Final Exam PPHA 34600

5/31/2022

I. Section 03: Keiser, David A. and Joseph S. Shapiro 2018. "Consequences of the Clean Water Act and the demand for water quality".

I.A: Describe the research question in this paper in words. Explain, in words and math, the ideal experiment one might want to use to answer this question. Explain, in words and math, the naïve estimator. Provide three concrete examples of why the naïve estimator is unlikely to provide an unbiased parameter estimate in this setting, and explain whether each would bias your estimate upwards or downwards relative to the truth.

(1) The research question

The research paper poses several research questions:

1. What are the effects of the Clean Water act in water pollution concentration?
2. **Did the Federal grants for municipal wastewater plant treatment have an effect in water pollution?**
3. What is the benefit-cost ratio of the Clean Water Act?

Given that the authors state, *"Because this article exploits the timing and location of grants to identify the effect of the Clean Water Act's grants program"* (pg. 357) and the cost-effectiveness analysis is also based on grant-allocation, I consider the 2nd research question as the main question that motivates the paper. Since $650 billion of the $1 trillion invested by the Clean Water Act was destined to municipal grants for water treatment, the authors seek to measure whether these grants had an effect in water pollution concentration.

(2) The ideal experiment

In the ideal experiment, we would run a **Randomized Controlled Trial (RCT)**, where municipalities are *randomly assigned* into receiving federal grants for wastewater plant treatments. Those municipalities that receive grants would be the treatment group, which we would compare against those municipaliies that do not receive grants (ie. the control group).

Using the **potential outcomes framework**, we can expect the potential water pollution for a grant-recipient municipality to be the same level of water pollution for a control municipality that did not receive a grant, had that grant-recipient municipality never received a grant.

If an RCT is properly randomized (which can be checked with a balance table pre-treatment), we effectively solve the *selection bias problem*, since treated and control municipalities would be statistically similar in observable and unobservable characteristics. This means that the following **identifying assumption** for an RCT holds:

$$E[\varepsilon_i | D_i] = 0$$

Where:

- D refers to treatment status: $D = 1$ for grant-recipient municipalities and $D = 0$ for municipalities that do not receive grants
- i is each individual municipality

This means that the expectation of the error term, conditional on receiving a grant, is zero. In other words, if we find a treatment effect, we can attribute it solely to the grant for water treatment and not to other unobservable characteristics that might be correlated with the grant allocation and water pollution.

With this *ideal RCT*, we can estimate the **Average Treatment Effect (ATE)** of the *average municipality grant* on water pollution:

$$\tau^{ATE} = E[Y_i(1)] - E[Y_i(0)]$$

Where:

- Y is the outcome variable of interest, water pollution
- i is each individual municipality.

Under randomization, the **expected** water pollution for municipalities that are recipients of grants (ie. $Y_i(1)$) is equivalent to the **mean** water pollution for grant-receiving municipalities, and likewise for municipalities in the control group. We would then estimate ATE by taking the difference in means between both groups:

$$\tau^{ATE} = \overline{Y_i}(D = 1) - \overline{Y_i}(D = 0)$$

Where:

- D refers to treatment status: $D = 1$ for grant-recipient municipalities and $D = 0$ for municipalities that do not receive grants
- Y is water pollution, measured by the quantity of pollutants in the water
- i is each individual municipality

In the ideal experiment, there would be **perfect compliance** such that only municipalities randomly assigned to receive a grant actually receive it and vice versa. There would also be **no spillover effects** such that only the waters in municipalities that receive grants benefit from the wastewater treatment plants built with those grants and waters in control municipalities do not have access to those treatment plants.

(3) The naive estimator

In the absence of randomization, the naive estimator consists of a simple difference in means between **observed outcomes** instead of **potential outcomes**:

$$\tau^N = \overline{Y}(1) - \overline{Y}(0)$$

Where:

- Y is the outcome variable, ie. water pollution

This naive estimator is very likely to be **biased** because of selection into treatment. Three examples of why it might be problematic:

1. It could be the case there is a corruption issue within municipalities that receive the grants. For example, if municipality politicians redirect the funds designated for water treatment plants for personal gain or other projects in their political interest, then we could have a **downward bias** assuming that these grants don't work in reducing water pollution when in fact, the funds where never destined to address the issue.

2. Some municipalities may have a concentration of water-polluting factories while other municipalities do not. If coincidentally, those municipalities with *very limited* water-polluting factories are the ones that receive the grants (due to environmentally-focused politicians that lobby for the grants or respond to climate-conscious citizens that hold them accountable), then we would have an **upward bias** of the effect of the grants, incorrectly concluding that they are more effective than they actually are because we are observing municipalities whose waters were unpolluted to begin with.

3. There could be governance problem or a delay in between grants being allocated and grants actually being received by the municipalities. If it takes too long to receive the grants or they money comes in fractions over time, it would take longer than 2-10 years to construct water treatment plants. Then, we would have a **downward bias** where we conclude the grants are not effective in reducing water pollution when in reality, those grants are not received entirely, take too long or the operation costs of transferring them reduce the quantity received.

I.B: Copy down the authors' main regression specification (and be sure to list which equation number this is in the paper). Explain, in words and math, what treatment parameter the authors are recovering with their main specification. Does this approach recover the population average treatment effect? If yes, why? If no, why not? What assumptions are required for this regression to recover the causal effect of interest? Do you think these assumptions are likely to be satisfied in this context? Why or why not? Include references to evidence presented in the paper to support your conclusion.

For the second research question, which I consider to be the main question of the paper, ie. *did the Clean Water Act grants for municipal wastewater plant treatment have an effect in water pollution?*, the authors employ a **Triple difference-in-differences DDD** design with the following regression (page 367, regression #3):

$$Q_{pdy} = \gamma G_{py} d_d + X'_{pdy}\beta + \eta_{pd} + \eta_{py} + \eta_{dwy} + \epsilon_{pdy}$$

Where:

- Q_{pdy} is water pollution, which varies by treatment plan, year and whether the water is upstream or downstream
- G_{py} is the cumulative number of grants than each plant received by year
- γ is the **main coefficient of interest**, and represents the mean effect of each cumulative grant on downstream water pollution
- X'_{pdy} are covariates that are controlled for, such as temperature and precipitation
- η_{pd} are **plant x downstream** fixed effects
- η_{py} are **plant x year** fixed effects
- η_{dwy} are **plant x downstream x year** fixed effects
- p is each individual treatment plant
- y is each year
- d is downstream

With this regression specification, the authors include time (year y) and unit (treatment plant p) fixed effects as part of the standard DD estimator, and add a third difference where they additionally compare between downstream (d) and upstream areas. In this case, upstream areas *"act as a counterfactual for downstream areas"* (pg. 353). The focus is on areas 25 miles downstream of a treatment plant because downstreamis the direction that pollution travels in a water source.

We can also rewrite this regression as follows:

$$\tau^{DDD} = \beta_0 + \beta_1 CumGrant_{pi} + \beta_2 Post_y + \beta_3 Downstream_{pj} + \eta_1(CumGrant_{pi} \times Post_y) +$$
$$\eta_2(Post_y \times Downstream_{pj}) + \eta_3(CumGrant_{pi} \times Downstream_{pj}) +$$
$$\tau(CumGrant_{pi} \times Post_y \times Downstream_{pj}) + \varepsilon_{ijy}$$

Where:

- $CumGrant$ refers to the cumulative number of grants per year by plant, because the authors measure grants as investment in durable capital (instead of the dollar value of each grant)
- $_{pi}$ is plant i and $_{pj}$ is plant j
- $_y$ is time measured in years

- τ is the triple DDD estimate *(Equation was adapted from slide 13 of Lecture 13 and the regression in the paper.)*

With their main specification, the authors are recovering the **Average Treatment Effect** of the average cumulative grant on downstream water pollution:

$$\tau^{DDD} = E[Y_{pyd}(1)] - E[Y_{pyd}(0)]$$

Where:

- Y is downstream water pollution
- p is each individual plant
- y is each year
- d is 1 if dowwnstream, 0 otherwise

The DDD estimator effectively recovers the population ATE if the following identifying assumption is met:

$\rightarrow$ Conditional on changes over time, within treatment plants and within downstream areas, grant-recipient and non grant-recipient plants would have been trending similarly before, PRE, grants being allocated. This is the **similar difference in trends over time** identifying assumption, which is a weaker assumption than the parallel trends assumption for the standard DD estimator because a more narrow comparison is being made.

One reason why this identifying assumption may not hold is due to **coincident treatments**, which is mentioned by the authors. They refer to air pollution regulations that happened at the same time as the Clean Water Act, where *"less acid rain may have led to less acidic rivers and lakes".* (pg 352) Another coincident phenomenon is climate change, which is inversely correlated with a oxygen saturation deficits.

