

# Program Evaluation Final Exam

Hye-Min Jung

6/4/2020

## I.A

**Describe the research question in this paper in words.**

- The research question is “Default effects on program(time-varying electricity pricing) adoption” in the context of a residential electricity-pricing program. More specifically, the analysis is based on a field experiment run by the Sacramento Municipal Utility District (SMUD) in 2011-2013.

**Explain, in words and math, the ideal experiment one might want to use to answer this question.**

- In a completely unconstrained world, where I could do anything with almighty ability, I would run randomized controlled trial, to estimate “Default Effects on Program Adoption”. Because under the randomized controlled trial, we can easily estimate the impact of treatment. ATE is simply the difference in means between treated and control groups.
- Define:  $D_i \in \{0, 1\}$  as the treatment indicator for household  $i$ 
  - when household  $i$  is treated,  $D_i = 1$
  - when household  $i$  is not treated,  $D_i = 0$
- Define:  $Y_i(D_i)$  as the electricity usage(kWh) as a function of  $D_i$ 
  - when household  $i$  is treated, we observe electricity usage(kWh)  $Y(1)$
  - when household  $i$  is not treated, we observe electricity usage(kWh)  $Y(0)$
- Then, the impact of treatment for electricity usage(kWh) is:  $\tau_i = Y_i(1) - Y_i(0)$
- To put it simply in words, impact of treatment is a parameter that measures ‘What is the outcome of being invited to opt-in/opt-out to a new time-varying pricing structure, compared to the outcomes that we would have observed without the treatment?’
  - Impacts is ‘changes in outcomes caused by the policy’ whereas, outcomes is ‘things that we could “potentially” observe’.
  - We need to consider all possible outcomes we could have observed, spans both actual and alternative programs, to explain the impact of treatment.

## I.B

**Explain, in words and math, what treatment parameter the authors are recovering in Equation (1).**

- The author estimates a difference-in-differences (DID) specification using data from the pre-treatment and treatment periods to identify the average intent to treat (ITT) effect.
  - This gives estimation of ITT for experimental treatment groups
- In math,  $\tau^{ITT} = \bar{Y}(R_i = 1) - \bar{Y}(R_i = 0)$

**Describe each term in the estimating equation in words, including a discussion of the unit of analysis.**

- $y_{it} = \alpha + \beta_{ITT}Z_{it} + \gamma_i + \delta_t + \varepsilon_{it}$ 
  - Unit of analysis  $i$  is household
  - $y_{it}$  measures hourly electricity consumption for household  $i$  in hour  $t$
  - $\beta_{ITT}$  is coefficient of interest
  - $Z_{it}$  is an indicator variable equal to 1 starting on 2012/6/1 if household  $i$  was encouraged to be in the treatment group, and 0 otherwise.
  - $\gamma_i$  is a household fixed effect that captures systematic differences in consumption across households, and  $\delta_t$  is an hour-of-sample fixed effect.
  - $\varepsilon_{it}$  is an error term

**Explain how – if at all – this estimate differs from the population-wide average treatment effect of time-varying electricity pricing on electricity consumption.**

- This estimate is same with the population-wide average treatment effect of time-varying electricity pricing on electricity consumption.

**What assumptions are required in order for Equation (1) to recover the causal effect of interest?**

- In order for us to get an unbiased estimate and recover the causal effect of interest, we need:  $E[Y_i|R_i = 1, D_i = 0] = E[Y_i|R_i = 0, D_i^* = 0]$ .
  - In other words, assignment to treatment didn't affect the likelihood of non-compliance.

**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.**

- This assumption is likely to be satisfied in this context.
  - The exclusion restriction implies that always takers in the opt-out group are responding to the time-varying rates in the same way as their counterparts in the opt-in group. Under this assumption, differences in these estimated ITT effects across the opt-in and opt-out groups are driven by a demand response among complacents. We have also estimated the opt-in and opt-out equations jointly so that we could test equality of the coefficients. We can reject equality with at least 95% certainty in all cases except for event day TOU, where  $p = 0.055$ . (page 17).

## I.C

**Explain, in words and math, what treatment parameters the authors are recovering in Equation (2).**

- The author recovers Local Average Treatment Effect, which is the effect of treatment for the compliers.
- In math,  $\tau^{LATE} = \frac{\bar{Y}(R_i=1) - \bar{Y}(R_i=0)}{\pi^C}$

**Describe each term in the estimating equation in words, including a discussion of the unit of analysis.**

- $y_{it} = \alpha + \beta_{LATE} Treat_{it} + \gamma_i + \tau_t + \varepsilon_{it}$ 
  - Unit of analysis  $i$  is household
  - $Treat_{it}$  is an indicator variable equal to 1 starting on 2012/6/1 if household  $i$  was actually enrolled in treatment, 0 otherwise (estimated separately for the opt-in and opt-out groups)
  - $Z_{it}$  randomized encouragement to the corresponding treatment and is instrumented for  $Treat_{it}$
  - $\beta_{LATE}$  coefficient captures the LATE

**Explain how – if at all – this estimate differs from the population-wide average treatment effect of time-varying electricity pricing on electricity consumption.**

- Eventhough the LATE estimate for opt-inner is bigger than opt-outer, because there are many more households who stay in the opt-out treatment, then who stay opt-in to opt-in treatment, so the average reductions for the opt-out group are nearly 3 times larger than the average reductions for the opt-in group for CPP and 2 times larger for TOU.

