T - i)y_{0i} \\ &= y_{0i} + T_i(y_{1i} - y_{0i}) \\ &= E[Y_{0i}] + \rho T_i + y_{0i} - E[y_{0i}] \\ &= \alpha + \beta_2 T_i + u_i \end{aligned}$$

where $\alpha = E[y_{0i}]$
Evaluating the conditional expectation of this equation with treatment status switched on and off, we obtain that

$$\begin{aligned} E(Y_i|T_i = 1) &= \alpha + \rho + E[u_i|T_i = 1], \\ E(Y_i|T_i = 0) &= \alpha + E[u_i|T_i = 0] \end{aligned}$$

so that

$$E(Y_i|T_i = 1) - E(Y_i|T_i = 0) = \rho + E[u_i|T_i = 1] - E[u_i|T_i = 0]$$

The above implies that selection bias
$E[u_i|T_i = 1] - E[u_i|T_i = 0]$ equal the correlation between the regression error term ui and the regressor T_i. Furthermore, since

$$E[u_i|T_i = 1] - E[u_i|T_i = 0] = E[y_{0i}|T_i = 1] - E[y_{0i}|T_i = 0]$$

this correlation reflects the difference in (no-treatment) potential outcomes between those who get treated and those who do not.

Propensity Score Matching Method $y_{0i}, y_{1i} \perp T_i|x_i$ implies that $y_{0i}, y_{1i} \perp T_i|x_i p(x_i)$ where

$$p(x_i) = P(T_i = 1|x_i)$$

something like this

$$\begin{aligned} &P[T_i = 1|y_{0i}, y_{1i}, p(x)] = E[T_i|y_{0i}, y_{1i}, p(x)] \\ &= E\{E[T_i|y_{0i}, y_{1i}, p(x), x]|y_{0i}, y_{1i}, p(x)\} \\ &= E\{E[T_i|y_{0i}, y_{1i}, x]|y_{0i}, y_{1i}, p(x)\} \\ &= E\{E[T_i|x]|y_{0i}, y_{1i}, p(x)\} \\ &= E\{E[p(x)]|y_{0i}, y_{1i}, p(x)\} \\ &= p(x) \end{aligned}$$

Propensity Score Matching Method

$$\tau \equiv E(TE_i) = E[y_{1i} - y_{0i}]$$

Under the so-called "unconfoundedness" condition , one can estimate the ATE by first estimating the conditional treatment effects given a vector of covariates xi:

$$\begin{aligned} T(x) &\equiv E[y_{1i} - y_{0i}|x_i = x] \quad (2) \\ &= E[y_i|T_i = 1, x_i = x] - E[y_{0i}|T_i = 0, x_i = x] \quad (3) \\ &= E[y_i|T_i = 1, x_i = x] - E[y_i|T_i = 0, x_i = x], \quad (4) \end{aligned}$$

where the second equality follows from the "unconfoundedness" assumption which states that conditional on the covariates xi, the treatment assignment ti and the outcome yi are independent of each other.
Define $m_0(x_i) = E[y_i|T_i = 0, x_i]$, then by (11), it is straightforward to see that

$$\tau \equiv y_i = m_0(x_i) + T_i\tau_i + \epsilon_i$$

where $\tau_i = \tau(x_i)$, and $E(\epsilon_i|x_i, t_i) = 0$ by construction.
Under the "overlap" assumption $(0 < p(x) < 1)$, one can average the conditional treatment effects to obtain ATE:

$$\tau = E[\tau(x_i)]$$

It can be shown that an alternative expression for ATE is given by

$$\tau = E\frac{(T_i = \mu_i)y_i}{\mu_i(1 - \mu_i)}],$$

where

$$\mu_i \equiv \mu(x_i) \equiv P(T_i = 1|x_i) = E(T_i|x_i)$$

is the propensity score. To obtain a feasible ATE estimate, we need to replace μ_i by an estimator.