While we can never truly observe the counterfactual, that is, how grant-recipient plants would have trended had they not received grants, the researchers include **event study graphs** in evidence that grant-recipient and non grant-recipient plants were trending similarly before grants were allocated. In equation (4) on page 369, they specify the following event-study specification:

$$Q_{pdy} = \sum_{\tau=-10}^{\tau=25} \gamma_\tau 1[G_{p,y-\tau} = 1]d_d + X'_{pdy}\beta + \eta_{pd} + \eta_{py} + \eta_{dwy} + \epsilon_{pdy}$$

Where:

- τ refers to the event study periods, from -10 years before grants were allocated, to 25 years since the grants were allocated
- $1[G_{p,y-\tau} = 1]$ is a dummy that equals 1 for each of the years before/since the grants were allocated and 0 otherwise
- All other variables are the same as described above

Based on the corresponding event-study graph on page 377, we can observe that in the years before the grants were allocated, water pollution (measured as dissolved oxygen deficit and share of not fishable waters) in downstream and upstream areas were on similar trends. According to the authors, *"In years before a grant, the coefficients are statistically indistinguishable from zero, have modest magnitude, and have no clear trend."*(pg. 377) This suggests that statistically, plants in downstream and upstream areas were on similar trends in the years **before** grants were allocated, which provides evidence in favor that the identifying assumption holds and that the authors are recovering the ATE from their triple DDD estimator. We cannot definitively proof that the identifying assumption but the provided event-study graph supports that the assumption holds.

Additionally, to recover a precise ATE, it is necessary for the research design to have **clustered standard errors** before claiming statistical significance. The authors pose they cluster standard errors by watershed (which are water treatment plants that receive NPDES permits, pg 356). That is, SE's are clustered at the

unit level, which in this case are the plants and since the analysis is at the national level, we can expect there to be more than 45 clusters. In all, the evidence in the paper supports the identifying assumption.

I.C: Describe the main results of the paper. Include a discussion of (at a minimum) one table and one figure, in which you interpret the estimated coefficients and describe their magnitudes. What is the main policy take-away of the paper?

Based upon the three different research questions, these are the main results of the paper:

1. The authors find a substantial decline in water pollution concentration from 1962-2001 but, this decline was smaller over time (marginal returns to scale). From 1972-2001, the share of waters safe for fishing increased by *12 percentage points*. However, the authors also note that there has been a decrease in air pollution over the same time period (ie. **coincident treatments**), which explains less acidic pH levels in lakes and rivers (and therefore the decline in water pollution may not be solely attributed to the the Clean Water Act). (pgs 349, 352)

2. The Clean Water Act's grants destined for municipal wastewater treatment plans caused part of the decline in water pollution concentration. The authors employ a *triple difference-in-differences DDD design* comparing water pollution 1) pre and post grant allocation, 2) upstream vs downstream of wastewater plants, and 3) across plants. The authors find that on average, each grant reduces the probability that a downstream area violates the regulation's standards by a *0.5 percentage point*. This effect is found within 25 miles downstream of treatment plants and are estimated to last for 30 years. The authors also find that municipal had near complete pass-through (that is, it wasn't misplaced). $1 of a federal grant led to $1 of spending on municipal sewers. (pgs 349, 352)

Table II. interpretation, pg. 378

- Holding all else constant, on average, each cumulative grant is estimated to:
- Decrease *dissolved oxygen deficit* by -0.68 percentage points, statistically significant at all conventional levels of significance.
- Decrease the *share of not fishable waters* by -0.7 percentage points, statistically significant at all conventional levels of significance
- Decrease in the *share of not swimmable waters* by -0.4 percentage points, statistically significant at the 10% level.
- Decrease in *biochemical oxygen demand (BOD)* by -0.10 milligrams of oxygen per liter, statistically significant at the 5% level.
- Decrease of -204.06 *fecal coliforms* per 100 milliliters of water, statistically significant at the 5% level.
- Decrease of -0.5 by milligrams of *total suspended solids TSS* per liter of water (mg/L), statistically insignificant.
- In general, the DDD estimates show a decrease in the different types of pollutants and share of non-swimmable and non-fishable waters, statistically significant for 6 of the 7 measures. The magnitude of the coefficients are argued by the researchers to be sizable in terms of water pollution decline, especially considering that $650 billion were destined to municipal grants. If, each average cumulative grant causes a decrease of -0.0.68 pp in dissolved oxygen deficit, this is a quite large effect given the amount of investment destined to grants. Additionally, all 7 coefficients have a negative sign, as one would expect, showing each average cumulative grant has a negative relationship with water pollution.

3. The estimated effect of the Clean Water Act's grants on housing values was less than the cost of the grants themselves. To estimate the benefits, the authors ask residents within a 25-mile radius of affected rivers, how much they valued these grants. This is essentially estimating the demand curve to understand the value or utility derived from residents from these grants. The authors find that on average, the benefit over home values was 25% the cost of the grant. These results are mostly homogenous across areas and types of grants. The researches find that the average grant was $31 million and the estimated effect on housing value was $7 million, which represents a benefit of 22.5% of the grant's cost. (pgs 349, 352)

Graph IV. interpretation, pg. 385

- The graph shows the event-study results, both for homes very close to the treatment plants (within 0.25 miles) and for homes further away (25 miles of the downstream river).
- On panel A, for **homes close to treatment plants**, we can observe that in the years before grants were allocated, home values were in a down-sloping trend. After grant allocation, there is an immediate steep increase in home values, however, the trend in home values varies up and down over the following years. This panel barely provides evidence in support of grant allocation increasing house values for households near treatment plants.
- On panel B, for **homes further from treatment plants**, there is no clear indication that grants are associated with an increase in house values. There is up and down variation that could be due to many reasons other than the grants.
- It looks as though residents near the plants have a slight benefit, as measured by a modest increase in home values, but this could well be the result of other factors. The authors mention that treatment plant construction as a result of the grants, could also lead to increased taxes, which would reduce home values (pg. 389). They also pose that it's possible that the price function may have shifted, leading to willingness to pay (WTP) estimates being underestimated. (pg. 389) While there are different explanations for the directions of possible correlations, the event study graphs do not support an effect in home values.

Main policy take-away

This paper provides evidence that the Clean Water Act did lead to a decline of water pollution concentration. However, the benefit-cost ratio is small because of the huge costs (estimated at $1 trillion invested in 2014 dollars) and benefits that do not seem as large as are expected.

As a main policy take-away, the Clean Water Act may not be the policy instrument best-suited nor the most cost-effective to decrease water pollution at the desired levels for several reasons:

1. The Clean Water Act is effectively a type of **subsidy**, by allocating municipal grants and subsidizing pollution control equipment. Economic theory shows that market-incentives are often more cost-effective. As an example, the authors compare the cap-and-trade system behind the Clean Air Act.(pg. 391)

2. The scope of the policy was limited to only certain sources of pollution. It only addressed industrial pollution or "point sources" and excluded urban and agricultural pollution, or "nonpoint sources." (pg. 356) Agriculture is a major water pollution source, according to the authors. In contrast, the Clear Air Act included all major sources of pollution.

3. Most of the benefit from unpolluted water is recreational trips and aesthetics, which people can substitute easily by traveling to clean rivers instead of polluted rivers. (pg. 391)

4. In terms of health, people drink treated and filtered water, even if it was previously polluted, so they are less likely to view polluted water sources as an issue. This is different with breathing polluted air which has direct health implications. (pg. 392)

Half of these issues are specific to the way the Clean Water Act was design (subsidy design and the scope of pollution sources) and the other half are related to the inherent differences between water and air that change how people value these unpolluted resources. As the most important **take-away**, the authors note that *"allowing the stringency of pollution regulation to vary over space has potential to increase social welfare".* This means that the net benefits of a policy like the Clean Water Act vary by location, population density and how popular water recreation is, so this policy should be adapted to those areas in order to maximize net benefits. (pg. 392)

II. A local activist group, the Universal Committee to Heighten Interest in Cars Against Greenhouse Outlays (UCHICAGO) promotes the adoption of electric vehicles in Chicago. They would like you to help them design a pilot program to demonstrate the impacts of a new electric vehicle (EV) subsidy they are trying to roll out at scale.

II.A: UCHICAGO hypothesizes that providing homeowners with subsidies for electric vehicles will increase miles driven in EVs. Using the potential outcomes framework, describe the impact of treatment - a 20% EV subsidy – on electric vehicle adoption, household electricity use, and electric vehicle driving at the household level. EV adoption is a binary variable; household electricity use is measured in kWh, and electric vehicle driving is measured in eVMT. Explain to UCHICAGO what the ideal experiment would be for answering this question. Describe the dataset you'd like to have to carry out this ideal experiment and use math, words, and the potential outcomes framework to explain what you would estimate and how you would do so. Make sure to be clear about the unit of analysis (ie, what is i here?)

Using the **potential outcomes framework**, we are interested in measuring the change in potential outcomes (ie. electric vehicle adoption, household electricity use, and electric vehicle driving) caused by a 20% EV subsidy (ie. treatment). Here, potential outcomes refer to *all possible outcomes* we could have observed with the subsidy and in the absence of the subsidy.

