ECON 8000/9000 Empirical Energy Economics
PS3: Replication Exercise: RD and RDiT
Christy Zhou

Instruction and Notes:

- Deadline: Sunday 2/8 @ 5pm
- Submission: Please compile all results into a single PDF.
- Group: Everyone can work together, but only **up to two students** can submit the assignment together. If so, make sure you include both names at the top when you submit.
- Here I write instructions using Stata. You can use other software if you like.
- This study has a replication package online. I recommend you follow my instructions below, which are slightly less complicated than the online package.

Q1. Additional Exercise for Englmaier et al. (2024) "Price Discontinuity in Online Market for Used Cars" Mgm Science
Q1.A Manipulation Test
Last time in PS2, we replicated Table 3 Column 1. We covered in class that this specification is parameterized in the same way as in an RD design, but the estimation strategy is likely not a true RD design due to potential bias around the cutoff. Here we will test if such a manipulation exists.

Again, same info:

- Dataset: Same as in PS2
- Key running variable: mileage

Tasks:

- 1. Run the manipulation test for the cutoff 10000 km using rddensity. To test this cutoff, limit observations to vehicles with mileages +/- 5000 from the cutoff, as in the commands above. Save the test figure.
  - ```
    preserve
        keep if inrange(mileage, 5000, 15000)
        rddensity mileage, c(10000) plot nlocalmin(10) all
    restore
    ```
  - Note: rddensity is available in R and Python as well. You don't have to do this assignment in Stata. More on rddensity: https://rdpackages.github.io/rddensity/
- 2. Redo exercise #1 for the cutoffs 10000, 20000, …, and 8000 in a loop. Save these figures in your loop.
- 3. To understand the test results, do the following:
  - (a) For the visual output, study if the left CI overlaps with the right CI on the figures generated.
  - (b) For the test statistics, study (i) the effective bandwidth on the left & right, (ii) the effective sample size on the left & right, (iii) the manipulation test t-stat and p-value, also reported as e(T_q) and e(pv_q).
- 4. Interpret the test results:
  - Does the RD regression pass the manipulation test?
  - Does Figure 3 foreshadow these results?
  - What do the manipulation results tell you about how you can interpret Table 3 that you did in PS2?

Q2. Replication Exercise for Smith (2016) "Spring Forward at Your Own Risk: Daylight Saving Time and Fetal Vehicle Crashes", AEJ:AE
In this exercise, we will reproduce the main RDiT regressions in Smith (2016), as well as additional tests and visualization exercises. We will also reproduce the FE estimation.
    ```

TABLE 1—RD ESTIMATES OF THE IMPACT OF ENTERING DST ON FATAL CRASHES (*Spring*)

	(1)	(2)	(3)	2002–2006 (4)	2007–2011 (5)	Placebo (6)
DST	0.0649***	0.0499***	0.0626***	0.0941***	0.0375	0.000536
	(0.0231)	(0.0176)	(0.0215)	(0.0302)	(0.0361)	(0.0225)
Bandwidth selector	CCT	IK	CV	CCT	CCT	CCT
Observations	550	966	670	235	265	550

TABLE 2—RD ESTIMATES OF THE IMPACT OF LEAVING DST ON FATAL CRASHES (*Fall*)

	(1)	(2)	(3)	2002–2006 (4)	2007–2011 (5)	Placebo (6)
Leaving DST	0.00114	−0.000182	0.00630	0.0274	−0.00260	0.0361*
	(0.0236)	(0.0153)	(0.0242)	(0.0265)	(0.0327)	(0.0218)
Bandwidth selector	CCT	IK	CV	CCT	CCT	CCT
Observations	381	850	347	215	225	381

Info:

- Dataset: "PSdata_Smith_AEJ_main.dta"
- Key vars:
  - outcomes: ln_fatals (main outcome), ln_vmt
  - date variables: year, month, day
  - main running variable using relative dates: days_from_spr_tran, days_from_fa_tran
  - additional vars: sp_tran_date, sp_tran_date
  - vars for FE regressions: sp_dst, first_six, second_8, rest_sp, remainder_fa_dst

Tasks:

- For tasks 1-4, you can go over yourself w/o reporting back to me.
- 1. Understand Stata format (You can skip this if you use another program)
  - Run codes such as the following to understand the date formats
    format fa_tran_date %tdCCYY-NN-DD
    format sp_tran_date %tdCCYY-NN-DD
    format fa_tran_date %10.4g
    format sp_tran_date %10.4g
    format fa_tran_date %d
    format sp_tran_date %d
- 2. Create some date-specific fixed effects:
  - gen date = mdy(month, day, year)
    gen day_of_year = doy(date)
    gen day_of_week = dow(date)
  - You can refer to all date functions (link) in case you'd like to in the future for your own work.