Difference-in-differences

$$\Theta_{DD} = (\bar{y}_{t=1}^{T=1} - \bar{y}_{t=0}^{T=1}) - (\bar{y}_{t=1}^{T=0} - \bar{y}_{t=0}^{T=0})$$

where $\bar{y}_{t=1}^{T=1} - \bar{y}_{t=0}^{T=1}$ denotes the change in outcomes (between pre-treatment and post-treatment) fro the treated group; $\bar{y}_{t=1}^{T=0} - \bar{y}_{t=0}^{T=0}$ denotes the change in outcomes for the control group; $\bar{y}_{t=1}^{T=1}$ denotes the average outcome in the post-treatment period for the treated group; $\bar{y}_{t=1}^{T=0}$ denotes the average outcome in the post-treatment period for the control group; $\bar{y}_{t=0}^{T=1}$ denotes the average outcome in the pre-treatment period for the treatment group; $\bar{y}_{t=0}^{T=0}$ denotes the average outcome in the pre-treatment period for the control group.

Regression Analogue

$$y_{it} = \alpha + \beta P_t + \gamma T_i + \Theta P_t T_i + \epsilon_{it}$$

- when $P_t = 1, T_i = 1$, $y_{i1}^{T=1} = \alpha + \beta + \gamma + \theta + \epsilon_{i1}^{T=1}$
- when $P_t = 1, T_i = 0$, $y_{i1}^{T=0} = \alpha + \beta + \epsilon_i^{T=0}$
- when $P_t = 0, T_i = 1$, $y_{01}^{T=1} = \alpha + \theta + \epsilon_{i0}^{T=1}$
- when $P_t = 0, T_i = 0$, $y_{01}^{T=1} = \alpha + \epsilon_{i1}^{T=1}$

where

- β is the change in outcome for the control group from the pre-treatment period to the post-treatment period.
- γ is the change in outcome between the treatment group and control group in the pre-treatment period.
- θ is the diff-in-diff effect.

The expected change over time for the control group, which is

$$E(y_{i1}^{T=0} - y_{i0}^{T=0}) = \beta + E(\epsilon_{i1}^{T=0} - \epsilon_{i0}^{T=0})$$

and the expected change over time for the treated group is

$$E(y_{i1}^{T=1} - y_{i0}^{T=1}) = \beta + E(\epsilon_{i1}^{T=1} - \epsilon_{i0}^{T=1})$$

Under the identifying assumption, the expected change over time for the control group provides a good counterfactual for the expected change over time for the treated group in the absence of treatment. This implies that

$$E(\epsilon_{i1}^{T=0} - \epsilon_{i0}^{T=0}) = E(\epsilon_{i1}^{T=1} - \epsilon_{i0}^{T=1})$$

which is also called common trends assumption.
Note that by law of large numbers, the diff-in-diff estimator in (1) is

$$\begin{aligned} \Theta_{DD} \xrightarrow{\text{p}} &E(y_{i1}^{T=1} - y_{i0}^{T=1}) - E(y_{i1}^{T=0} - y_{i0}^{T=0}) \\ &= \theta + E(\epsilon_{i1}^{T=1} - \epsilon_{i0}^{T=1}) - E(\epsilon_{i1}^{T=0} - \epsilon_{i0}^{T=0}) \\ &= \theta \end{aligned}$$

where the last equality is due to the common trends assumption.
Clustering SEs ADD A NOTE ON CLUSTERING SEs
Instrument Variable

$$\tau = y_{1i} - y_{0i}$$

$$S_i = \begin{cases} 1, \text{ if go to college} \\ 0, \text{ if not go to college} \end{cases}, Y_i = \begin{cases} Y_{1i}, \text{ if go to college} \\ Y_{0i}, \text{ if not go to college} \end{cases} \quad (5)$$

$$\begin{aligned} &\underbrace{E[Y_i|S_i = 1] - E[Y_i|S_i = 0]}_{\text{observed difference in earnings}} = E[Y_{1i}|S_i = 1] - E[Y_{0i}|S_i = 0] \\ &= \underbrace{E[Y_{1i}|S_i = 1] - E[Y_{0i}|S_i = 1]}_{\text{ATT}} + \underbrace{E[Y_{0i}|S_i = 1] - E[Y_{0i}|S_i = 0]}_{\text{Selection Bias}} - E[Y_{0i}|S_i = 0] \end{aligned}$$

if we simply compare the earnings of those who do not go to college and those who do not, we would get a biased estimate - going to college is NOT random assigned. The left-hand side of the above formula is also the coefficient ρ in the following regression:

$$Y_i = \alpha + \rho S_i + \eta_i$$

We suspect that $cov(s_i, \eta_i) \neq 0$

Omitted Variable Bias

$$Y_i = \alpha + \rho S_i + \eta_i$$

$$Y_i = \alpha + \rho S_i + A_i'\gamma + \epsilon_i$$

where α, ρ, γ are the population coefficients, and by construction, S_i, A_i are uncorrelated with ϵ_i. That is
$\eta_i = A_i'\gamma + \epsilon_i$
Let's refer to the first ρ as ρS (biased) and the second as ρL rearraging the first equaiton

$$\rho S = \frac{cov(Y_i, S_i)}{var(S_i)}$$

plugging in $Y_i = \alpha + \rho_L S_i + A_i'\gamma + \epsilon_i$ in $cov(Y_i, S_i)$

$$\begin{aligned} cov(Y_i, S_i) &= cov(\alpha + \rho_L S_i + A_i'\gamma + \epsilon_i, S_i) \\ &= cov(\alpha, S_i) + \rho_L cov(Si, S_i) + cov(A_i'\gamma, S_i) + cov(\epsilon_i, S_i) \\ &= \rho_L cov(S_i, S_i) + cov(A_i'\gamma, S_i) \\ &= \rho_L var(S_i) + cov(A_i'\gamma, S_i) \end{aligned}$$

now plugging these back in

$$\rho_S = \frac{\rho_L var(S_i) + cov(A_i'\gamma, S_i)}{var(S_i)} = \rho_L + \frac{cov(A_i'\gamma, S_i)}{var(S_i)}$$

where the second term on the right-hand side is the omitted variable bias! Equation connects the population coefficients ρS in and ρL in together, where ρL in has a causal interpretation, while ρS in does not.
Instrument Variable (IV) So, S_i is endogenous which leads to omitted variable bias. For Z_i to be an instrument for S_i

- Relevance: $cov(Z_i, S_i) \neq 0$, i.e there is correlation (or relevance) between Z_i and S_i
- Exclusion: $cov(Z_i, \eta_i) = 0$, Z_i is uncorrelated with the error term.

The above two conditions imply that, the instrument variable Z_i will affect the outcome variable Y_i ONLY through the endogeneous variable S_i, and Z_i cannot affect Y_i directly.
IV Estimation first assume 1 endog variable and 1 IV. Using $cov(Y_i, Z_i) = cov(\alpha + \rho S_i + A_i'\gamma + \epsilon, Z_i) = \rho cov(S_i, Z_i)$ implies causal ρ

$$\rho S = \frac{cov(Y_i, Z_i)}{cov(S_i, Z_i)} = \frac{cov(Y_i, Z_i)/var(S_i)}{cov(S_i, Z_i)/var(S_i)}$$

Using what we have learned in the first part of this course, the numerator on the right-hand side is the population coefficient of running OLS of Yi on Zi, and the denominator is the population coefficient of running OLS of Si on Zi. This leads to the following regression

$$Y_i = \alpha_1 + \beta_1 Z_i + \zeta_{1i}, \text{ (reduced form)}$$

$$S_i = \alpha_2 + \beta_2 Z_i + \zeta_{2i}, \text{ (first stage)}$$

which implies

$$\rho = \frac{\beta_1}{\beta_2}$$

This suggest that, to get a consistent estimator of the causal effect of schooling on earnings (denoted by ρ), we only need to get consistent estimators β_1 and β_2, then $\hat{\rho} = \frac{\hat{\beta_1}}{\hat{\beta_2}}$ will be consistent by Slutsky theorem. As Zi is exogeneous in both the reduced-form and first-stage equations, running OLSs on these two equations will give us unbiased and consistent estimators $\hat{\beta_1}$and $\hat{\beta_2}$, and therefore consistent$\hat{\rho} = \frac{\hat{\beta_1}}{\hat{\beta_2}}$.
Slutsky Theorem note: Slutsky's Theorem allows us to make claims about the convergence of random variables. It states that a random variable converging to some distribution X, when multiplied by a variable converging in probability on some constant a, converges in distribution to a x X. Similarly, if you add the two random variables, they converge in distribution to a plus X.