**Describe two different procedures to estimate this treatment parameter.**

- Under constant treatment effects: equal to estimating procedure of ATE, ATT
- With heterogeneous treatment effects: equal to estimating procedure of ATT
- Estimation can be also done by using “instrumental variable”.  $\hat{\tau}_{ITT}/\hat{\pi}^C$

**What assumptions are required in order for Equation (2) to recover the causal effect of interest?**

- To interpret  $\beta_{LATE}$  as a causal effect, there should be no defiers and we must invoke an exclusion restriction, which requires that the encouragement (i.e., the offer to opt in or default assignment into treatment with the ability to opt out) affects electricity consumption only indirectly via an effect on participation.
- We also invoke a monotonicity assumption which requires that our encouragement weakly increases (versus reduces) the participation probability for all households.

**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.**

- No defiers assumptions are not likely to be satisfied in this context.
  - Because in the real world, we cannot rule out the possible existence of single defier who is being untreated if put in the treatment group or treated if put in the control group.
  - If you refer to the fraction of participants in table 4 and Figure 4 that shows the grid of the identification of always takers, complacents, and never takers, we can know that there is defiers in the population.

## I.D

**Describe the main results of the paper.**

- Overall, the average reductions for the opt-out group are nearly 3 times larger than the average reductions for the opt-in group for CPP and 2 times larger for TOU.
- Eventhough unit specific effect for opt-inner is big, there are many more households who stay in the opt-out treatment, then who stay opt-in to opt-in treatment, so the average reductions for the opt-out group are nearly 3 times larger than the average reductions for the opt-in group for CPP and 2 times larger for TOU.
  - While the unit specific treatment effect for the opt-in group is way bigger than opt-out group, there are not many of opt-inners. (You can imagine that opt-inners are energy-consious households who tend to paying attention to the program. Whereas, opt-outers are likely to be households who does not pay attention to the letters carefully.)
  - While the unit specific treatment effect for the opt-out group is way smaller than opt-in group, there are many of opt-outers, who stays under the opt-out category.
  - Complacents, households who followed what researchers told them to do, regardless of what group they are put into(not actively changing their status), we don't see huge treatment effect.
  - LATE for opt-inner and opt-outers are potentially different: LATE for compliers in opt-inners are bigger and LATE for compliers in opt-outers are smaller.
- The fact that the effects are smaller in opt-outers but because there are just so many of them, the total effect for the opt-out group turns out to be larger than the opt-in group, has important implications for how to implement this type of policy.

**Include a discussion of (at a minimum) Tables 3 and 4, in which you interpret the estimated coefficients and describe their magnitudes.**

- Table 3 Intent to treat effects (Encouragement (CPP) row under Critical event column):
- Notice how treatment effect for opt-in unit is different from treatment effect for opt-out unit.
- The average reductions for the opt-out group are nearly 3 times larger than the average reductions for the opt-in group for CPP and 2 times larger for TOU.
  - Providing households the opportunity to opt-in to the CPP treatment leads to an average reduction in electricity consumption of 0.129 kWh during peak hours of event days (averaged across all household that received the opt-in offer).
  - Providing households the opportunity to opt-out to the CPP treatment leads to an average reduction in electricity consumption of 0.305 kWh during peak hours of event days (averaged across all household that received the opt-out offer).
  - This treatmnet effect difference persists across all the treatment assignments.
  - There are 2 underlying facts for this finding: (1) It's the difference in the number of people who opt-in vs. opt-out, (2) it's treatment effect for the people who stayed under the both opt-in, opt-out treatment.
- Table 4 Average treatment effects shows how complacent households reduced consumption by less during critical peak periods. You can know that unit specific treatment effect for the opt-in group is way bigger than opt-out group and we don't see huge treatment effect in complacents.
  - The ATE of encouragement assignment(CPP) on average hourly electricity usage in kilowatts in critical event hour for opt-inners is  $-0.658$ .
  - The ATE of encouragement assignment(CPP) on average hourly electricity usage in kilowatts in critical event hour for opt-outers is  $-0.330$ .
  - The ATE of encouragement assignment(CPP) on average hourly electricity usage in kilowatts in critical event hour for complacents is  $-0.242$ .

- This treatment effect difference persists across all the treatment assignments.

**What is the main policy take-away of the paper?**

- One must consider not only the initial choice subject to the default manipulation, but also any “follow-on” behaviors that can depend on the initial choice.
- Given that many default manipulations aim to induce changes in follow-on behavior, it is important to account for both direct and indirect impacts of default manipulations on economic outcomes.