Hypothetically, we could measure this as:

$$\tau_{householdi} = Y_{householdi}(D=1) - Y_{householdi}(D=0)$$

$$\tau_{householdi} = Y_{householdi}(1) - Y_{householdi}(0)$$

Where:

- $D_i = 1$ for households that receive the subsidy and $D_i = 0$ for households that do not *$Y_i(D_i)$ are the different *potential* outcomes of interest (ie. electric vehicle adoption, household electricity use, and electric vehicle driving) for each household i with treatment status D_i
- $\tau_{householdi}$ is the treatment effect for an individual household i.

This treatment effect for each **individual household is impossible to estimate** because we would need to observe two potential outcomes for each individual household under both states of the world: as a recipient of the EV subsidy and as a non-recipient of the EV subsidy *at the same time.* This is the **fundamental problem of causal inference** because we cannot ever observe both $Y_{householdi}(1)$ and $Y_{householdi}(0)$ simultaneously for the same household. This means, we cannot actually estimate the treatment effect of the subsidy at the *individual household level.*

Ideal experiment

In the ideal experiment, we would randomize homeowners who are recipients of the 20% EV subsidy (ie. treatment group) and homeowners that do not receive the subsidy (ie. control group). There would be **perfect compliance** in the ideal experiments, such that only those homeowners assigned to receive the treatment actually receive it, and viceversa for the control group.

We could answer UCHICAGO's question by instead estimating the effect of the **average 20% EV subsidy** on each of the potential outcomes, by estimating the **Average Treatment Effect (ATE)**. Since there are three outcomes of interest, we would calculate three separate ATE's of the following form:

$$\tau^{ATE} = E[Y_i(1)] - E[Y_i(0)]$$

Where:

- Y is the outcome variable of interest, 1) electric vehicle adoption as a binary variable, 2) household electricity use measured in kWh, and 3) electric vehicle driving measured in eVMT
- i is each individual homeowner

Under randomization, the **expected** potential outcome for homeowners that receive the EV subsidy (ie. $Y_i(1)$) is equal to the **mean** outcome of subsidy-recipient homeowner, and likewise for homeowners in the control group. Then, we can estimate the three different ATE's by taking the difference in means between both groups:

$$\tau^{ATE} = \overline{Y_i}(D = 1) - \overline{Y_i}(D = 0)$$

Where:

- D refers to treatment status: $D = 1$ for homeowners that receive the subsidy and $D = 0$ for those who don't
- Y is each of the 3 outcome variables, previously described
- i is each individual homeowner

The dataset I would use to estimate the ATE through an RCT, would include a **cross-sectional data** of multiple homeowners and their baseline characteristics, which would allow me to check the randomization process was effective through a balance table. That is, the dataset should include a number of observable characteristics that could potentially be correlated with the different outcomes, such as the homeowner's wealth, whether they already own an EV, average driving, etc. I would also need the dataset to include measures of three outcome variables of interest: binary variable on EV adoption, household electricity use in kWh and electric vehicle driving in eVMT.

II.B: UCHICAGO have secured funding to run a randomized trial to test their hypothesis. However, their funder is worried about implementing an individually-randomized design. In particular, they are concerned that control group individuals may be less likely to purchase an EV if all of the cars on the market are bought by the treatment group. Is this a problem for an RCT that is randomized at the individual level? If yes, explain why, and describe what this would do to your treatment effects relative to the truth. If no, explain why not. Assume for the remainder of Question II that everyone offered a subsidy by UCHICAGO purchases an EV.

If effectively, control group individuals are less likely to buy an EV because *all of the cars on the market are bought by the treatment group*, this is would be a problem. This is because our treatment is an EV subsidy with the goal of measuring how that subsidy impacts electric vehicle adoption, household electricity use, and electric vehicle driving. This means that we want to compare the behavior towards EV adoption and driving between individuals that received and that did not receive the subsidy.

If all cars on the market are bought by the treatment group, we have inadvertently created a **spillover**, an artificial situation where the control group cannot act as they normally would have. In the absence of this situation, some individuals in the control group may opt to buy an EV at full price (without receiving the subsidy). However, since there are no EV's available for them to buy, they will be unable to acquire one. We would therefore be recover an **upwards biased ATE** because we would observe that the whole control group did not acquire an EV, not because they did not want to but because there were no vehicles available as a by-product of the research design.

This is a violation of the **Stable Unit Treatment Value Assumption (SUTVA)**:

$$D_i \neq D_i' \rightarrow Y_i(D) \neq Y_i(D')$$

Where treatment status of individuals receiving the 20% EV subsidy (D_i), affects potential EV adoption for individuals that did not receive the subsidy ($Y_i(D')$). This is the case of **spillovers**. Since the EV subsidy impacts the availability of EV's for non subsidy recipients:

$$(D_i = 0) \downarrow$$

When we estimate ATE by taking the difference in means, we are **overestimating the effect of the subsidy**:

$$\tau^{ATE} = \overline{Y}(1) - \overline{Y}(0) \downarrow$$
$$\tau^{ATE} \uparrow$$

Where:

*Y is one of three outcomes variables (EV adoption, household electricity and EV driving)

As such, we would incrrectly conclude that the 20% EV subsidy is more effective than it actually is because *only subsidy-recipient individuals bought an EV.* Randomization makes treated and control groups statistically similar on observable and unobservable charactersitics so that the only difference in outcomes between them can be attributed to the treatment (ie. the EV subsidy). However, this problem is not solved by randomization because we are creating a situation that restricts the control group's choices in ways that would not be restricted in the absence of the experiment.

If this design were to be implemented no matter what, then we could treat the estimated ATE as a upper bound (ie. we know it is overestimated so it can never be higher than our estimate). Alternatively, the RCT could be designed to avoid this spillover effect or to actually measure it (if it were in UCHICAGO's interest).

II.C: The founder is adamant that they will not support an individually-randomized design. They would like UCHICAGO to instead provide subsidies at the neighborhood level. Will this address the concern they raised in IIA? If yes, explain how. If no, explain why this will be no better than the individually randomized design.

If instead, subsidies are provided at the neighborhood level, this would indeed solve the issue raised in (II.B) as **SUTVA will not be violated**. We would no longer have unintended **spillover effects** from subsidy-recipient individuals buying all available EV's from the stores and leaving none to buy for the control group because the whole neighborhood gets a subsidy and we assume they all buy an EV.

We are effectively randomizing entire neighborhoods into receiving the subsidy and entire neighborhoods to not receive the subsidy (control group), which gets rid of the spillover effects. Control neighborhoods who are not offered the subsidy are free to act as they normally would, so we can expect some individuals to buy EV's and others who won't, according to their WTP for an EV at full-price. Since we are now assuming that *"everyone offered a subsidy by UCHICAGO purchases an EV"*, we would be comparing the impact of the subsidy on the three outcomes in neighborhoods that are randomized into receiving the subsidy vs. those randomized not to receive it.

II.D: After listening to the founder's feedback, UCHICAGO has decided that they are actually interested in studying how providing EV subsidies to some people in a neighborhood impacts EV adoption, home electricity use, and EV driving for people who don't get the subsidy. The founder is excited about this too, and is willing to devote substantial funding to the project. Given a large budget, describe an RCT design that will allow you to measure treatment effects on both households who get a subsidy from UCHICAGO and households who don't. Make sure to describe any necessary steps, and clearly lay out any treatment arms. A tree diagram may be helpful. Use words and math to explain what treatment parameters you can estimate with this design. Be sure to be clear about the comparisons you are making. Write down a regression equation that you would use to estimate these treatment parameters, and describe how you would interpret any coefficients you recover. Finally, describe how you would use your estimates to recommend to UCHICAGO whether they should scale their subsidy program or not.

To measure treatment effects on both households who get a subsidy and households who do not (control group), I would propose running the following **Stratified Design**:

For the previous question (II. C), I would have conducted a cluster-randomized RCT, where the subsidy was randomized at the neighborhood level in order to correct for spillovers. However, since UCHICAGO is specifically interested in "how providing EV subsidies to *some people in a neighborhood* impacts EV adoption, home electricity use, and EV driving for *people who don't get the subsidy"*, this requires comparing the outcomes of treated and untreated household in the *same neighborhoods.*

As such, I propose a stratified design that **randomizes the subsidy within neighborhoods**. The steps included in the diagram are the following:
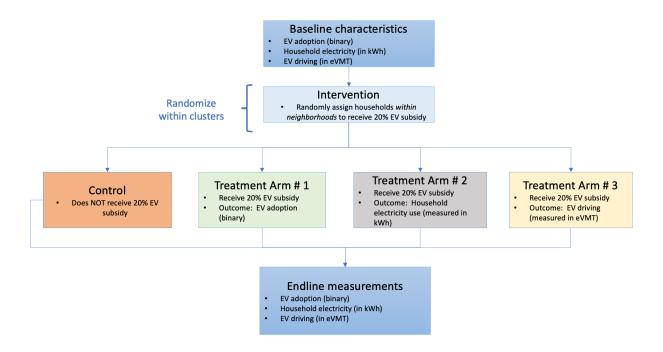
Figure 1: Stratified RCT Design, Tree Diagram