Another way: matrix form

$$X = \begin{bmatrix} 1 & S_1 \\ 1 & S_2 \\ . & . \\ . & . \\ . & . \\ 1 & S_n \end{bmatrix}, Z = \begin{bmatrix} 1 & Z_1 \\ 1 & Z_2 \\ . & . \\ . & . \\ . & . \\ 1 & Z_n \end{bmatrix}$$

pre-multiply Z' ib bith sides of $Y = X\delta + \epsilon$ where $\delta = \begin{bmatrix} \alpha \\ \rho \end{bmatrix}$ (this is the matrix form of IV)

$$Z'Y = Z'X\delta + Z'\epsilon$$

Running OLS on this

$$\hat{\delta} = (Z'X)^{-1}Z'Y$$

we can show that $\hat{\delta}$ is consistent, but not necessarily unbiased

$$\begin{aligned} \hat{\delta} &= (Z'X)^{-1}Z'Y \\ &= (Z'X)^{-1}Z'(X\delta + \epsilon) \\ &= \delta + (Z'X)^{-1}Z'\epsilon \\ &= \delta + (\frac{Z'X}{N})^{-1}\frac{Z'\epsilon}{N} \\ &\xrightarrow{\text{p}} \delta + E(Z_i X_i')^{-1}E(Z_i\epsilon_i) \\ &= \delta \end{aligned}$$

where in the second to last equality we use the law of large numbers, and in the last equality we used the fact that $cov(Z_i, \epsilon_i) = 0$. If endog vars are not equal to number of IVs, you have to use 2SLS

2SLS Suppose two instruments

$$S_i = \alpha_2 + \beta_2 Z_{1i} + \gamma_2 Z_{2i} + \zeta_{2i}$$

plugging into the above equation

$$\begin{aligned} Y_i &= \alpha + \rho[\alpha_2 + \beta_2 Z_{1i} + \gamma_2 Z_{2i} + \zeta_{2i}] + \epsilon_i \\ &= \alpha + \rho[\alpha_2 + \beta_2 Z_{1i} + \gamma_2 Z_{2i}] + [\rho\zeta_{2i} + \epsilon_i] \end{aligned}$$

where $\alpha_2 + \beta_2 Z_{1i} + \gamma_2 Z_{2i}$ is the population fitted value from the first-stage regression of S_i on Z_{1i} and Z_{2i}. Because Z_{1i} and Z_{2i} are uncorrelated with the reduced-form error $\rho\zeta_{2i} + \epsilon_i$, the coefficient on $\alpha_2 + \beta_2 Z_{1i} + \gamma_2 Z_{2i}$ in the population regression of Yi on $\alpha_2 + \beta_2 Z_{1i} + \gamma_2 Z_{2i}$ equals ρ, the causal effect of schooling on earnings. This suggests us that, we could first run the first-stage regression of Si on Z_{1i} and Z_{2i}, i.e., equation (9), get fitted value

$$\hat{S_i} = \hat{\alpha_2} + \hat{(}\beta_2)Z_{1i} + \hat{\gamma_2}Z_{2i}$$

Then run the second-stage regression:

$$Y_i = \alpha + \rho\hat{S_i} + \epsilon_i$$

The resulting estimator $\hat{\rho}$will be consistent for ρ.

Measurement Error in the Independent Variable Covariate an imprecise measure of x, and $x* = x + u$, where u is the measurement error, and $u \sim (0, \theta_u^2)$. We assume that the measurement error u is uncorrelated with the unobserved variable x (This is called classical measurement error problem).

$$y = \alpha + \beta x* + \eta$$

where $\eta = e - \beta u$, endogeneity issue because $cov(x*, \eta) = cov(x*, e-\beta u) = cov(x+u, e-\beta u) = -\beta var(u) \neq 0$