- 3. Study how relative dates are created
  - Here we have supplied days_from_spr_tran and days_drom_fa_tran
  - Run di dmy(1,1,2002)-dmy(7,4,2002). Then see how this compares to the first observation's sp_tran_date
    - This basically is how these dates are created.
    - You can gen a new var by gen my_days_from_spr_trans = date – sp_tran_date just to see this.
    - Realistically, the relative date is compared to the nearest transition, so the relative date of 2002-1-1 to Fall DST is actually compared to fall 2001 not fall 2002.
- 4. Generate outcome var for the RD regression.
  - The author takes out some seasonality by demeaning the outcome on year FE and day-of-week FE.
  - Run reghdfe ln_fatals, abs(i.year i.day_of_week) vce(robust) res

- Use the predict command to generate residuals as the demeaned log(fatals), called demeaned_ln_fatals.
- Do the same exercise for ln_vmt and generate demeaned_ln_vmt

- 5. Re-produce Table 1 columns 1-3 using rdrobust
  - We will produce 3 versions of bandwidth selection. They will not be exactly the same as Smith (2016) but should be pretty close. The reason that the number will be a bit different is that Cattaneo and co-authors made new updates on the algorithms of bandwidth selection.
  - For all versions, the basecode will be rdrobust demeaned_ln_fatals days_from_spr_tran, kernel(uni)
  - For the 1st version, add option h(27 27), which is the baseline bandwidth as Smith (2016)
  - For the 2nd version, add option bwselect(mserd)
  - For the 3rd version, add option bwselect(cerrd)
  - Report all 3 results together in one table.
    - Take a look at the effective observations from the left and the right-hand side of the cutoff. These are equivalent to the observations reported in Table 1. Report them in your table as well.
    - Take a look at the bandwidth on the left & right-hand side. Report them as well.
    - In terms of main estimates, you will find versions 1 and 3 quite close.
    - Summarize what you learn from your table.
  - Note: In the original version, author effectly ran bwselect(cct), bwselect(ik), and bwselect(cv). The former two have now been updated to the MSE-RD and CER-RD versions of bandwidth selection. People don't use cross-validation much now.
- 6. Re-produce Table 2 columns 1-3
  - Aka, redo task #5 above using days_from_fa_tran.
  - Summarize what you learn from your table. Also, compare the results in task #5 vs #6.
- 7. Re-run Table 1 using an alternative outcome ln(VMT), aka ln_vmt
  - Given your result, do we have any evidence that individuals adjust VMT before and after Spring DST that may bias your results in Table 1?
- 9. Addition RD-plot
  - In this exercise, we will produce an RD-plot that is consistent with Table 1 and another for Table 2. Here I plot +/- 50 days, the 7 bins on each side mean I create roughly one CI for a week.
  - For spring transition, run rdplot demeaned_ln_fatals days_from_spr_tran if inrange(days_from_spr_tran, -50, 50), c(0) kernel(uni) h(50 50) ci(95) nbins(7 7) p(6)
  - Run the same for the fall transition, days_from_fa_tran
  - Save these two figures. Examine if CIs overlap more or less around the cutoff. Explain what the plot informs you.
- 8. Understand why the density manipulation test doesn't work.
  - In Q1.A, we did a manipulation test.
  - Here we have RDiT and this test doesn't work anymore. Redo the same test using something like this:
  - keep if inrange(days_from_spr_tran, -50, 50)
  - rddensity days_from_spr_tran, c(0) plot nlocalmin(10) all
  - Save the figure and report it. What does your figure show? Does it surprise you?
- 9. Now we will move on to the FE regressions in Table 5
  - We will re-run columns 1 and 3.
  - For column 1, the RHS includes sp_dst remainder_fa_dst
  - For column 3, the RHS includes first_six second_8 rest_sp remainder_fa_dst
  - Run both with reghdfe. You should absorb 3 sets of FE, including year FE, day of the week FE, and day of the year FE.
  - What do your results show and compare to the RD results?
- 10. How that you have seen how RDiT works. Can you think of a research question that you can use an RDiT design?
  - It does not need to be limited to an energy-related topic, although that will be encouraged.
  - If RDiT is tricky, you are welcome to think of a regular RD design as well.