**Describe the results in Table 6.**

- Table 6 analyzes the pricing programs from the perspective of the utility, comparing the costs of enrolling participants and implementing the program against the benefits (i.e., costs avoided when peak consumption is reduced).
- In the final column of Table 6, net benefits are reported. Researchers estimated that both opt-out programs would be cost-effective. The CPP opt-in program is estimated to be marginally cost-effective. The TOU opt-in program, which led to much smaller demand reductions than the CPP program, is projected to incur costs in excess of savings.

## II.A

**Describe this comparison in math and words.**

- FUNACRONYM would like to compare the average number of public goods (roads, schools, public buildings, et cetera) in towns with female-headed governments as compared with towns that have male-headed governments.
- To do this:
- Define:  $D_i \in \{0, 1\}$  as the treatment(female\_headed) indicator for town i
  - when town i is female-headed(treated),  $D_i = 1$
  - when town i is male-headed(not treated),  $D_i = 0$
- Define:  $Y_i(D_i)$  as the number of public goods as a function of  $D_i$ 
  - when town i is female-headed(treated), we observe number of public goods  $Y(1)$
  - when town i is male-headed(not treated), we observe number of public goods  $Y(0)$
- Then, the impact of female-headed government(treated) for number of public goods is:  $\tau_i = Y_i(1) - Y_i(0)$
- To put it in words, impact of treatment is a parameter that measures ‘What is the outcome of female-headed governments, compared to the outcomes that we would have observed under the male-headed governments?’
  - Impacts is ‘changes in outcomes caused by the policy’ whereas, outcomes is ‘things that we could “potentially” observe’.
  - We need to consider all possible outcomes we could have observed, spans both actual and alternative programs, to explain the impact of treatment.

**Under what conditions would this comparison estimate the causal effect of female leaders on public goods provision?**

- Conditional expectation of Y should be same as the unconditional expectation of Y, in order for this naive estimator  $\tau^N = \overline{Y(1)} - \overline{Y(0)}$  to estimate the causal effect of female leaders on public goods provision. This underlying assumption is very unlikely to happen in the real world.

**Provide two concrete examples of reasons why this comparison may be problematic.**

- There are many circumstances that involves selection problem, which makes the naive estimator different from the ATE:
  - For instance, suppose among all towns, only highly-educated, wealthy, already well established(abundant public good) towns tend to welcome, trust female-head running the government. This implies treated and untreated towns being different even under the absence of the female-head. Under this scenario, because our key assumption is broken, the naive estimator is not going to work.
- Under the scenario with this type of selection problem, we cannot distinguish between effects of the female-head and the town’s characteristics from our estimate, resulting naive estimator to be different from ATE.

## II.B

Describe, using math and words, a comparison between female and male-headed towns which leverages these administrative 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.

- $Y_i = \beta + \tau D_i + \varepsilon_i$  where,
  - $\beta$ : mean (expectation) of  $Y_i(0)$
  - $\tau$ : constant treatment effect,  $\tau_i = \tau = Y_i(1) - Y_i(0)$
  - $\varepsilon_i$ : random component of  $Y_i(0) : Y_i(0) - E[Y_i(0)]$
- Leveraging on town characteristics from the administrative data, I can use OLS regression to estimate treatment parameters.

**Under what conditions would this comparison estimate the causal effect of female leadership on public goods provision?**

- There should be no selection bias for this comparison to estimate the causal effect of female leadership on public goods provision. In math,  $E[\varepsilon_i | D_i] = 0$ . In other words, we should observe everything that is correlated with treatment and affects  $Y_i$ .

**Provide two concrete examples of reasons why this comparison may be problematic (different from what you described above).**

- Again, selection bias is a big problem for estimation.
  - For instance, suppose among all towns, only family friendly towns with above average public goods among all the towns tend to welcome, trust female-head running the government. This implies treated and untreated towns being different even under the absence of the female-head. Under this scenario, because our key assumption is broken, the naive estimator is not going to work.
- Under the scenario with this type of selection problem, we cannot distinguish between effects of the female-head and the town's characteristics from our estimate, resulting the estimator to be biased.

## II.C

**FUNACRONYM in-house machine learning experts tell you that they can use this same administrative data to solve your issues. Do you agree? Why or why not? Be specific.**

- I disagree, because machine learning is not designed for estimating  $\hat{\tau}$ , we cannot directly use ML for what we want to estimate. So machine learning experts cannot solve issue.
  - FUNACRONYM's questions is focused on the :causal effect of D on Y". Estimation in causal inference, we highly focus on the unbiasedness.
  - But machine learnin focuses on answering "the best guess of some outcome". This is prediction method, answering "given a dataset with outcome Y and covariates X, what function f (X) best predicts Y?".
  - So this what machine learning can answer is different from answering causal inference question.
  - Although I am skeptical about machine learning to answer causal inference question, I do admit that machine learning offers some opportunities for data generation, heterogeneity analysis and estimating  $\hat{\tau}$ .