$$\begin{aligned} \hat{\beta} &= \frac{c\hat{o}v(x*,u)}{v\hat{a}r(x*)} \\ &= \frac{c\hat{o}v(x*,\alpha+\beta x*+\eta)}{v\hat{a}r(x*)} \\ &= \beta + \frac{c\hat{o}v(x*,\eta)}{v\hat{a}r(x*)} \\ &= \beta - \beta\frac{v\hat{a}r(u)}{v\hat{a}r(x*)} \\ &= \beta[1 - \frac{v\hat{a}r(u)}{v\hat{a}r(x*)+v\hat{a}r(u)}] \\ &= \beta[1 - \frac{v\hat{a}r(u)}{v\hat{a}r(x*)+v\hat{a}r(u)} \\ &= \beta\frac{v\hat{a}r(x)}{v\hat{a}r(x*)+v\hat{a}r(u)} \end{aligned}$$

$$plim\hat{\beta} = \beta\frac{v\hat{a}r(x)}{v\hat{a}r(x*)+v\hat{a}r(u)}] < \beta$$

Measurement Error in the Dependent Variable imprecise measure of y, and $y = y* + u$, where u is the measurement error, and $u \sim (0, \sigma_u^2)$.

$$y* = \alpha + \beta x + \eta$$

where $\eta = e - u$

$$\begin{aligned} \hat{\beta} &= \frac{c\hat{o}v(x,y*)}{v\hat{a}r(x)} \\ &= \frac{c\hat{o}v(x,\alpha+\beta x+\eta)}{v\hat{a}r(x)} \\ &= \beta + \frac{c\hat{o}v(x,\eta)}{v\hat{a}r(x)} \end{aligned}$$

$$plim\hat{\beta} = \beta$$

because $plim\frac{c\hat{o}v(x,\eta)}{v\hat{a}r(x)} = 0$ because $cov(x, \eta) = cov(x, e - u) = 0$

IV for measurement error in independent varibales As discussed previously, measurement error in the independent variable will bias the OLS estimator, however, the OLS estimator will remain unbiased for the case of the measurement error in the dependent variable.One solution to the first case is using the instrument variable. Suppose we have a valid instrument variable z for x*. For z to be valid, it needs to satisfy two conditions: 1 $cov(z, x) \neq 0$ and 2. $cov(z, e) = 0$. Now applying IV estimator which is consistent

$$cov(z, y) = cov(z, \alpha + betax^* + \eta) = \beta cov(z, x^*)$$

$$\hat{\beta_{IV}} = \frac{c\hat{o}v(z,y)}{c\hat{o}v(z,x)}$$

Wald Estimator For the IV estimator, when the instrument is a dummy/binary variable (i.e., the instrument only takes two values, 0 or 1), the IV estimator is also called the wald estimator. Suppose we have a binary instrument variable Zi (e.g., draft eligibility), and the IV estimator (which is given in the chapter 4 (1)) is the sample analogue of:

$$\rho = \frac{cov(Y_i, Z_i)}{cov(S_i, Z_i)} = \frac{cov(Y_i, Z_i)/var(S_i)}{cov(S_i, Z_i)/var(S_i)}$$

Note that the numerator on the right-hand side is the population coefficient of running OLS of Yi on Zi, and the denominator is the population coefficient of running OLS of Si on Zi. This leads to the following regression:

$$Y_i = \alpha_1 + \beta_1 Z_i + \zeta_{1i}, \text{ reduced form}$$

$$S_i = \alpha_2 + \beta_2 Z_i + \zeta_{2i}, \text{ first stage}$$

When Zi is a binary variable,
$$
$$

$$\begin{aligned} \beta_1 &= E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0), \\ \beta_2 &= E(S_i|Z_i = 1) - E(S_i|Z_i = 0) \end{aligned}$$

which implies

$$\rho = \frac{E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0)}{E(S_i|Z_i = 1) - E(S_i|Z_i = 0)}$$

The numerator E(Yi|Zi = 1) - E(Yi|Zi = 0) is called the intent-to-treat (ITT) effect, and the denominator is called take-up ratio. Note that under the monotonicity assumption, the denominator is always positive. The sample analogue is

$$\rho = \frac{\bar{Y_Z} = 1 - \bar{Y_Z} = 0}{\bar{S_Z} = 1 - \bar{S_Z} = 0}$$

Monotonicity Assumption: the presence of instrument does not dissuade people from taking treatment (or there is no defier), i.e, $Si|Zi = 1 - Si|Zi = 0 \geq 0$. In treatment literature, there are usually four types of people: 1. Always-takers: regardless of being offered treatment or not, always take treatment. 2. Never-takers: regardless of being offered treatment or not, never take treatment. 3. Compliers: take treatment when being offered treatment, and don't take treatment if not being offered treatment. 4. defiers: take treatment when not being offered treatment, and don't take treatment if being offered treatment. The presence of defiers usually complicates the analysis, and we assume it away here.

