Problem Set 3

This assignment applies the RD ideas discussed in lectures to Clark and Royer (2013), henceforth CR.

Background and data

In April 1947, the British government raised the minimum school-leaving age from 14 to 15. People born before April 1933 could leave at 14, whereas those born in April 1933 or later had to remain until 15. Not everyone would otherwise have left at 14, so the reform did not increase everyone's schooling by one year. Education is measured by age left full-time education, a multivalued treatment. Increasing this age by one year is interpreted as an additional year of education.

The data extract on the course website (`ps3_rd.dta`) combines Health Survey of England (HSE) and General Household Survey (GHS) data. It contains 81,415 UK-born respondents from birth cohorts within ten years of the threshold. The variables are:

- `r`: birth month relative to April 1933 ($r = -1$ is March 1933; $r = 0$ is April 1933).
- `educend`: age left full-time education; `smoke_ever`: ever smoked (0/1).
- `sex`: male (1) or female (0); `ghs`: GHS (1) or HSE (0).
- `yintb`, `mintb`: interview year and month; `mob`: calendar month of birth.
- `age`: age in months at interview.

Let $D_i \equiv$ `educend`$_i$, $Y_i \equiv$ `smoke_ever`$_i$, and $Z_i \equiv 1(r_i \geq 0)$. Here Z denotes exposure to the new minimum leaving age, not educational attainment. Create the following variables:

```
use ps3_rd.dta, clear
assert _N == 81415
gen byte z = (r>=0) if !missing(r)
gen double zr = z*r
```

```
gen double r2 = r*r
gen double zr2 = z*r2
gen double r3 = r2*r
gen double zr3 = z*r3
gen double age2 = age*age/100
gen double age3 = age2*age/100
```

We rescale age2 and age3 to make estimates easier to interpret.

Important!

- Use heteroskedasticity-robust standard errors throughout.
- Report probability effects as percentage points, not percent changes.
- Note that an h-month observation window is defined as $-h \leq r \leq h$.

Q1: OLS estimates

(a) Follow CR and restrict the sample to $-79 \leq r < 0$ and $D \in \{14, 15\}$. Regress ever smoking on education and an intercept. Interpret the schooling coefficient. Why is this a useful baseline estimate for thinking about the causal effect of education induced by the reform?

(b) Re-estimate the OLS relationship controlling for sex, a third-order polynomial in age, interview-year and interview-month indicators, the GHS indicator, and calendar-month-of-birth indicators. Compare the estimate with Q1(a) and CR's OLS result in panel A of Table 5a (OLS column for "ever smoke"). By how much does adding these controls change the estimated relationship between education and smoking?

(c) Describe two unobserved characteristics that could bias the OLS estimate in Q1(b). For each characteristic, explain how it might be related to education and smoking, and how you would expect the estimate to change if you could control for it.

Q2: Assessing RD validity

(a) Can you think of any reasons why the key continuity assumption underlying RD might fail in this context? That is, why might individuals born in March 1933 differ systematically from those born in April 1933 in ways that affect their smoking behavior independently of their education? Describe at least one scenario that could violate this assumption, even if you do not find it to be plausible. [Hint: Here's an example scenario (that you are not allowed to use!). Suppose that a sudden change in medical practices in April 1933 affected the survival chances of babies born from that month onward. This could change the composition of surviving individuals born on either side of the threshold, potentially creating systematic differences in their later smoking behavior.]

(b) One way to assess RD validity is to examine whether the density of the running variable changes discontinuously at the threshold. Another way is to examine whether characteristics determined before the reform change discontinuously at the threshold. Explain how CR do these checks and what they find. [Hint: See Appendix Figure A1 of CR, available at `https://www.aeaweb.org/articles/materials/2394`.]

(c) The data used in this exercise come from surveys conducted many decades after the individuals were born. Explain why this could generate a sample-selection (i.e., attrition) problem, even if the continuity assumption held in the original population.