## II.D

**They suggest that you use an instrumental variables approach leveraging this new piece of information. Describe, in math and words, what this IV approach would look like.**

- Let,  $Z_i$  (reservation) be instrument variable for  $D_i$
- First stage: Regress endogenous  $D_i$  on all exogenous variables, including  $Z_i$ 
  - $D_i = \alpha + \gamma Z_i + \beta X_i + \eta_i$
  - Store the predicted values of  $D_i$ ,  $\hat{D}_i$
- Second stage: Regress outcome  $Y_i$  on predicted  $D_i$  and other Xs
  - $Y_i = \alpha + \tau \hat{D}_i + \delta X_i + \varepsilon_i$
  - $\hat{\tau}$  in this equation is our IV estimate

**Under what conditions would this approach estimate the causal effect of female leadership on local public goods provision?**

- $\text{Cov}(Z_i, D_i) \neq 0$ .
  - In other words,  $Z_i$  and  $D_i$  are related
- $\text{Cov}(Z_i, \varepsilon_i) = 0$ 
  - In other words,  $Z_i$  and  $\varepsilon_i$  are not related.
  - ( $Z_i$  only affects  $Y_i$  through  $D_i$ , so  $Z_i$  is as good as randomly assigned!)

**Provide two reasons why these conditions may not be satisfied in this setting.**

- Reservation status should only affect public goods through female-head leadership effect. But under certain settings, this condition may not be satisfied.
  - First, when ruling party abide by political campaign to build highway after the re-election, the re-election affects public goods provision directly. I
  - Second, when re-elected ruling party turn interest to other town so decide to allocate all the resources to public goods in other town, whereas not investing any on public goods in re-elected town. Then the re-election affects public goods provision directly in negative way.

## II.E

**Describe, in math and words, the research design you would use to leverage this new information. Be sure to include a regression equation.**

- I would like to implement the fuzzy regression discontinuity design, using new information that the treatment status doesn't change by 100% at the cutoff. (the top 500 towns on the list are required to reserve the female leadership positions and it is not perfectly followed)
- Fuzzy regression discontinuity, we know that  $\Pr(D_i = 1|X_i \geq c) - \Pr(D_i = 1|x_i < c) = k$ , where  $0 < k < 1$ 
  - This implies that with  $X_i < c$  may get treated; more with  $X_i \geq c$  do.
  - To estimate treatment effect with a fuzzy RD, we need to account for incomplete change in treatment by estimating:



- In math,  $Y_i = \alpha + \theta \mathbf{1}[X_i \geq c] + \nu_i$  for  $c - h \leq X_i \leq c + h$ 
  - The effect of crossing the threshold on treatment status is:  $\hat{\gamma} = \bar{D}(c \leq X_i \leq c + h) - \bar{D}(c - h \leq X_i < c)$
  - In regression,  $D_i = \alpha + \gamma \mathbf{1}[X_i \geq c] + \eta_i$  for  $c - h \leq X_i \leq c + h$
  - $\hat{\gamma}$  estimates the change in probability of treatment from crossing  $c$
  - The fuzzy RD estimator gets you:  $\hat{\gamma}^{FRD} = \frac{\bar{Y}(c \leq X_i \leq c + h) - \bar{Y}(c - h \leq X_i < c)}{\bar{D}(c \leq X_i \leq c + h) - \bar{D}(c - h \leq X_i < c)} = \frac{\hat{\theta}}{\hat{\gamma}}$

**Under what conditions would this approach estimate the causal effect of female leadership on public goods provision? For whom is this causal effect identified?**

- We need the standard IV assumptions, to estimate the causal effect of female leadership by Fuzzy RD.
  - First stage:  $E[D_i | X_i \geq c] \neq E[D_i | X_i < c]$  for some  $i$
  - Independence:  $Y_i(D_i, \mathbf{1}[X_i \geq c]), D_i(X_i \geq c), D_i(X_i < c) \perp \mathbf{1}[X_i \geq c]$
  - Exclusion restriction:  $Y_i(X_i \geq c, D_i) = Y_i(X_i < c, D_i)$  for  $D_i \in \{0, 1\}$
  - Monotonicity:  $|D_i(X_i \geq c) - D_i(X_i < c)| \geq 0$  for all  $i$

