Regression Discontinuity Approach

A before/after study design

What is it?

- Used when you are interested in the causal outcome of an intervention/change in policy.

- Technique applied to non-experimental observational data where treatment is observed whether an observed 'assignment' variable exceeds a known cut-off.

  - Example: Did a change in the state pension age from 60 to 65 for women impact on health outcomes.

- The effects of the treatment may be heterogeneous between individuals

Types of RD

- Sharp RD is when all eligible individuals for the treatment take the treatment

- Fuzzy RD is when **NOT** all eligible individuals take the treatment but there is an added incentive for them to do

What type of RD design would investigating the change in state pension age on women's health outcomes be?

A) Sharp RD

B) Fuzzy RD

0   0

A   B
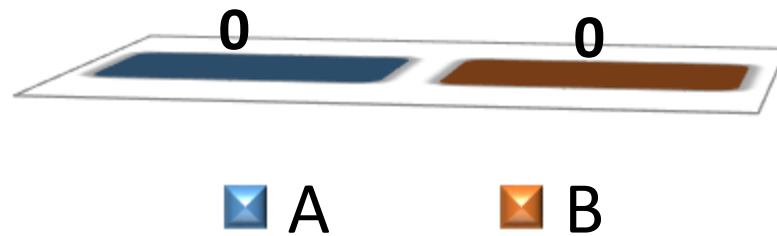
Fuzzy RD approach

- Some individuals may choose not to retire and carry on working.

- These non-compliers can be represented by:
$$\Pr(D_i = 1) = p(V_i)$$

- The probability of retirement V_i is a continuous function except at the point at V_0 which is the age where women can receive their full pension.

$MH_i = f(V_i) + \beta V_i + \epsilon_i$ (structural equation)

$MH_i = f(V_i) + \pi_2(V_i \geq V_0) + \xi_{2i}$ (reduced form)

$MH_i = g(V_i) + \pi_1(V_i \geq V_0) + \xi_{1i}$ (first stage)

Fuzzy RD approach continued

- The cut-off indicates a change in probability of being affected by a change in the retirement age.

- Some women may be unaffected by the treatment, requiring re-scaling to account for this. Making it a standard IV model:

$$\beta = \frac{\pi_2}{\pi_1}$$

- A complier is a woman who retires at 60 if she has the option and will not take it if it is not an option.
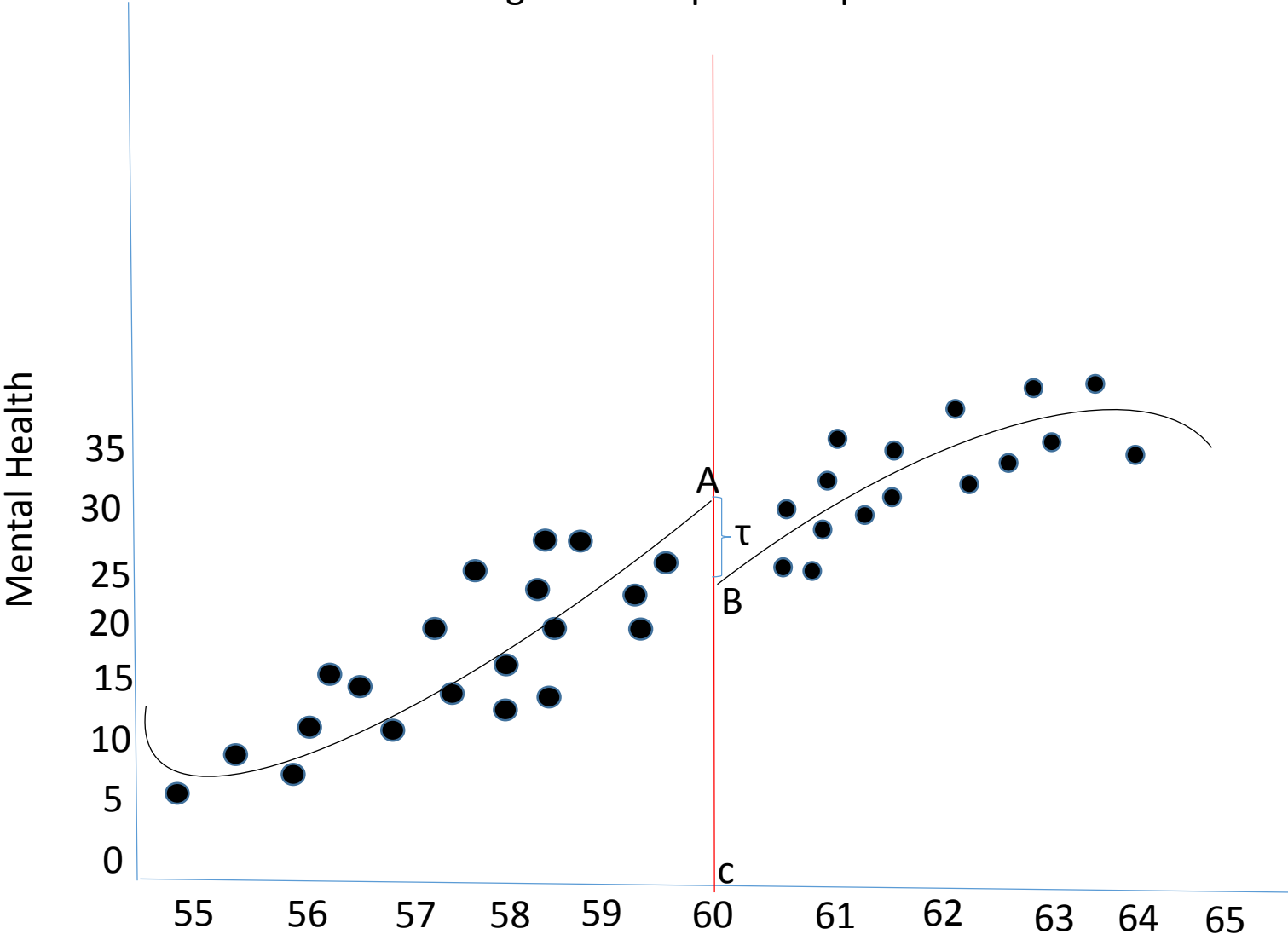
Figure 1: Assignment Probabilities of Women Retiring
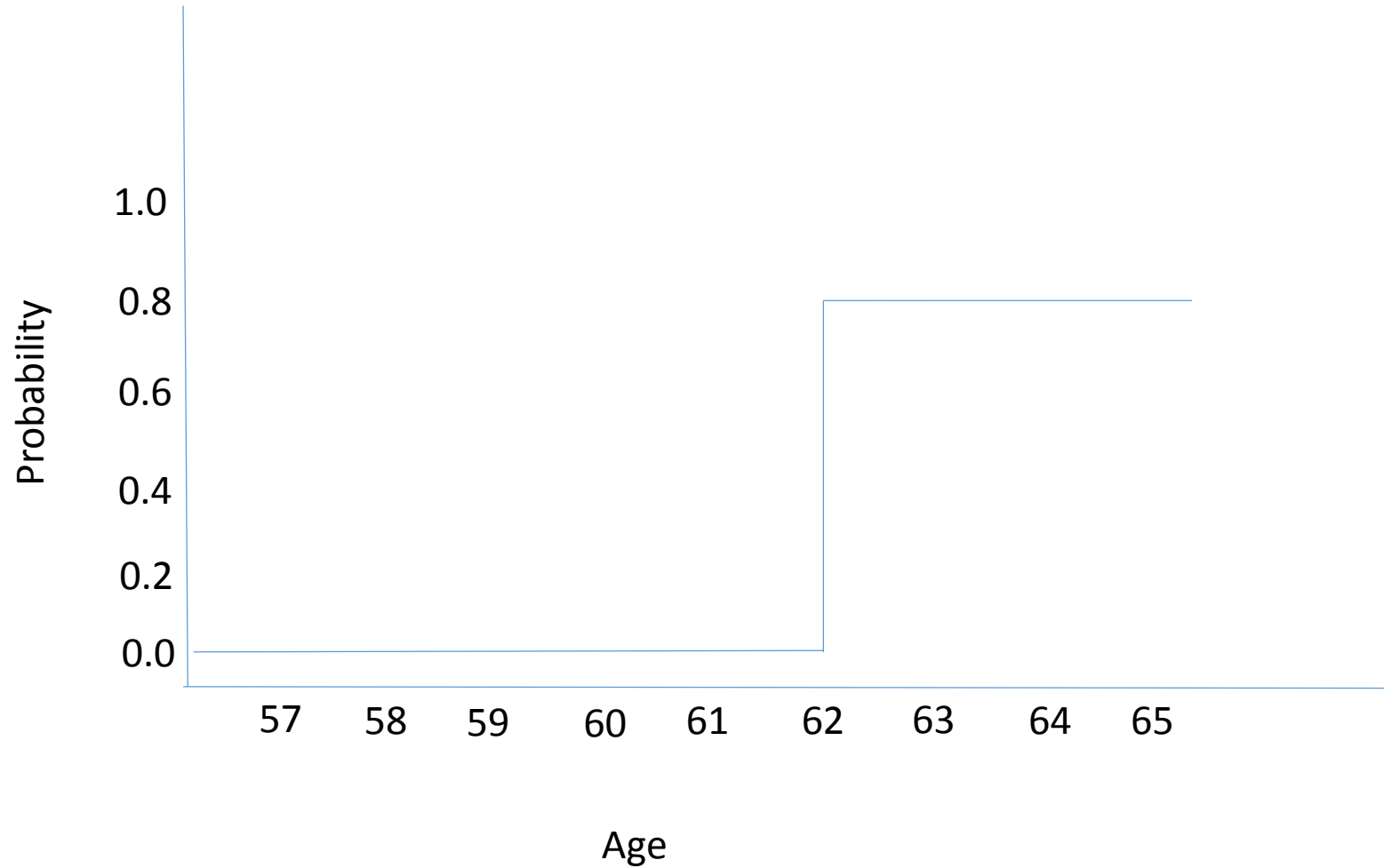
Deciding on an outcome measure

