

Friday Exam 2

1. a. Our colleague suggests comparing outcomes for those who did not go to mental health court, or

$$E[Y'|D=1] - E[Y^0|D=0],$$

which is just the SDO. This may be decomposed into

$$SDO = ATE + E[Y^0|D=1] - E[Y^0|D=0] + (1-\pi)(ATT-ATU),$$

It now becomes clear that for this to be a causal effect, we need the biases to zero out. We achieve this by assigning treatments independent of their potential outcomes ("randomize"), i.e.

$$E[Y^0|D=1] = E[Y^0|D=0] \text{ and}$$

$$E[Y'|D=1] = E[Y'|D=0]$$

If this is the case, then selection bias ($E[Y^0|D=1] - E[Y^0|D=0]$) becomes zero. Note ATT-ATU (Heterogeneous Treatment Effect Bias) is:

$$E[Y'|D=1] - E[Y^0|D=1] - (E[Y'|D=0] - E[Y^0|D=0])$$

which goes to zero by plugging in our "randomization" assumptions.

In that case,

$$SDO = ATE$$

In context of the problem, this method of comparing recidivism for those who did not go through MHC requires that assignment to MHC is independent of likelihood of committing another crime post-punishment. If this is true, then our colleague's method will produce the average treatment effect of assignment to the MHC on recidivism, which is causal under that assumption.

- b. I believe my colleague is wrong. Comparing the two groups would not be causal because the groups of convicts (those above 60 and those below) are likely very different on both observable and unobservable factors. It is likely that those above 60 are mentally ill, which directly effects recidivism independent of the courts; thus,

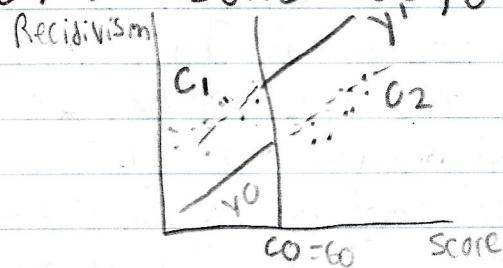
$$E[Y^0|D=1] \neq E[Y^0|D=0] \text{ and}$$

$$E[Y'|D=1] \neq E[Y'|D=0]$$

Mental illness effects both the score and probability of

Recidivism, so it constitutes a roundabout way of selecting observables — thus, treatment and outcomes are not independent, selection bias is present, and our colleague's strategy does not yield a causal effect.

C. The running variable for this project is the mental health score assigned by the therapist, with the cutoff being a score of 60. The counterfactual for an inmate in MHC is someone below the score of 60, but still in MHC, i.e.



That is, the counterfactual for someone in MHC (y^1) is some theoretical person who follows the curve through the cutoff. Similarly, the counterfactual for someone not in MHC is someone with a score above 60, but not in MHC (c_2). Note that it is best to compare observations and counterfactuals with close scores, i.e. 59 vs 61.

d. The identifying assumptions in RDD are the continuity assumptions, i.e. expected outcomes change smoothly as a function of the running variable, even through the cutoff. This assumption means, $E[y_i^0 | x=co]$ and $E[y_i^1 | x=co]$ are continuous in x at co , i.e. both treatment and control are smooth through the cutoff. In other words, y does not jump at x except by the treatment. However, this has several "tests" that can jeopardize identification if true. They are:

1) Assignment rule is known in advance

2) Agents are interested in adjusting

3) Agents have time to adjust

If true, agents can warp their running variable to get their preferred treatment. Thus, falsity of these criteria is a continuation of the identification assumptions.

c. If we fear therapists are manipulating the data, we can use a McFadden density test to confirm or deny our suspicions. First, we partition the assignment variable into bins and find frequencies for each bin. Then, we treat those frequencies as dependent variables in local linear regressions. We then seek to find the density surrounding the distribution. If we fail to reject the null of continuous density at the cutoff, we judge there to be no manipulation. If we reject the null in favor of the alternative that the densities differ on either side of the kink, i.e. we find that more or less observations fall in the bin to the right of the cutoff as compared to the left, and this appears to be a discontinuity in the density, then we are suspicious.