## II.F

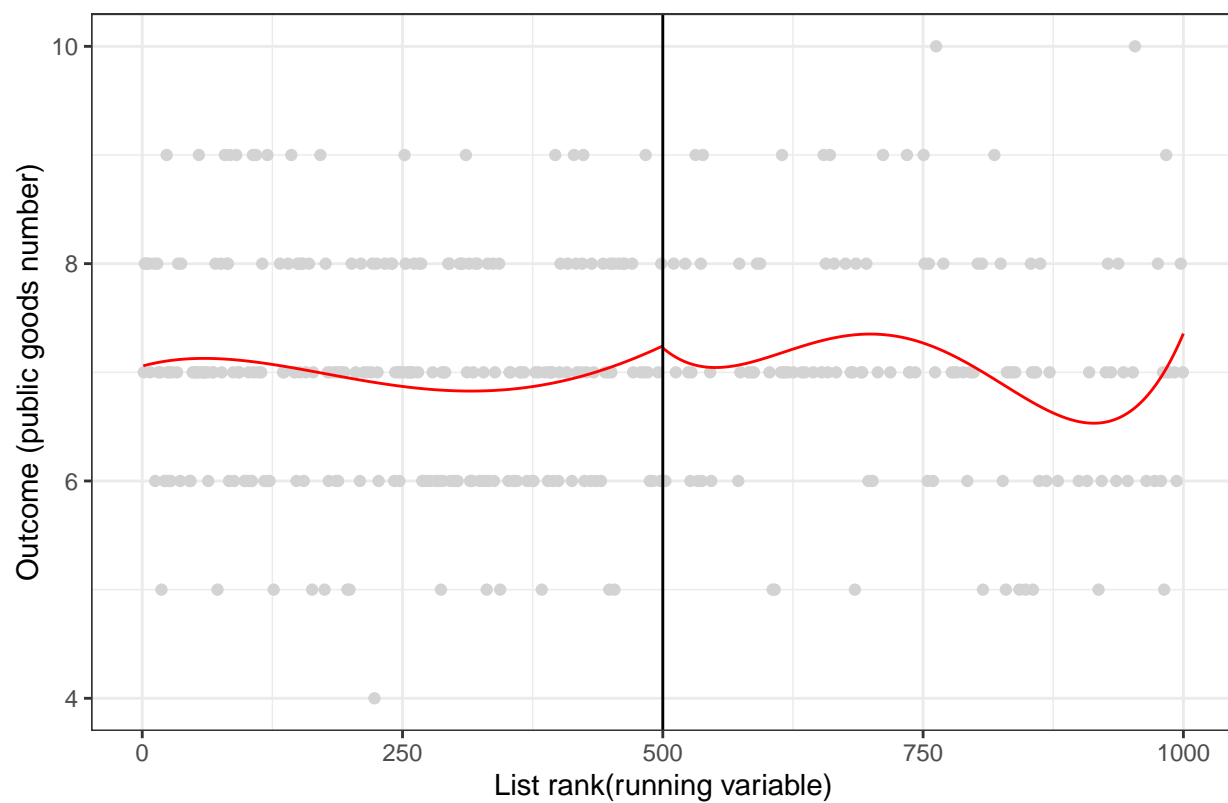
**What empirical tests would you like to perform, prior to attempting to estimate the effect of female leadership on public goods provision, to provide evidence in support of the identifying assumption(s)?**

- I would like to perform test to validate:
  - Independence:  $Y_i(D_i, \mathbf{1}[X_i \geq c]), D_i(X_i \geq c), D_i(X_i < c) \perp \mathbf{1}[X_i \geq c]$
  - Exclusion restriction:  $Y_i(X_i \geq c, D_i) = Y_i(X_i < c, D_i)$  for  $D_i \in \{0, 1\}$
  - Monotonicity:  $|D_i(X_i \geq c) - D_i(X_i < c)| \geq 0$  for all  $i$

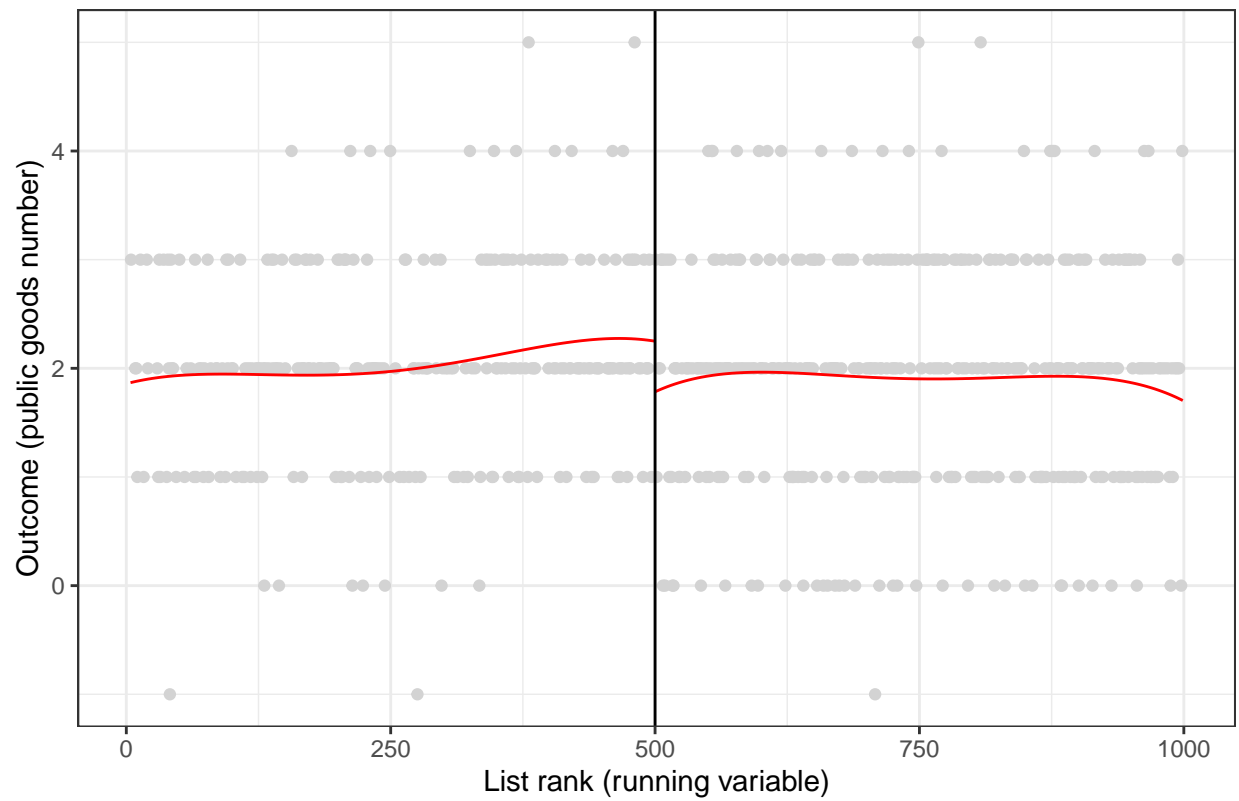
**Perform at least two tests (hint: these should be simple graphical exercises).**