- How will you measure health?

- What are you capturing?

- Why may this be important?

Figure 1: Graphical Representation of RD data

Relationship between health & retirement

- To estimate our model we need to be sure that health does not change around the retirement age.

- Then we can say that 'B' is a reasonable estimate of an individual's mental health given they were exposed to a change in the expected retirement age.

- 'A' would then be a reasonable guess of an individual's mental health if they were able to retire at the earlier expected age.

- From this we can say that the causal effect of changing the retirement age on mental health is B-A.

Applying in practice

- There are limits on how close you can get to 'c' with real data.

- To get a reasonable guess of the treated and untreated states at $X=c$ the only option is to use data away from the discontinuity (i.e. away from individuals that are 60).

- Presenting the results graphically will provide you with some evidence if there is a 'jump' at the cut-point and if the research question is worth pursuing using this methodological approach.

How to create figure 1

- Divide age into a number of equal size intervals which are often referred to as 'bins'.

- Start defining the bins at the cut-point (age 60) and work your way to the right and left to ensure that no bins straddle the cut-points.

- To decide the appropriate bandwidths, it is usually a process of trial and error by trying a few and identifying the ones which look the most sensible.

- The ideal bandwidth are narrow enough so that the patterns in the data are visible especially around the cut-point but are wide enough so that the noise in the data does not over power the signal.

How to create figure 1 (continued)

- Calculate the average value of the outcome measure (mental health score) and the midpoint value of the rating variable (age) for each bin and count the number of observations in each bin.

- Plot the average outcome variable for each bin on the Y-axis against the midpoint rating variable for each bin on the X-axis. The number of observations in each bin can be used as the weight so that the size of the plot reflects the number of data points in each bin.

- To help readers better visualise the data you can superimpose flexible regression lines such as by using the STATA command: *lowess* on top of the plotted data

Steps for estimation

- Decide if the RD approach is sensible for your research question

- Assess whether a sharp or fuzzy design is more appropriate. What is the probability of receiving treatment/being affected by the policy change.

- Conduct preliminary graphical and statistical analysis as described above (e.g construct a figure and organise the data).

- Unless evidence suggests otherwise use the simplest model when conducting your analysis and more complicated models for robustness checks.

Generalisability of Results

- The literature in this area argues that the comparison of the mean outcome for participants in the intervention and control group **ONLY** identifies the mean impact of the intervention at values close to the cut-point but not across the age spectrum.

- Because of the heterogeneous effects of the intervention/change in policy this effect near the cut-point may be very different to the impact away from the cut-point.

Group Exercise

- Can you think of any examples of when you would use a Sharp RD design?