Problem Set: Regression Discontinuity

BUSS975: Causal Inference in Financial Research

Problem 1 (What RD assumes, and what it gives you)

(a) Write down the estimand for a **sharp** design and for a **fuzzy** design. The fuzzy estimand is a ratio — say what sits in the numerator and what sits in the denominator, and explain why the sharp case is the special case where the denominator equals one.

(b) Every other design in this course has relied on **common support**: for a given covariate value we need both treated and untreated units. RD has *none* — at any X you observe only one treatment status. Explain why the design is nonetheless identified, and name precisely what replaces common support.

(c) The estimand is a **limit**. Explain why τ is not "the effect of the treatment," and give one example in which a policymaker's question and the RD estimand come apart.

(d) On the slide titled **"Simulation"** (§6.3, just before the coverage experiment), the true conditional expectations are **quadratic** on each side of the cutoff,

$$E[Y^0 \mid X = x] = 0.40 + 0.30\,x + 6.0\,x^2, \qquad E[Y^1 \mid X = x] = 0.60 + 0.80\,x - 4.0\,x^2,$$

so that $\tau = 0.20$ — while `rdrobust` fits a local **linear** polynomial by default. A student concludes that the design is misspecified and the estimate invalid. Evaluate this. State exactly what the continuity assumption requires, and say what the polynomial order p is doing if it is not modelling the truth.

(e) RD buys credibility and pays for it. State what you give up relative to a design that assumes unconfoundedness on covariates, and say when you would nonetheless prefer unconfoundedness.

Problem 2 (Choosing p, h and K)

(a) Take the approximate mean squared error of the RD estimator at bandwidth h,

$$MSE(h) \;\approx\; B^2 h^{2(p+1)} \;+\; \frac{V}{nh}.$$

Minimise with respect to h and show that

$$h_{MSE} \;\propto\; V^{\frac{1}{2p+3}}\; B^{-\frac{2}{2p+3}}\; n^{-\frac{1}{2p+3}}.$$

(b) The formula has three ingredients: the noise V, the curvature B, and the sample size n. Taking each in turn and holding the other two fixed, say whether raising it makes the optimal window **wider or narrower**, and give the intuition in one sentence.

(c) Unweighted OLS on a window is a **rectangular kernel**. Explain the boundary problem it creates — why a bin mean is the wrong estimate of the conditional mean *at the cutoff* — and how a triangular kernel helps.

(d) A referee asks you to "control for the running variable more flexibly" by adding a global cubic and quartic. Give two separate reasons to refuse, and say what you would do instead.

Problem 3 (Inference: three intervals, and one bad habit)

(a) **Computational.** Generate data from the design used in lecture:

$$X_i \sim U(-1,1), \qquad Y_i = m_{D_i}(X_i) + \varepsilon_i, \qquad \varepsilon_i \sim N(0, 0.25^2), \qquad D_i = \mathbf{1}\{X_i \geq 0\},$$

$$m_0(x) = 0.40 + 0.30x + 6.0x^2, \qquad m_1(x) = 0.60 + 0.80x - 4.0x^2,$$

so the true effect at the cutoff is $\tau = 0.20$. Take $n = 2{,}000$. For each of $h = 0.10$ and $h = 0.40$, draw 500 samples; in each, fit `rdrobust` with `p = 1`, a triangular kernel and that **fixed** h, and record whether each of the three intervals — conventional, bias-corrected, robust bias-corrected — contains 0.20.

(b) Conventional coverage is close to nominal at $h = 0.10$ and catastrophic at $h = 0.40$. Explain this using the two terms B and V.

(c) Bias-corrected coverage is roughly flat, and roughly 80%, at **both** bandwidths. Recentring fixed the location, so why is the interval still wrong? What does the robust version add?

(d) Lee and Card (2008) observed that with a **multi-valued** (discrete) running variable the local polynomial cannot pass through the conditional mean at every mass point, and that every observation sharing a mass point therefore shares the same specification error. State in one line why this is the structure of a group shock, and what they concluded you should cluster on.

(e) **Computational.** Using `data/hansen_dwi.csv`, estimate Hansen's recidivism specification

$$Y_i = \alpha + \tau\,\mathrm{DUI}_i + \beta_1 \mathrm{BAC}_i + \beta_2 \mathrm{BAC}_i \cdot \mathrm{DUI}_i + \varepsilon_i$$

on windows $|\mathrm{BAC} - 0.08| \leq h$ for $h \in \{0.05, 0.02, 0.01\}$. For each, report N, the number of distinct BAC readings G, the heteroskedasticity-robust standard error on DUI, and the standard error clustered on the distinct BAC reading. **Before you run it, predict the direction of the change.**

(f) You should find that clustering makes the standard error *smaller*, by more as the window narrows. Give the **two** separate things going wrong, and explain why the second means clustering here is not merely imprecise but conceptually wrong.