1. Collect household-level data on baseline characteristics for each of the three desired outcomes: EV adoption, household electricity use and EV driving PRE subsidies being allocated.
2. Randomize *within clusters* (ie. neighborhoods), such that some households will receive the subsidy (treatment group) and others will not (control group). Here, a balance table should be done by running t-tests to make sure that the clusteres are statistically balanced. As per UCHICAGO's specifications, the treatment itself does not change (ie. 20% EV subsidy) but there are different treatment arms for each of three outcomes of interest:

- EV adoption
- Household electricity usage
- EV driving

4. Collect end-line POST measurements at the household-level and conduct statistical analysis.

This design ensures balance between the neighborhoods and increases statistical power. With this design, we can recover the **ATE** of the average household that received the subsidy on EV adoption, household electricity use and EV driving:

$$\tau^{ATE} = E[Y_{ni}(1)] - E[Y_{ni}(0)]$$

Where:

- Y is each of the three outcome variables of interest, 1) EV adoption, 2) household electricity use, and 3) EV driving
- i is each individual household
- n is each neighborhood or strata

Under stratified randomization, in **expectation** the potential outcomes for households that receive the subsidy is equivalent to the average outcomes for households that received it. This expectation holds the other way around for households that do not receive the subsidy. Then, to estimate ATE, we need only take the difference in means between both groups:

$$\tau^{ATE} = \overline{Y_{ni}}(D=1) - \overline{Y_{ni}}(D=0)$$

Where:

- D refers to treatment status: $D=1$ for subsidy-recipient households within a neighborhood $D=0$ for households that do not receive subsidies within a given neighborhood.
- Y is the outcome variable of interest: EV adoption, household electricity, use and EV driving
- i is each individual household
- n is each neighborhood or strata

In order for the recovered ATE to be **unbiased**, we require as an identifying assumption that the potential outcome of households within a neighborhood strata are unaffected by other households within the same neighborhood, such that there is no **cross-neighborhood/cluster interference**.

Another way to estimate this difference-in-means to recover ATE, is to run a regression:

$$Y_i = \alpha + \tau Subsidy_i + (\gamma Subsidy_i \times X_n) + \delta X_i + \varepsilon_i$$

Where:

- Y_i is the outcome variable of interest: EV adoption, household electricity, use and EV driving
- τ is the ATE of the subsidy on the average household i
- $\gamma Subsidy \times X_n$ is an interaction term that accounts for heterogenous effects by neighborhood strata. X covariates that may be correlated with the outcome variables, vary by neighborhood n, whereas the treatment status into receiving subsidy varies by household i
- X_i refer to covariates that vary by household i
- i is each individual household
- n is each neighborhood/strata

If the stratified design is randomized, it is not necessary to include X_i covariates in the regression because we will still recover the ATE. Doing so will soak up noise and improve the ATE estimates.

I would interpret the ATE coefficient as follows: the average difference in household electricity use/EV driving between households that receive the 20% EV subsidy and households that do not, holding everything else constant, is predicted to be an (x coefficient value) increase/decrease (depending on a positive/negative sign) in kWh of household electricity consumed/ electrical driving eVMT. In the case of EV adoption, since the dependent and independent variables are dummies, it would be interpreted as the x% probability that an EV is adopted. I would note whether the coefficients are statistically significant, the sign of the coefficient and whether it is economically significant.

Assuming the identifying assumption holds and that the recovered ATE estimate is unbiased, I would recommend to UCHICAGO to scale their subsidy program if we obtain statistically significant estimates of a sizable magnitude for most (if not each) of the three outcome variables: EV adoption, household electricity use and EV driving. To determine whether the magnitude of the ATE estimates is significant or not, I would have to discuss with UCHICAGO experts based on their past experiences of how much is economically significant, or alternatively, research this in the literature.

III: Around the world, women are under-represented in politics. A bunch of people have hypothesized that having more women in positions of political power could improve well-being. A policy advocacy group, Powerful, Respected, and Organized Government Requires that All Men Elect Valuable Accomplished Ladies (PROGRAMEVAL), is interested in demonstrating the impact of women in government on local public goods provision in India. They've asked you for help with their analysis.

III.A: PROGRAMEVAL would like you to compare the average number of public goods (roads, schools, public buildings, etcetera) in towns with female-headed governments as compared with towns that have male-headed governments. Describe this comparison in math and words. Under what conditions would this comparison estimate the causal effect of female leaders on public goods provision? Provide two concrete examples of reasons why this comparison may be problematic.

If we simply compare the average number of public goods in towns with female-headed governments vs. male-headed governments, we would be obtaining the **naive estimator**:

$$\tau^N = \overline{Y}(1) - \overline{Y}(0)$$

Where:

- $Y(1)$ is the average number of public goods for *female-headed* governments
- $Y(0)$ is the average number of public goods for *male-headed* governments

This naive estimator is based soley on **observed public goods**, instead of the **potential public goods** for female-headed governments had they been male-headed governments instead. It is very, very unlikely that this comparison would ever estimate the effect of female leaders on the provision of public goods. For this to be the case, we would have to completely **rule out any selection bias on both observable and unobservable characteristics**, meaning that all differences between female-led and male-led governments are "as good as random".

In order to estimate the naive estimator, we could run a regression of the following form:

$$Y_i = \alpha + \tau D_i + \varepsilon_i$$

Where:

- D_i refers to treatment status $Di = 1$ for female-led towns and $D_i = 0$ for male-headed towns
- Y_i is the outcome variable, average number of public goods provided (roads, schools, public buildings, etc.)
- i is each individual town

This assumption is *highly unlikely* to hold since we can expected female-headed governments to be systematically different from male-headed governments in many ways other than in public good provision. For example, female-headed governments are likely to be more liberal, more developed and with more educated citizens, on average.

Two examples why the naive estimator would be problematic:

1. If female-headed governments also happen to be governments with better governance and a better public goods provision, then the naive estimator would lead us astray with a **upwards bias** over-estimation that female-led governments are better at providing public goods than they actually are, because we are only analyzing systematically more efficient governments.

2. Alternatively, if male-headed governments happen to be governments where public goods provision is much cheaper, like certain lower-income developing countries, then we might have a **downward bias** were we underestimate the true effect of female-led governments on public goods provision because we are comparing them to governments where the increased provision of public goods is a result of their low-cost and not because of governments being male-led.

III.B: PROGRAMEVAL gets it - this is not the best comparison. However, they have data on a bunch of other town characteristics: per-capita income, number of residents, year of incorporation, average population age, and share of gross city product devoted to manufacturing. Describe, using math and words, a comparison between female- and male-headed towns which leverages these administrative data. Under what conditions would this comparison estimate the causal effect of female leadership on public goods provision? Provide two concrete examples of reasons why this comparison may be problematic (different from what you described above).

Here we have a **cross-sectional** dataset for multiple towns in the same time period, that includes several observable characteristics.

Assuming we can only rely on this data and that an RCT is not possible, we can estimate ATE with a **Regression Adjustment as part of a Selection on Observables (SOO)** design where we essentially add controls.

To leverage this administrative data, we adjust the regular ATE regression by including controls in the form of additional covariates X_i's.

$$Y_i = \alpha + \tau D_i + \gamma X_i + v_i$$

Where:

- D_i refers to treatment status $Di = 1$ for female-led towns and $D_i = 0$ for male-headed towns
- X_i represents a series of control variables available to us in the administrative data, including:
    - per-capita income
    - number of residents
    - year of incorporation
    - average population age
    - share of gross city product devoted to manufacturing
- Y_i is the average number of public goods provided (roads, schools, public buildings, etc.)
- v_i is the error term, which captures anything else correlated with public good provision that is not being controlled for in the regression
- i is each individual town

Since we don't have random assignment into female-led governments, the error term conditional on treatment status is non-zero, ie. $E[\varepsilon \mid D_i] \neq 0$. Therefore, in order to recover an unbiased ATE from this adjusted regression, we need the following **four identifying assumptions** to hold:

1. **Conditional independence:** conditional on the controlled observables (eg.per-capita income, average population age, etc.), treatment assignment into a female-headed government is independent of potential public goods provision. in other words, all differences between $D_{town} = 1$ and $D_{town} = 0$ are assumed to be *as good as random.* This is a **very strong assumption** and can be mathematically expressed as:

$$(Y_i(1), Y_i(0)) \perp D_i \mid X_i$$

Where:

- D_i refers to treatment status $Di = 1$ for female-led towns and $D_i = 0$ for male-headed towns
- Y_i is the average number of public goods provided
- X_i represents control variables, such as: per-capita income, number of residents, year of incorporation, etc.
- i is each individual town

This assumption indicates that once we control for these covariates X_i's (eg.per-capita income, average population age, etc.), selection into a female-led town D_i is *as good as random* and we have essentially **eliminated selection bias.**

2. **Common support -** The probability that $D_i = 1$ for all levels of X_i is between 0 and 1:

$$0 < Pr(D_i = 1 \mid X_i = x) < 1$$

Where:

- D_i refers to treatment status $Di = 1$ for female-led towns and $D_i = 0$ for male-headed towns
- X_i represents control variables, such as: per-capita income, number of residents, year of incorporation, etc.
- i is each individual town

This means that for each covariate, there must be *both treated and untreated units* in the administrative (ie. female and male-headed towns, respectively.).