Regression Discontinuity Design second-best research design after the expriments. "assignment variable" (also called a "score", "running variable", or "forcing variable") exceeds a known cutoff point. 3 principles: a score, a cutoff, and a treatment.

Sharp RD Design treatment is assigned to those units whose score is above a known cutoff and not assigned to those units whose score is below the cutoff.The assignment of the treatment follows a rule that is known (to the researcher) and hence empirically verifiable. Assume that there are n units, each unit has a score or a running variable Xi, and c is a known cutoff. Units with $Xi \geq c$ are assigned the treatment, and units with $Xi < c$ are not. Let Ti denote the treatment assignment, then

$$T_i = 1(X_i \geq c),$$

where 1(.) is the indicator function. The assignment rule implies that the probability of treatment assignment as a function of the score changes discontinuously at the cutoff. Being assigned to the treatment condition, however, is not the same as receiving or complying with the treatment. If the treatment assigned and treatment received are identical (i.e., take-up rate is 1), then it is sharp design. Note that these outcomes are called potential because only one of them is observed (this is also the fundamental problem of causal inference). The observed outcome is

$$Y_i = (1 - T_i)Y_{i0} + T_i Y_{1i} = \begin{cases} Y_{i0}, \text{ if } X_i < \text{c} \\ Y_{i1}, \text{ if } X_i \geq \text{c} \end{cases} \quad (6)$$

$$E(Y_i|X_i) = \begin{cases} E(Y_{i0}|X_i) = , \text{ if } X_i < \text{c} \\ E(Y_{i1}|X_i), \text{ if } X_i \geq \text{c} \end{cases} \quad (7)$$

The Sharp RD design exhibits an extreme case of lack of common support condition, as units in the treatment group $(Ti = 1(Xi \geq c) = 1)$ and control group $(Ti = 1(Xi < c) = 0)$ cannot have the same value of running variable (Xi). However, the values of the average potential outcomes at c are not abruptly different from their values at points near c, the units with $Xi = c$ and $Xi = c - \epsilon$ would be very similar except their treatment status, and we could approximately calculate the vertical distance at c using observed outcomes. That is, we compare the outcomes of the units with scores exactly equal or barely above c to those with scores barely below c.

$$T_{SRD} = \lim_{x \to c^+} E(Y_i|X_i = x) - \lim_{x \to c^-} E(Y_i|X_i = x)$$

The above result tells us, if the average potential outcomes are continuous functions of the score at c, the difference between the limits of the treated and control average observed outcomes as the score converges the the cutoff is equal to the average treatment effect at the cutoff. In RD context, continuity means that as the score x gets closer and closer to its value at the cutoff, the average potential outcome $E(Yi0|Xi = x)$ gets closer and closer to its value at the cutoff, $E(Yi0|Xi = c)$ (analogously for $E(Yi1|Xi = x)$). Thus, continuity gives a justification for estimating the sharp RD effect by focusing on observations above and below the cutoff.

Idenifying assumptions As long as characteristics (observable and unobservable) are smooth and continuous across the cutoff, then we can get an unbiased estimate of the treatment by comparing the observed outcomes across the threshold, i.e., if there is any discontinuous jump in outcomes at the cutoff, it can be interpreted as the causal effect of admission to the treatment.

RD Notes 1 RD designs can be invalid if individuals can precisely manipulate the "assignment variable". 2 If individuals–even while having some influence-are unable to precisely manipulate the the assignment variable, a consequence of this is that the variation in treatment near the threshold is randomized as though from a randomized experiment.