- Independence and Exclusion restriction assumption together buy us, covariate smoothness
  - $E[Y_i(1) | X_i = x]$  and  $E[Y_i(0) | X_i = x]$  are continuous in  $x$
- So, I can simply look at the covariate smoothness to test for (1) Independence and (2) Exclusion restriction assumption

Test 1a. Covariate smoothness for female leader towns

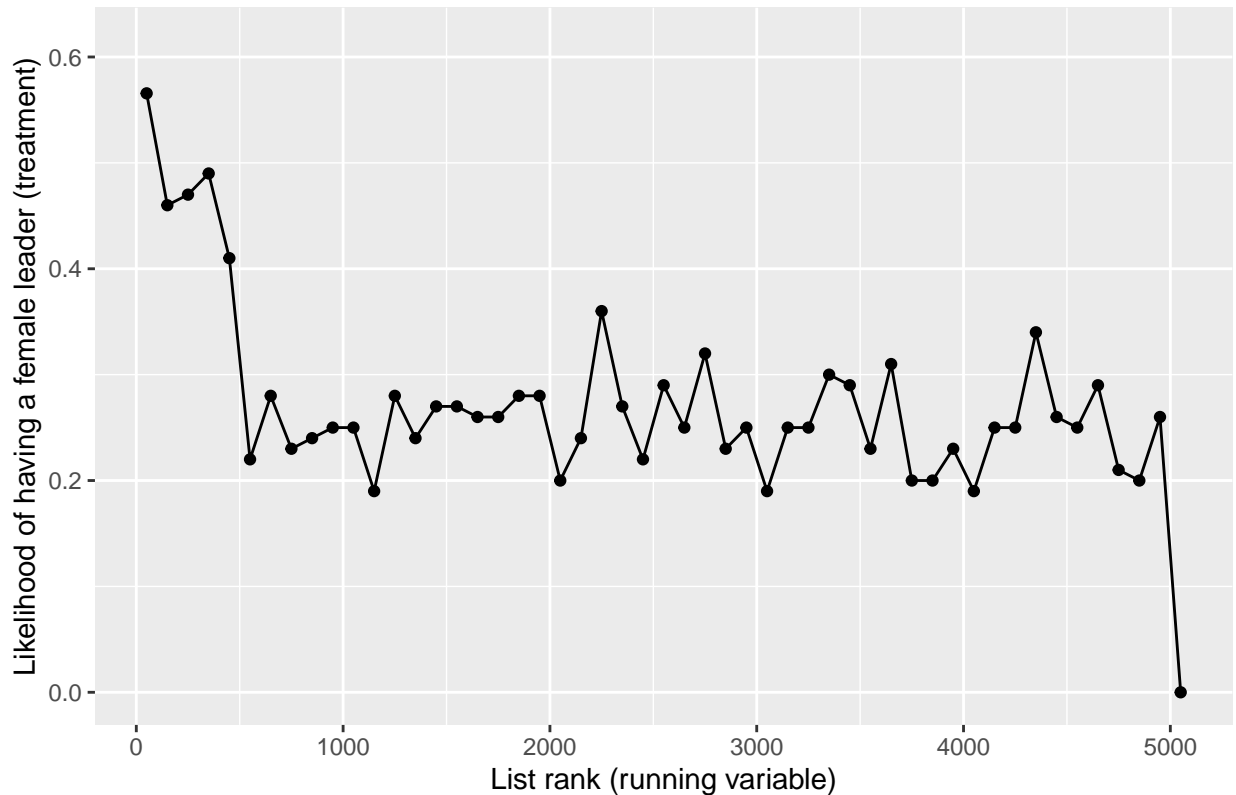


### Test 1b.Covariate smoothness for male leader towns



```
## No summary function supplied, defaulting to 'mean_se()'
## No summary function supplied, defaulting to 'mean_se()'
```

## Test 2: Monotonicity



- Testing for monotonicity plot shows that  $|D_i(X_i \geq c) - D_i(X_i < c)| \geq 0$  for all  $i$ .
  - Note that the this plot is graphed in decreasing manner, because our rank lists in the decreasing manner for the threshold (top 500 treated, bigger than 500 untreated).