3. **Close mean covariates between treated and untreated -** The mean covariates for female-headed towns must be **close** to the mean covariates for male-headed towns, ie: $\overline{X}_T$ close to $\overline{X}_U$. If the difference between the covariates is large, $\mid \overline{X}_T - \overline{X}_U \mid$, then the $\hat{\tau}$ estimate will be biased.

4. **Functional form adherence -** The true relationship is assumed to be of the following **functional form**:

$$Y_i = \alpha + \tau D_i + \gamma X_i$$

Where:

- Y_i is the average number of public goods provided
- D_i is the treatment status $Di = 1$ for female-led towns and $D_i = 0$ for male-headed towns
- X_i represents control variables, such as: per-capita income, number of residents, year of incorporation, etc.
- i is each individual town

NOTE: *(The write-up for these assumptions were adapted from my own work in Problem Set 2)*

Examples of why this SOO Regression Adjustment may be problematic

1. The **conditional independence** assumption is a very, very **strong assumption** and it is fundamentally untestable. It is *highly unlikely* that this assumption would hold because it requires essentially controlling for every single observable that might be potentially correlated with the provision of public goods. For instance, some observables that are likely correlated with public goods provision in any given town (and for which we do not have administrative data), include: town's government budget, the town's debt level, corruption, etc. A town with a higher budget allocation would have more funds that can be destined to public goods provision, likewise an indebted town is unlikely able to provide public goods if most of the budget is oriented to pay-off its debt. If a given town has a high degree of corruption, which can be measure in accounting terms, then public goods provision will be reduced. Essentially, out estimates are affected by **omitted variable bias** because we cannot account for every variable that is correlated with public goods provision.

2. Another example is that we might not have **common support** for all covariates in our administrative data. It's possible that there are no observations for female-led governments (since there are significantly fewer in quantity) for all covariates (per-capita income, share of gross city product devoted to manufacturing, etc). If this is the case, we will recover a **biased** ATE estimate because we will be unable to compare female-led vs. male-led towns in all observables. It could well be that the covariates on which we are controlling both groups are precisely those covariates for which female-led towns have a lower performance relative to male-led towns, leading us to underestimate the true effect of having a female-headed government. Since we don't have the full picture, we are making biased comparisons with part of the picture.

III.C: PROGRAMEVAL understands your concerns, but has some in-house machine learning experts. They tell you that they can use this same administrative data to solve your issues. Do you agree? Why or why not? Be specific.

In principle, machine learning methods cannot solve my concerns of using an Adjusted SOO Regression. Causal inference is focused on estimating the *causal effect of female-led towns in the provision of public goods*, that is, finding τ^{ATE}. The focus of program evaluation is recovering **unbiased estimates**.

On the other hand, Machine Learning is concerned with **prediction**, that is, finding the best guess for a given outcome $\hat{Y}$. For example, PROGRAMEVAL's experts might be able to predict what the *average number of public goods* is in a given town, given features/covariates X_i such as per-capita income or the number of residents in a town. The experts may employ different ML methods to find the function $f(X)$ that best predicts the outcome Y by minimizing bias without sacrificing variance too much, but ultimately their goal is prediction $\hat{Y}$, when I need the causal effects τ^{ATE}.

ML experts, however, could help to figure out which covariates actually matter in driving public goods provision. The root of the problem is we don't know if we are controlling for the right stuff. So, ML can help us by following four steps:

1. Predict public good provision $\hat{Y}$ as a function of the covariates X (ie. per-capita income, number of residents, etc.)

2. Predict treatment status, ie. female-headed town D as a function of the covariates X

3. Calculate the outcome and treatment status residuals:

$$Y^R = Y - \hat{Y}$$

$$D^R = D - \hat{D}$$

4. Estimate the treatment effects τ by running the following regression:

$$Y_i^R = \alpha + \tau D^R + \varepsilon_i$$

   Where:

   - Y_i^R are the public good provision residuals
   - D^R are the treatment status residuals, ie. female-led towns

Of these four steps, (1) and (2) can be done with any ML methods, ideally some that get the best bias-variance trade-off without overfitting the training data.

$\longrightarrow$ Nonetheless, for this approach to work, **SOO assumptions need to hold** and my main concern in part (III.B) is precisely that the Adjusted Regression identifying assumptions are *highly unlikely to hold*. Unfortunately, ML experts will *not be able to help me solve these issues*, even with the most sophisticated techniques available.

III.D: PROGRAMEVAL forgot to tell you that, in India, certain local government positions are "reserved" for women – meaning only women can run for office to fill these seats (this is, again, a Real Thing!). They inform you that towns are selected to have reserved seats based on their political party. In Uttar Pradesh, an Indian state, all towns are put on a list, ordered by the share of women in the population. Each election cycle, the top 500 towns on the list are required to reserve the leadership positions for women (though, be warned – official rules aren't always perfectly followed). PROGRAMEVAL asks you whether you can use Uttar Pradesh as a test case. Describe, in math and words, the research design you would use to leverage this new information. Be sure to include a regression equation. Under what conditions would this approach estimate the causal effect of female leadership on public goods provision? For whom is this causal effect identified?

Yes, Uttar Pradesh can be used as a test case. I would employ a **Regression Discontinuity Design RD** to leverage this new information. The benefit of an RD design is that it allows us to mimic random assignment

into female-led governments by comparing the towns that just made it into the top 500 towns, based on the lowest share of women in the town's population, with towns that just barely didn't make it to the cut-off list of top 500. This is "as good as random" because the cutoff point of top 500 towns is an arbitrary policy cut-off point.

Since we are warned that *"offical rules aren't always perfectly followed"*, this means there is **imperfect compliance** across the cut-off point. That is, not all towns that made it to the top 500 list reserve leadership positions for women and viceversa. Therefore, our design will be a **Fuzzy RD**, that focuses on the **probability** of having leadership positions reserved for women. Mathematically, this can be expressed as:

$$Pr(D_i = 1 | X_i \geq c) - Pr(D_i = 1 | X_i < c) = k$$

Where:

- D_i is treatment status, $Di = 1$ for leadership positions reserved for women and $D_i = 0$ for no leadership position reserved for women
- X_i is the running variable, which is the list rank
- i is each individual town
- c is the cut-off point, ie. rank 500 on the list
- k is the difference in probabilities, that can take any value between 0 and 1. If there was perfect compliance: k = 1

This means that the probability of being treated (ie. leadership positions reserved for women) is higher if the town is above the cutoff point (ie. in the top 500 list with the lowest share of women), and likewise, the probability of leadership positions being reserved for women is lower if the town did not make it to the top 500 list.

(1) Regression Equations

Since treatment status D_i (ie. leadership seats reserved for women) does not change by 100% at the cutoff, crossing the cutoff essentially changes the *probability of treatment.* To account for this incomplete change in treatment status, and in order to estimate treatment effects τ for a Fuzzy RD, we instrument on being below the cutoff point $\longrightarrow Z_i = 1[X_i \leq c]$. As with any **instrumental variable approach**, we can estimate treatment effects τ with a **2 Stage Least Squares (2SLS approach)**, which requires the following two regressions:

1. **First Stage -** This captures the effect of moving from above the threshold $X_i > c$ to below the threshold $X_i \leq c$ on treatment status D_i (ie. leadership seats reserved for women).

The First Stage regression allows for quantifying non-compliance and can be expressed as:

$$D_i = \alpha + \gamma 1[X_i \leq c] + v_i$$

For:

$$c - h \leq X_i \leq c + h$$

Which is the bandwidth at which we're close to the cutoff point c. This is because we purposefully compare units around the cutoff which we assume to be statistically similar.

Where:

- Y_i is the outcome, ie. number of public goods provided
- γ estimates the change in probability of leadership seats reserved for women from crossing the threshold c
- $1[X_i \leq c]$ is a dummy variable that equals 1 if the town is below the threshold (ie. list rank 500 and below)
- X_i is the running variable, ie. list rank
- c is the cut-off point, position 500 on the list

2. **Reduced Form -** Which refers to the effect of going from above the threshold $X_i > c$ to below the threshold $X_i \leq c$ on the number of public goods provided (outcome Y). *Note: I reversed the specification from the slides because in this scenario, treatment happens BELOW the cutoff point instead of above.* The reduced form regression would be:

$$Y_i = \alpha + \theta 1[X_i < c] + \varepsilon_i$$

For:

$$c - h \leq X_i \leq c + h$$

Which is the bandwidth at which we're close to the cutoff point c.

Where:

- θ measures the effect of crossing the threshold c on the number of public goods provided
- The other variables are the same as before

At the cutoff point, the true effect of crossing the threshold on the number of public goods provided is the same as the estimated effect $\theta = \hat{\theta}$

3. **Fuzzy RD estimator**

After estimating the first stage and the reduced form, we can recover the Fuzzy RD estimator by dividing one by the other:

$$\hat{\tau}^{FRD} \frac{\hat{\theta}}{\hat{\gamma}}$$

(2) Identifying assumptions

(2.1) RD assumption

In order for an RD design to recover the causal effect of female leadership on public goods provision, one identifying assumption needs to be met $\rightarrow$, which is that towns just below the cutoff (ie. top 500 list) are statistically similar in both observables and unobservables to towns that are just above the cutoff. This also means that the cutoff is *"as good as randomly assigned"* such that any differences in the level of public goods provision we observe between female-led and male-led towns can be uniquely attributed to treatment (ie. leadership positions reserved for women).