Fuzzy Regression Discontinuity Design Concept In the sharp RD design, we we assume perfect compliance, but sometimes imperfect compliance or non-compliance -although the probability of receiving treatment still jumps at the cutoff c, it no longer changes from 0 to 1 as in sharp RD case. We now introduce additional notation Di, which is a binary variable equal 1 if the treatment was actually received by unit i, and 0 otherwise (sometimes $T_i \neq D_i$. The observed is at $Di = TiD1i + (1 - Ti)D0i$. Exclusion effect: ssignment T does not have a direct effect on the outcome Y , instead, it affects the outcome Y only because this assignment induces a change in the actual treatment taken D, which in turn affects Y.

1. look at the effect of treatment assignment Ti on the outcome Yi, intent-to-treat effect captures local average of being assigned to the treatment

$$\tau_{ITT} \equiv \lim_{x \to c^+} E(Y_i|X_i = x) - \lim_{x \to c^-} E(Y_i|X_i = x)$$

2. look at the effect of the treatment assignment Ti on the treatment take-up Di (first stage effct) = average effect at the cutoff of being assigned to the treatment on receiving the treatment.

$$\tau_{FS} \equiv \lim_{x \to c^+} E(D_i|X_i = x) - \lim_{x \to c^-} E(D_i|X_i = x)$$

In particular, since we are assuming that Di is binary, τ_{FS} captures the difference in the probability of actually receiving the treatment at the cutoff between units assigned to treatment and control.

3. average treatment effect at the cutoff can be recovered as

$$\tau_{FRD} = \frac{\tau_{ITT}}{\tau_{FS}}$$

which is the ratio between the average treatment effect τ_{ITT} and the average effect of the treatment assignment on the treatment take-up,τ_{FS}, both at the cutoff.

Synthetic Control Method

- Suppose that we observe J +1 units in periods $1, 2, \ldots T$.
- Without loss of generality, we assume the first unit is treated during periods $T_0 + 1, \ldots, T$.
- The remaining J units are an untreated reservoir of potential controls (called "donor pool").
- Let Y_{it}^N be the outcome that would be observed for unit i at time t in the absence of the it intervention.
- Let Y_{it}^I be the outcome that would be observed for unit i at time t if unit i is exposed to the intervention in periods T0 + 1 to T .
- We aim to estimate the effect of the intervention on the treated unit (or the first unit):

$$\alpha_{1t} \equiv T_{1t}^I - Y_{1t}^N = Y_{1t} - Y_{1t}^N$$

for t > T0, and Y1t is the outcome for unit one at time t.

- Suppose that Y_{it}^N is given by a factor model

$$Y_{it}^N = \delta_t + \Theta_t Z_i + \lambda_t \mu_i + \epsilon_{it}$$

where δ_t is an unknown common factor with constant factor loadings, Z_t is a vector of observed covariates, λ_t is a vector of unobserved factors, and μ_i is a vector of unknown factor loadings.

- Let $W = (w2, \ldots, w_{J+1})'$ with $w_j \geq 0$ for $j = 2, \ldots, J + 1$ and $w_2 + \cdots + w_{J+1} = 1$. Each value of W represents a potential synthetic control.
- Abadie et al. (2010) show that if $Y_{1t} - \sum_{j=2}^{J+1} Y_{jt}$ is close to zero if $\sum_{j=2}^{J+1} w_j * Y_{jt} = Y_{1t}, \sum_{j=2}^{J+1} w * Z_j = Z, t = 1, 2, \ldots, T$.
- Let X_1 be a $k1$ vector of pre-intervention characteristics for the treated unit. Similarly, let X_0 be a kJ matrix which contains the same variables for the untreated units.
- The vector $W^* = (w_2^*, , w_{j+1}^*)'$ is chosen to minimize $||X_1 - X_0 W||$, subject to $w_2 \geq 0, \ldots, w_{J+1} \geq 0, w2 + \cdots + w_{J+1} = 1$. Typically,

$$||X_1 - X_0 W|| = (\sum_{h=1}^{k} v_h(X_{h1} - w_2 X_h 2 - \ldots w_{J+1} X_{hJ+1})^2)^{1/2}$$

where the positive constants $v1, \ldots, vk$ reflect the predictive power of each of the k predictors on $Y_{1t}N$, and they can be chosen by the researcher or selected via out-of-sample validation.

- Let Y_{jt} be the value of the outcome for unit j at time t. For a post-intervention period $t(t \geq T_0)$, the synthetic control estimator is:

$$\hat{\tau}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}$$