(d) Plot a one-month-bin histogram of respondents' birth cohorts for $-60 \leq r \leq 60$, marking $r = 0$. Does this provide any evidence of differential attrition? Next plot the fraction male by birth cohort over the same range. On the microdata, estimate

$$\texttt{sex}_i = a + \delta Z_i + br_i + cZ_i r_i + v_i$$

using $-46 \leq r \leq 46$. Report the jump and its robust SE. Do the graph or regression estimate provide any evidence of differential attrition?

Q3: First-stage estimates

(a) Plot average age left education by birth month using the full sample, marking the cutoff (i.e., similar to the graph in the top-right of Figure 2 of CR). Describe the pattern you observe. Is there evidence of a discontinuity at the cutoff?

(b) Using the individual data (46-month window and individuals with nonmissing education), regress D on $1, Z, r, Zr$. Interpret the estimated discontinuity in terms of years of education. Repeat this regression after adding sex, interview-year and interview-month indicators, the GHS indicator, calendar-month-of-birth indicators, and a third-order age polynomial. Report the result and check it is close to the estimate reported by CR in Table 2 column 1 (top row).

(c) Using the same 46-month sample and covariates as in (b), compare the covariate-adjusted interacted-linear estimate with the following alternatives:

- A common quadratic, D on $1, Z, r, r^2$ and the same covariates
- An interacted quadratic, adding Zr and Zr^2 and retaining the same covariates

Explain the restrictions imposed by each functional form and assess the robustness of the estimated discontinuity.

(d) Summarize the evidence on whether exposure to the reform affected education.

Q4: Reduced-form estimates

In CR, the preferred windows differ by outcome: 79 months for smoking and 46 months for schooling. For this question, use an observation window of 79 months and continue to exclude observations with missing education information.

(a) Graph average ever-smoking rates by birth month using the full sample. Describe the pattern you observe. Is there evidence of a discontinuity at the cutoff?

(b) Use individual data to estimate the linear-interacted reduced form of Y on $1, Z, r, Zr$. Repeat using the same 79-month sample after adding sex, interview-year and interview-month indicators, GHS, calendar-month-of-birth indicators, and a third-order age polynomial. Check that your covariate-adjusted estimate and standard error are close to those reported by CR (Table 5a, RF column). Interpret both estimated discontinuities and robust SEs. Compare their statistical significance at the 5% significance level.

(c) Using the same 79-month sample and covariates as in (b), compare the covariate-adjusted interacted-linear estimate with the following alternatives:

- A common quadratic, Y on $1, Z, r, r^2$ and the same covariates
- An interacted quadratic, adding Zr and Zr^2 and retaining the same covariates

Compare the estimated discontinuities and robust SEs across the three specifications.

(d) Summarize the evidence on whether exposure to the reform affected smoking.

Q5: Fuzzy RD and instrumental variables

Throughout this question, use a 46-month window and restrict to observations with nonmissing education and smoking. Use the interacted-linear specification, including r and Zr, and the same covariates as the adjusted specifications in Q3 and Q4: sex, interview-year and interview-month indicators, the GHS indicator, calendar-month-of-birth indicators, and a third-order polynomial in age. Restrict to observations with nonmissing values of all required variables, and use exactly this common sample and the same covariates in the first-stage, reduced-form, and IV regressions.

(a) Estimate the education first stage and the smoking reduced form on this sample. Report the two discontinuities and calculate their ratio.

(b) Estimate the model by instrumental variables, instrumenting education with the cutoff indicator and including r, Zr, and the same covariates as in part (a), on the same common sample. Report the estimate, a robust SE and 95% confidence interval. Verify the estimate equals the ratio reported in part (a).

(c) Compare your IV estimate with the OLS estimates from Q1. What do you conclude about the effect of education on smoking?

References

Clark, D. and Royer, H. (2013). The effect of education on adult mortality and health: Evidence from Britain. *American Economic Review*, 103(6):2087–2120.