(2.2) IV assumptions

Additional to the RD assumption, we also need the instrumental variable identifying assumptions to hold:

First Stage: The probability of treatment (ie. leadership seats reserved for women) conditional on the share of women *below* the threshold $X_i \leq c$ must be different from the probability of treatment conditional on the share of women *above* the threshold $X_i > c$ for at least some towns. In other words, we need that crossing the threshold c actually changes treatment status for some towns.

$$E[D_i \mid X_i \leq c] \neq E[D_i \mid X_i > c]$$

Here, our instrument is an indicator function for being below the threshold $Z_i = 1[X_i \leq c]$ that takes the value of 1 if below the threshold and 0 otherwise.

NOTE *I use below the threshold instead of above because in this example, towns get seats of leadership reserved for women if their share of women is in rank 500 or BELOW*

Independence: The potential outcomes Y_i and potential treatment status D_i need to be independent of treatment assignment (ie. being below the threshold $1[X_i \leq c]$).

$$Y_i(D_i, 1[X_i \leq c]), D_i(X_i \leq c), D_i(X_i > c) \perp 1[X_i \leq c]$$

Exclusion Restriction: The only way in which crossing the threshold affects the number of public goods provided Y_i is through its effect on treatment status ie. leadership seats reserved for women D_i. That is to say, the instrument is *as good as randomly assigned.*

$$Z_i \rightarrow D_i \rightarrow Y_i \tag{1}$$

And is not directly related to public goods provision Y_i:

$$Z_i \not\rightarrow Y_i \tag{2}$$

This occurs when the instrument $Z_i = 1[X_i \leq c]$ is uncorrelated with the error term ε_i.

$$Cov(Z_i, \varepsilon_i) = 0$$

Since the policy cutoff point at 500 is arbitrary, we can expect this assumption to hold.

Monotonicity: Towns below the threshold have higher treatment status than towns that are above the threshold for all the towns. There cannot be some towns that are less likely to have seats reserved for women as they move *below the cutoff point.*

$$\mid D_i(X_i \leq c) - D_i(X_i > c) \mid \geq 0$$

That is, moving across the cutoff point need not move *every single town's* treatment status (ie. having leadership seats reserved for women), but *all affected towns* need to move in the **same direction** (whether positive or negative).

(3) For whom the causal effect is identified

Provided that the 5 identifying assumptions hold, the treatment effect recovered is a **Local Average Treatment Effect (LATE)** specifically for *complier* towns right below and right above the cutoff that do not move across the threshold.
$$\tau^{FRD} = \tau^{ATE}$$

This τ^{LATE} can be different from the ATE or LATE at other cutoff points (for example if the cutoff was designated at the top 1000 list of towns with the lowest share of women).
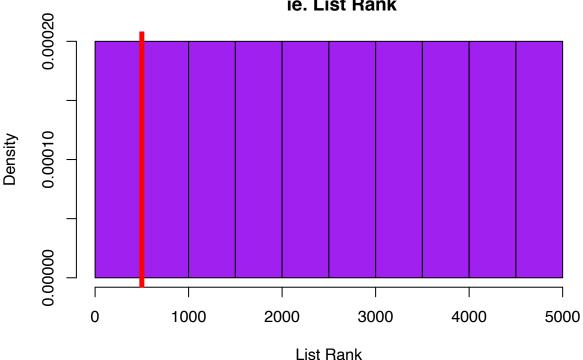
III.F: PROGRAMEVAL likes this idea, and is willing to share data with you to try this out. Use the dataset contained in final_exam_2022.csv. What empirical tests would you like to perform, prior to attempting to estimate the effect of female leadership on public goods provision, to provide evidence in support of the identifying assumption(s)? Perform at least two tests (hint: these should be simple graphical exercises). What do they tell you about the validity of the identifying assumption(s) in this case?

Before attempting to employ a **Fuzzy RD design**, we need to run empirical tests of the identifying assumptions, in graphical form.

1. **Density of the running variable** Towns could manipulate whether they receive treatment. For example, in a town where the patriarchy dominates, politicians may lobby for their town not to have leadership positions reserved for women, even if the town is in the top 500 list with the lowest share of women in the population. Likewise, more progressive towns might manipulate the system to get leadership positions reserved for women, even if they didn't make it to the top 500 list. With this graphical test, we can visualize if there is smooth density across the cutoff point. If it is smooth, then the cutoff is essentially random.

```
# Plot density of running variable, ie. share of women
hist(data$list_rank, main = "Density of Running Variable
     ie. List Rank", breaks = 15, freq = FALSE,
     xlab = "List Rank", col = "purple")
abline(v = 500, lwd=5.0, col = "red")
```


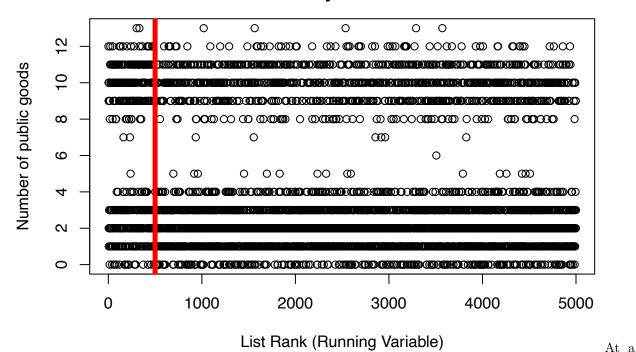**Density of Running Variable
ie. List Rank**

From
the graph, we can see that the density of the running variable (ie. the list rank) is uniform across the sample.
In particular, it is exactly smooth below and above the cutoff point of 500. This provides evidence in favor of
the threshold point being *essentially random*.

2. **Outcome across the running variable** This test allows us to see the potential for an RD design
   because it will show whether there is a discontinuous jump in public goods provision between female-led
   and male-led towns. If there is no jump, then there is no potential for this design.

```
# Bin observations by share of women, this is the running variable
# that determines if a town gets reserved seats for women or not
data$bin = floor((data$list_rank - 500/10))
bin_data = data %>% group_by(bin) %>% summarise(mean_outcome =
mean(public_goods_number), med_rank = median(list_rank))

# Plot
plot(bin_data$med_rank, bin_data$mean_outcome, main="Public Goods
     Provision by Share of Women", xlab="List Rank (Running Variable)",
     ylab = "Number of public goods")
abline(v=500, lwd=5.0, col = "red")
```
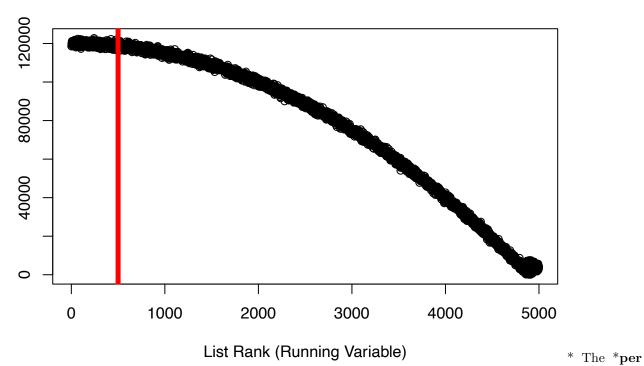
Public Goods
Provision by Share of Women

List Rank (Running Variable)

At a first glance, there does not appear to be a very dramatic point to the left of the cutoff. However, if we focus on the observations to the left of the cutoff point, we can see that the lower rank in the list (ie. the lower the share of women in a town's population), the *modestly* higher the number of public goods provided. As we move to the left, there seems to *some* increase in public goods provision, even if not very pronounced. This supports employing an RD design for this case study.s
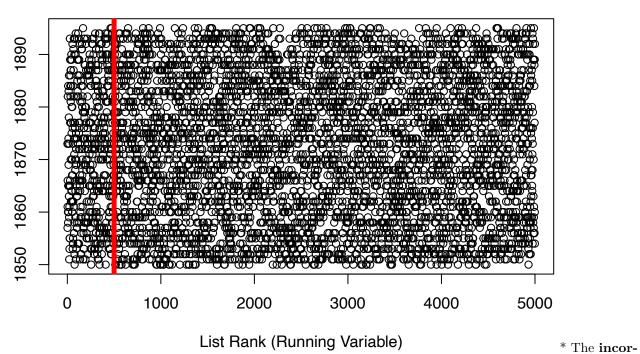
3. **Continuity in Covariates** If we observe continuity in the observables X_i across the threshold, then we can attribute any discontinuity in public goods provision Y between female-led and male-led towns solely to the treatment (leadership seats reserved for women). This acts as a *proxy test* for the RD identifying assumption, since we cannot observe nor plot potential outcomes of treated towns in the absence of treatment.

```
# Use bins
covar = data %>% group_by(bin) %>% summarise(mean_income = mean(per_capita_income_rupees),
    mean_incorp = mean(incorp_year),
    mean_manuf = mean(manufacturing_product_share),
    mean_res = mean(number_of_residents),
    med_rank = median(list_rank))

# (1) Plot per capita income
plot(covar$med_rank, covar$mean_income, main = "Per capita income in rupees",
    xlab = "List Rank (Running Variable)",
    ylab = "")
abline(v = 500, lwd=5.0, col = "red")
```