This does not prove manipulation, but it alerts us that something could be amiss.

f. In this context, a balance test would begin by identifying observed covariates that are exogenous to, but associated with potential outcomes (i.e. noncolliders). Then, we treat these variables, Z , as we did our outcomes, y . We create a graph of the running variable against Z , estimate it using parametric or nonparametric means, and anything else we did to examine. We can even draw a confidence interval around our regression lines on either side of the cutoff. Essentially, we aim to determine if these exogenous, noncollider covariates also jump after the cutoff. If there is a jump, smoothness is not supported and

Our data may be unbalanced and unfit to use. If no jump is evident, it supports our assumption of continuity and smoothness.

Q. In order to combat nonlinearities, we have two methods.

First, we have "higher order polynomials". First, suppose that

$$E[Y_i | x_i] = f(x_i)$$

for a smooth $f(x_i)$. Then our regression is

$$Y_i = f(x_i) + \delta D_i + \eta_i$$

We now approximate $f(x_i)$ for $x_i > c_0$ with a p^{th} order polynomial

$$Y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_p x_i^p + \delta D_i + \eta_i$$

Now, let x_i differ on either side of the "cutoff":

$$E[Y_i^0 | x_i] = \alpha + \beta_{01}(x_i - c_0) + \dots + \beta_{0p}(x_i - c_0)^p$$

$$E[Y_i^1 | x_i] = \alpha + \beta_{11}(x_i - c_0) + \dots + \beta_{1p}(x_i - c_0)^p$$

Then we have

$$Y = E[Y_i | x_i] E[Y^0 | x_i] + [E[Y^1 | x_i] - E[Y^0 | x_i]] D_i$$

$$Y = \alpha + \beta_{11}(x_i - c_0) + \dots + \beta_{0p}(x_i - c_0)^p + \delta D_i + (\beta_{11} - \beta_{01})(x_i - c_0) + \dots + (\beta_{1p} - \beta_{0p})(x_i - c_0)^p + \varepsilon_i$$

However, this method leads to overfitting and bias, leading to a rise in popularity of the second method, Nonparametric kernels.

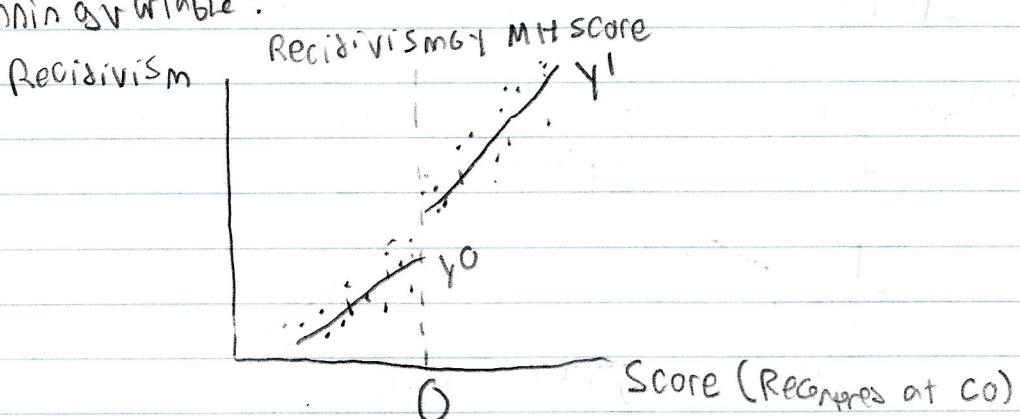
Implementing local linear nonparametric regression reduces the bias.

It does a weighted regression in a fixed window, and the kernel tells us the weights by:

$$(\hat{a}, \hat{b}) \equiv a, b \sum_{i=1}^n (Y_i - a - b(x_i - c_0))^2 K\left(\frac{x_i - c_0}{h}\right) \mathbb{1}(x_i > c_0)$$

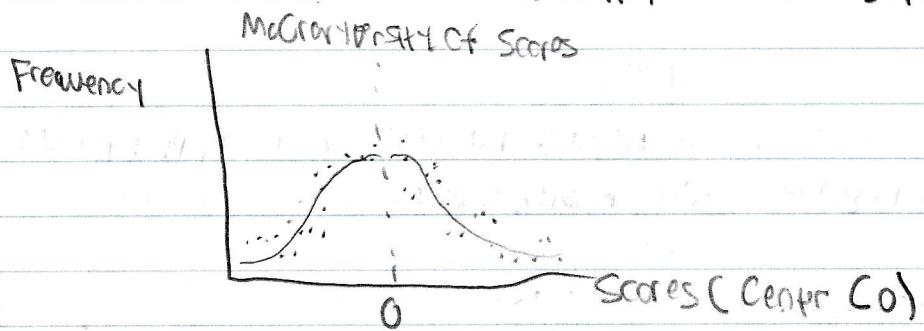
with K a kernel function, h the bandwidth size. This gets around kernel's issues with boundary bias in kernel estimation.