What do they tell you about the validity of the identifying assumption(s) in this case?

- Validity tests suggests that the threshold only have direct effect on treatment but not having direct effect on anything else. In other words, threshold affects outcome only through treatment.
  - Independence assumption holds:  $Y_i(D_i, \mathbf{1}[\text{listrank}_i \geq 500]), D_i(X_i \geq c), D_i(\text{listrank}_i < 500) \perp \mathbf{1}[\text{listrank}_i \geq 500]$
- Monotonicity assumption holds:  $|D_i(\text{listrank}_i \geq 500) - D_i(\text{listrank}_i < 500)| \geq 0$  for all  $i$

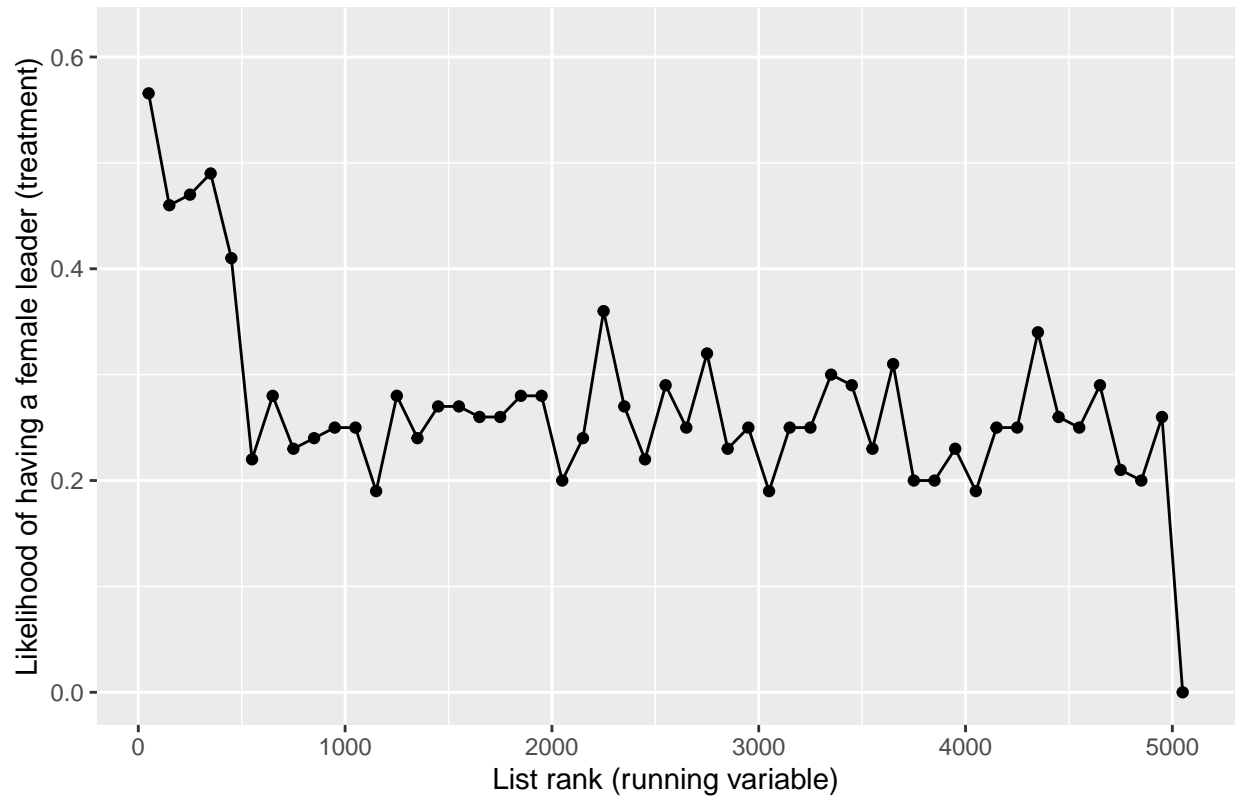
## II.G:

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.