Problem 4 (Manipulation, heaping, and what a test can clear)

(a) Continuity fails in five ways. Conditions 1–3 — a known rule, an interest in adjusting, and the opportunity to adjust — are together the recipe for sorting. For each setting below, say which of the three hold and whether you expect sorting:

i. A driver blowing into a police breathalyzer at the roadside.
ii. A firm reporting earnings per share relative to the analyst consensus.
iii. A student sitting a national university entrance exam with a published admission cutoff.
iv. A bank reporting total assets relative to a regulatory threshold that triggers supervision.

(b) State precisely what the McCrary / `rddensity` test tests. Then consider a running variable that is **heaped**: a large pile of observations is recorded at *exactly* the cutoff c_0 because the variable is reported coarsely — rounded to a round number — and **not** because anyone changed sides. No unit that truly belonged below c_0 has been shifted above it, or the reverse; the pile is an artefact of how the variable is recorded, not of sorting. Does the density test detect this? Answer at two levels — what happens to the **estimand** (the two one-sided limits), and what happens in any **implementation** that has to put that pile on one side of c_0. Say which of the two a researcher should worry about.

(c) In Almond et al.'s setting, babies recorded at exactly 1,500g are systematically frailer than their neighbours, and 1,500g is the VLBW threshold below which babies get extra care. Explain why this manufactures an apparent *benefit* of being below the cutoff, and why the same artefact is waiting at 1,400g and 1,600g.

(d) Why is the argument in (c) a statement about **composition** rather than about treatment? A colleague says "fine, but the effect is still real for the babies at 1,500g." Respond.

(e) The donut-hole RD drops the suspect observations and re-estimates. State the justification, and then state precisely what happens to the **estimand**. Why is "the estimate halved when we donut" not by itself evidence that the original estimate was wrong?

Problem 5 (RD in finance: how is each study performed?)

The lecture slide **"Where the cutoffs are"** lists four studies:

Paper	Running variable	Cutoff	Design
Keys, Mukherjee, Seru and Vig (2010)	FICO credit score	620	sharp-ish
Malenko and Shen (2016)	ISS voting-policy score	policy threshold	sharp
Almeida, Fos and Kronlund (2016)	EPS relative to consensus	0	sharp
Cuñat, Giné and Guadalupe (2012)	shareholder vote share	50%	fuzzy

(a) For **each** of the four papers, answer the following.

i. **The research question.** What causal effect does the paper want, and why is it hard to get without an RD? Name the specific confounder or reverse-causality problem the design is solving.

ii. **The rule.** What exactly happens at the threshold, and **who sets the rule** — a statute, a regulator, a private firm's published policy, or an industry convention? Say how the agents being studied learn about the rule.

iii. **The design.** Running variable and cutoff; sharp or fuzzy, and *why*; what is being compared with what; and what the first stage is (if the design is fuzzy, or if treatment is assignment to something other than the outcome of interest).

(b) **Audit the table.** The "Design" column above labels three of the four **sharp**. For each paper, ask the question that actually settles it: *does the probability of treatment jump to one at the cutoff?* Where it does not, report the size of the first stage as the paper estimates it. Then write the corrected column. One of the four does not report a first stage at all — identify it, and explain why that omission is not a minor gap but the hinge on which the paper's main claim turns.

(c) In (a.ii) you will have found that the four rules differ sharply in **who sets them**. Explain why that matters for the credibility of the design, and rank the four from most to least credible on the no-manipulation assumption.

(d) One of the four papers is built on the fact that firms **do** manipulate their position relative to the threshold — that is the paper's finding, not a nuisance. Identify it, and explain the trick: how is the running variable constructed so that the design survives the very manipulation it is documenting? State the assumption that this construction requires, and one way it could fail.

(e) RD estimates come from a neighbourhood, not a sample. Hansen had 214,558 stops. A typical close-vote governance study has a few hundred proposals in total, perhaps forty within a few points of 50%. Explain what this does to the power of the design, and state what an underpowered RD looks like in a published table — and why that pattern is dangerous rather than merely disappointing.