Per capita income in rupees

List Rank (Running Variable)

* The *per capita income (in rupees)** looks pretty smooth just below and just above the cutoff.

```
# (2) Plot incorporation year
plot(covar$med_rank, covar$mean_incorp, main = "Incorporation Year",
     xlab = "List Rank (Running Variable)",
     ylab = "")
abline(v = 500, lwd=5.0, col = "red")
```
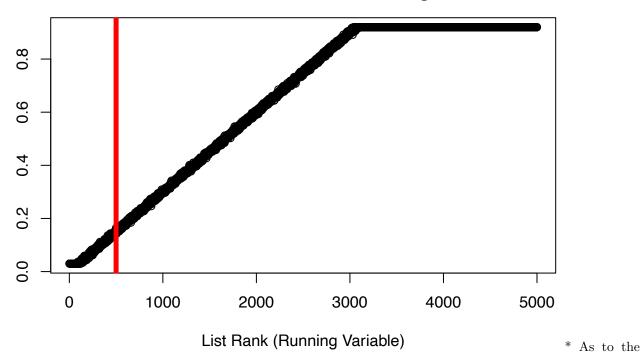
Incorporation Year

List Rank (Running Variable)

* The **incorporation year** appears to be *mostly smooth* just below and just above the threshold. Overall, there does not appear to be a very marked difference in continuity in this covariate.
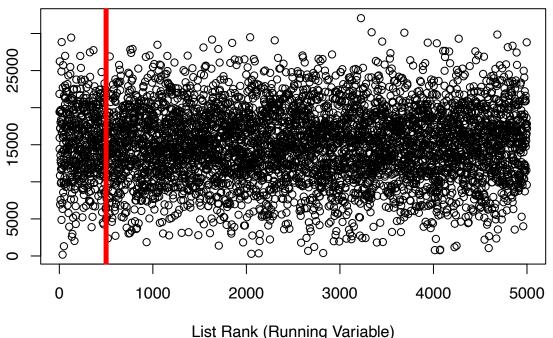
```r
# (3) Plot share of gross city product devoted to manufacturing
plot(covar$med_rank, covar$mean_manuf, main = "Share of town product
     devoted to manufacturing",
     xlab = "List Rank (Running Variable)",
     ylab = "")
abline(v = 500, lwd=5.0, col = "red")
```

Share of town product
devoted to manufacturing

List Rank (Running Variable)

* As to the **share of town product devoted to manufacturing**, the difference in continuity immediately to the left vs. immediately to the right of the threshold, is *not very smooth*. Just below the threshold there is a lower share of town product devoted to manufacturing, vs. the opposite above the threshold.

```
# (4) Plot number of town residents
plot(covar$med_rank, covar$mean_res, main = "Number of residents",
     xlab = "List Rank (Running Variable)",
     ylab = "")
abline(v = 500, lwd=5.0, col = "red")
```

Number of residents

List Rank (Running Variable)

* In terms of the **number of residents**, this covariate appears to be *smooth* just below and just above the threshold.

With the exception of share of town product devoted to manufacturing, all other covariates are smooth right below and above the cutoff point, providing evidence in support of the identifying assumption, based on the analyzed covariates. While we cannot definitively test for the RD identifying assumption, these graphs act as a proxy test, suggesting that among the available covariates in the dataset, towns just below the cutoff point are very similar to towns just above the cutoff point.

Conclusion:

In all, the performed graphical tests suggest that the RD identifying assumption is valid for the state of Uttar Pradesh in India, since 1) the density of the running variable (ie. list rank) is completely smooth around the cutoff point, 2) there is a *relative* visible change in outcome, where we see higher observations in the number of public goods provision as we move to the left of the threshold, and 3) out of 4 of the available covariates, 3 are smooth just below and just above the cutoff point.

As to the IV identifying assumptions, the **independence and exclusion restriction** give us **covariate smoothness**, which we checked for in these graphs. Even though we have not explicitly tested for the IV assumptions (ie. the exclusion restriction is fundamentally untestable), these visual graphs so far suggest the identifying assumptions are valid and that as part of a natural experiment, the cutoff point is essentially random. Therefore, we can expect the exclusion restriction to hold. Overall, we can feel more comfortable about the identifying assumptions holding and proceed to implement a Fuzzy RD design.

III.G: Plot the relationship between a town's position on the list and its likelihood of having a female leader. Describe what you're plotting, using a definition from the course. Plot the relationship between the probability of having a female leader and public goods provision. Describe what you're plotting, using a definition from the course. Informed by these plots, write down your preferred regression equation(s) for estimating the causal effect of female leadership on public goods provision. Defend your choice of bandwidth and any functional form choices you make.

In the data we are given, we are provided with the variable *reservation*, which is a dummy that takes a value of 1 for all treated towns (ie. all those with rank 500 and below in the list).
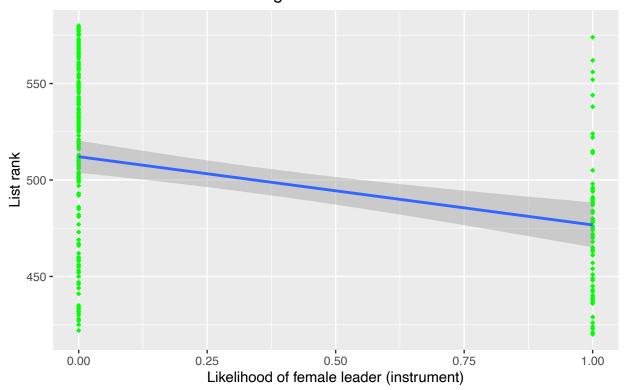
We are also provided with the variable *female_leader*, which I will assume represents the probability of having a female leader. In other words, this represents our **instrument**, because the likelihood of having a female leader is contingent upon moving below the threshold: $\rightarrow Z_i = 1[X_i \leq c]$

As we can observe from the data, some towns in the top 500 list have a 0 likelihood of having a female leader, which reflects the non-compliance in the sample.

1. **(1) Plot relationship between a town's position on the list and its likelihood of having a female leader**

```r
# Subset data by a bandwidth h = 80, where lower bound = 420 and upper bound =
# 580
filtered_data <- data %>% filter(list_rank %in% 420:580)
# Plot list rank against likelihood of female leader
ggplot(filtered_data, aes(x=female_leader, y=list_rank)) +
geom_point(shape=18, color="green") + geom_smooth(method=lm) +
ggtitle("Relationship between a town's position on the list and
        its likelihood of having a female leader at +/- 80 from the cutoff") +
  theme(plot.title = element_text (hjust = 0.5)) + xlab("Likelihood of female leader (instrument)") +
        ylab("List rank")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

Relationship between a town's position on the list and its likelihood of having a female leader at +/− 80 from the cutoff

Focusing around $+/-$ 80 bandwidth from the cutoff point of 500, this plot represents our **First Stage**, because it indicates the effect of the instrument (ie. moving from above to below the threshold $X_i > c$ to $X_i \leq c$) ON our treatment status D_i, which is partially determined by the town's rank in the list.

The line of best fit, comes from a regression of our instrument (likelihood of female leader) on treatment status. We can observe a downward sloping relationship between list rank and the likelihood of a female leader, where, from list rank 500 and above, the likelihood of female leaders is essentially 0. Conversely, from list rank 500 and below, we see a higher probability of having female leaders.