- This is "First stage" plot with relationship between treatment assignment by running variable is visualized.

```
## No summary function supplied, defaulting to 'mean_se()'
## No summary function supplied, defaulting to 'mean_se()'
```

### First stage: List rank and its likelihood of having a female leader

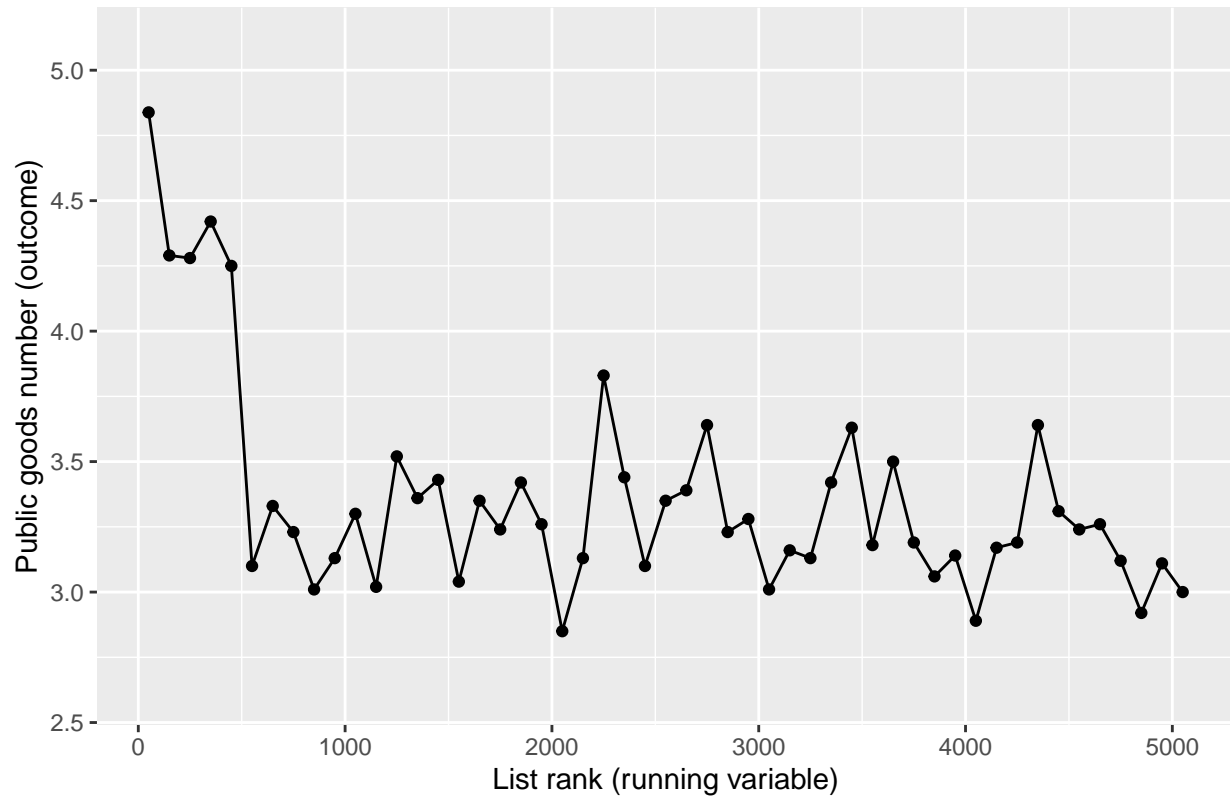


Plot the relationship between the town's position on the list and public goods provision. Describe what you're plotting, using a definition from the course.

- This is "Reduced form" plot with relationship between outcome by running variable is visualized.

```
## No summary function supplied, defaulting to 'mean_se()'
## No summary function supplied, defaulting to 'mean_se()'
```

### Reduced form: List rank and public goods provision



Informed by these plots, write down your preferred regression equation(s) for estimating the causal effect of female leadership on public goods provision.

- Fuzzy RD using the first stage and reduced form,
  - First stage:  $D_i = \alpha + \gamma \mathbf{1}[X_i \geq c] + \nu_i$
  - Reduced form:  $Y_i = \alpha + \theta \mathbf{1}[X_i \geq c] + \varepsilon_i$
  - $\hat{\tau}^{FRD} = \frac{\hat{\theta}}{\hat{\gamma}}$

Defend your choice of bandwidth and any functional form choices you make.

- Bandwidth: 499. In order to choose bandwidth, I've used popular method by implementing rdrobust package, using rdbwselect.
  - rdbwselect implements bandwidth selectors for local polynomial Regression Discontinuity (RD) point estimators and inference procedures developed in Calonico, Cattaneo and Titiunik (2014a), rdbwselect.
  - rdrobust implements local polynomial Regression Discontinuity (RD) point estimators with robust bias-corrected confidence intervals and inference procedures developed in Calonico, Cattaneo and Titiunik (2014a), Calonico, Cattaneo and Farrell (2018), Calonico, Cattaneo, Farrell and Titiunik (2019), and Calonico, Cattaneo and Farrell (2020).
  - Resource: <https://cran.r-project.org/web/packages/rdrobust/rdrobust.pdf>
- For functional form choice, comparison between end point, I checked that the slope between the threshold was not very different in F.

- So, I avoided using anything with a non-linear term because endpoints have an outsized impact on polynomials, being pulled up by the high/low points in the data.

## II.H

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

- Estimating the causal effect of female leadership on public goods provision by first stage regression and reduced form:
  - I find that public goods provision is 0.18762488 higher for female leadership town than male leadership town.
  - But this is only local effect around the threshold. This may not be generalized to the entire towns.
- Therefore, I suggest FUNACRONYM to expand female leadership to all towns, in order to improve well-being (measured by public goods number).