```
female = nrow(subset(data, female_leader==1))
male = nrow(subset(data, female_leader== 0))
total = female + male
cat("The overall probability of having  a female leader is", female/total)
```

```
## The overall probability of having  a female leader is 0.2676
```

```
cat("\nThe overall probability of having  a male leader is", male/total)
```
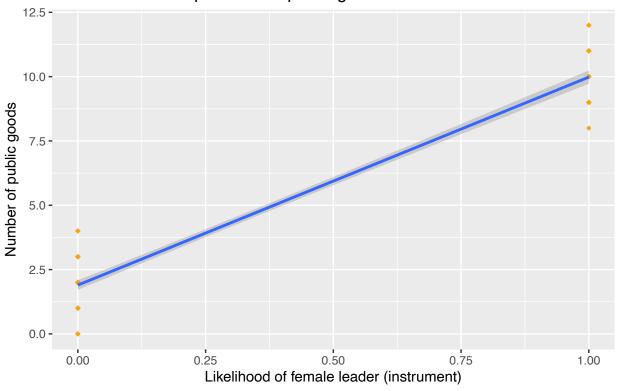
```
##
## The overall probability of having  a male leader is 0.7324
```

1. **(2) Plot relationship between the probability of having a female leader and public goods provision**

```
ggplot(filtered_data, aes(x=female_leader, y=public_goods_number)) +
geom_point(shape=18, color="orange") + geom_smooth(method=lm) +
ggtitle("Relationship between the likelihood of having a female leader
        and the provision of public goods at +/- 80 from the cutoff") +
  theme(plot.title = element_text (hjust = 0.5)) + xlab("Likelihood of female leader (instrument)") +
        ylab("Number of public goods")
```

Relationship between the likelihood of having a female leader and the provision of public goods at +/− 80 from the cutoff

This second graph reflects the **Reduced Form**, since we are plotting our instrument (ie. probability of having a female leader by moving from above to below the threshold $X_i > c$ to $X_i \leq c$) against our outcome Y, which is the number of public goods provided.

From the graph, we can observe a very clear positive relationship between Z and Y, such that when there is a high likelihood of having a female leader, the number of public goods provided is higher, and conversely, when there is no probability of having a female leader, there is a low amount of public goods provision. This graph appears to support the notion that having more women in positions of political power can improve welfare, as measured in terms of public goods provision.

Preferred regression equations for estimating the causal effect of female leadership on public goods provision

Based on these graphs, my preferred equations for estimating the causal effect of female leadership on public goods provision is essentially to estimate the First Stage and the Reduced Form in order to get the ratio between both and recover $\tau^{FRD} = \tau^{ATE}$.

1. As to the **First Stage**, I would run the following regression equation:

$$D_i = \alpha + \gamma 1[X_i \leq c] + v_i$$

Where:

- Y_i is the outcome, ie. number of public goods provided
- γ estimates the change in probability of leadership seats reserved for women from crossing the threshold c
- $1[X_i \leq c]$ is the instrument of moving below the threshold, in this example the likelihood of a female leader

- X_i is the running variable, ie. list rank
- c is the cut-off point, 500

Limited to the following bandwidth:

$$500 - 80 \leq X_i \leq 500 + 80$$

$$420 \leq \text{list rank} \leq 580$$

There is a bias-variance trade-off with the choice of bandwidth, h. Ideally we want a small bandwidth because it can yield a more unbiased estimate (since we are limiting the analysis to towns very close to the cutoff point, which are good counterfactuals for each other). Essentially, the closest to the cutoff point, the more statistically similar are towns just above to towns just below the threshold.

However, the closer we get to the threshold, the fewer observations we have to compare with, which increases the standard errors, decreasing precision. For this reason, I chose a bandwidth of 80 because it is close enough to the cutoff point to observe effects, without being too close to risk decreasing precision too much.

2. Secondly, I would run the following **Reduced Form**, regression:

$$Y_i = \alpha + \theta 1[X_i \leq c] + \varepsilon_i$$

Where:

- θ measures the effect of crossing the threshold c on the number of public goods provided
- The other variables are the same as above

Again, this would be estimated for the following bandwidth:

$$420 \leq \text{list rank} \leq 580$$

3. With out γ and θ estimates, we can proceed to recover our *Fuzzy RD estimator* by dividing the Reduced Form by the First Stage:

$$\hat{\tau}^{FRD} = \frac{\hat{\theta}}{\hat{\gamma}}$$

III.H: Finally, estimate the causal effect of female leadership on public goods provision. What do you find? Interpret your results. Advise PROGRAMEVAL: should they expand female leadership to all towns?

1. **First Stage Regression and Interpretation**

```
# Run First Stage
first_stage <- lm(reservation ~ female_leader, data = filtered_data)
gamma <- first_stage$coefficients[[2]]
cat("The effect of crossing the threshold on treatment status (gamma) is",gamma)
```

```
## The effect of crossing the threshold on treatment status (gamma) is 0.4509434
```

```
summary(first_stage)
```

```
##
## Call:
## lm(formula = reservation ~ female_leader, data = filtered_data)
##
## Residuals:
```

```
##     Min     1Q  Median     3Q     Max
## -0.8000 -0.3491  0.2000  0.2000  0.6509
##
## Coefficients:
##             Estimate Std. Error t value        Pr(>|t|)
## (Intercept)  0.34906    0.04417   7.902 0.000000000000428 ***
## female_leader 0.45094    0.07558   5.967 0.000000015196636 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4548 on 159 degrees of freedom
## Multiple R-squared:  0.183,  Adjusted R-squared:  0.1778
## F-statistic:  35.6 on 1 and 159 DF,  p-value: 0.0000000152
```

Our First Stage estimate, $\hat{\gamma}$ shows that the estimated difference between the likelihood of having a female leader (which is determined by crossing from above $X_i > c$ to below the cutoff point $X_i \leq c$), and having a male leader, ON the likelihood of having leadership positions reserved for women) is *approximately 45%*. This coefficient is highly statistically significant at all levels of conventional significance. The sign of the coefficient is positive, as expected. That is the likelihood of female leaders increases if the town is in the top 500 list, where seats are often reserved for women (not always).

Also, the F-statistic $2495 > 20$, shows the instrument (likelihood of female leader) satisfies the first stage assumption.

2. **Reduced Form Regression and Interpretation**

```
# Run Reduced Form
reduced_form <- lm(public_goods_number ~ female_leader, data = filtered_data)
theta <- reduced_form$coefficients[[2]]
cat("\n The effect of crossing the threshold on public goods provision (theta)
    is", theta)
```

```
##
##  The effect of crossing the threshold on public goods provision (theta)
##     is 8.085592
```

```
summary(reduced_form)
```

```
##
## Call:
## lm(formula = public_goods_number ~ female_leader, data = filtered_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9818 -0.8962  0.1038  1.0182  2.1038
##
## Coefficients:
##             Estimate Std. Error t value        Pr(>|t|)
## (Intercept)   1.8962     0.0946   20.04 <0.0000000000000002 ***
## female_leader 8.0856     0.1619   49.95 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.974 on 159 degrees of freedom
## Multiple R-squared:  0.9401, Adjusted R-squared:  0.9397
## F-statistic:  2495 on 1 and 159 DF,  p-value: < 0.00000000000000022
```

This regression provides a * $\hat{\theta}$ estimate, which captures the effect of crossing the threshold (ie. likelihood of female leader) on the outcome (ie. public goods provision). The estimated difference between the likelihood of having a female leader (which is determined by crossing from above $X_i > c$ to below the cutoff point $X_i \leq c$), and having a male leader, ON public goods provision is approximately *8 additional public goods being provided.*

This coefficient is highly statistically significant at all levels of conventional significance. Moreover, given that the mean provision of public goods is 4.12, this is a very, very meaningful coefficient.

```
mean_goods = mean(data$public_goods_number)
cat("The mean number of public goods provided is", mean_goods)
```

The mean number of public goods provided is 4.12

the instrument (randomized electricity price variation) on the outcome (groundwater consumption). As described in question (4), this is the **Reduced Form**.

Every additional USD dollar in electricity price variation, is associated with a - 147.50 acre-feet decrease in groundwater consumption. This $\hat{\theta}$ estimate is highly statistically significant at all conventional levels of significance. The sign of the estimate is logical by economic theory, since we can expect an increase in electricity prices to reduce consumption.

2. **Recover Fuzzy RD LATE estimate**

```
# Fuzzy RD LATE
tau_late= theta/gamma
cat("\n The LATE effect of female leadership on public goods provision for
    complier around the cutoff is estimated to be ", tau_late)
```

```
##
##  The LATE effect of female leadership on public goods provision for
##     complier around the cutoff is estimated to be  17.93039
```

After estimating the First Stage and the Reduced Form, we obtain that:

$$\hat{\tau}^{FRD} = \frac{\hat{\theta}}{\hat{\gamma}}$$

$$\hat{\tau}^{FRD} = \frac{8.0\hat{8}56}{0.4\hat{5}094}$$

$$\hat{\tau}^{FRD} = 17.93 = \tau^{LATE}$$

Therefore, our estimated causal impact of female leadership on public goods provision for *complier towns* around the cutoff with a bandwidth of +/- 80 is estimated to be τ^{FRD} is 17.93, which is *quite a sizable effect.* It should be highlighted that this LATE is specifically for complier towns and should the threshold change OR should the bandwidth be adjusted, we can expect this effect to change.

4. **Advise PROGRAMEVAL** Based on the results obtained from the Fuzzy RD design, there is evidence to support that having female-headed government is correlated with an increase in public goods provision, and therefore in social welfare. However, I would emphasize that the finding from this analysis is *specific and particular* to 1) complier towns around the cutoff, (that is to say that towns that didn't make it to the top 500 list did not reserve leadership positions for women and viceversa), and 2) towns within +/- 80 rank positions around the cutoff.

Therefore, I cannot advise PROGRAMEVAL to expand female leadership to ALL towns, because that would require estimating the **Average Treatment Effect (ATE)**, which we cannot do with this research design. If that is PROGRAMEVAL's interest, we would have to employ a different method such as an RCT. However, what I can advise PROGRAMEVAL is to increase the cutoff point to at least 580 instead of 500 to benefit more towns that can stand to gain from female-led governments. To determine the ideal cutoff point would require adjusting the bandwidth in coordination with the state of Uttar Pradesh.