h. The first figure you need for an RDD is outcome by running variable:



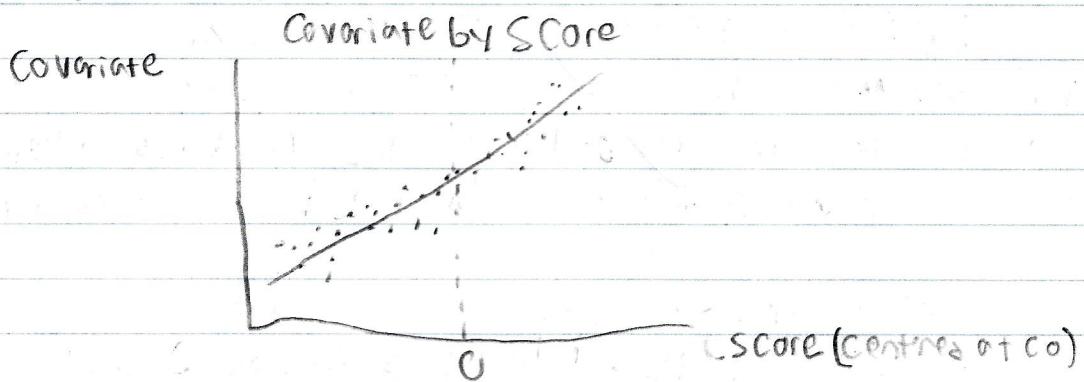
This figure shows that the expected potential outcomes change smoothly, except for a jump at the cutoff C₀ (0 here by recentering). If there is a clear jump at C₀, this is consistent with our hypothesis of continuity except for the ATE at C₀. If there is no jump, this is inconsistent with our hypothesis that the mental health Court effects rates of recidivism. A lack of smooth trends would also be inconsistent with the assumption of continuity.

The second figure required is McCrary density of the running variable:



This figure shows the density of the scores, allowing us to examine the distribution of the scores. If this is smooth and continuous, it supports the validity of the continuity assumptions and gives credence to the idea that this is not subject to manipulation or heaping. If there is a discontinuity at C₀, or if there appear to be many peaks, then this is evidence of rounding or manipulation in the data; it would disprove the assumption that people have imprecise control over the running variable.

The third requisite picture is Covariates by running Variable (for balance tests):



This graph shows the relationship between MH Score and another Noncollider/exogenous Covariate (i.e. Credit Score, etc.)

If the regression line is smooth even through the cutoff, this supports the assumption of smoothness/continuity from earlier.

If there is a JUMP at 0, then this provides evidence that the Potential Outcomes might not be jumping at 0 solely due to the treatment; it makes us doubt the Validity of the Continuity assumption.

2. a. The instrumental variable for this project is the Propensity of the assigned judge to declare inmates ill enough to be seen by the mental health court.

b. The first assumption with heterogeneous treatment effects is the Stable Unit Treatment Value Assumption (SUTVA):

$$I + Z_i = z_i \text{, then } D_i(Z) = D_i(z)$$

$$\text{If } z_i = z_i' \text{ and } D_i = D_i', \text{ then } Y_i(D, Z) = Y_i(D', Z)$$

In Context, this means that the assignment of an inmate to MH (or not) is unrelated to the assignments of other inmates, i.e. one individual's potential outcomes aren't effected by others' treatment statuses.

The Second Assumption is "Random Assignment," AKA the Independence or AS-Grid-AS-Random Assumption:

$$\{y_i(D_i^1, 1), y_i(D_i^0, 0), D_i^1, D_i^0\} \perp\!\!\!\perp Z;$$

In context, this means the propensity to assign MHC-level severity to an inmate is independent of that inmate's recidivism (Potential outcomes) and independent of assignment to an MHC (Potential treatment assignments).

The Third Assumption is the Exclusion Restriction:

$$v(D, Z) = v(D, Z') \text{ for all } Z, Z', \text{ and } D$$

In context, this means that the propensity to assign MHC-level severity of any therapist only effects recidivism through its effect on assignment to MHC, i.e. a different therapist with different "severity" would have no effect on recidivism unless she made a different "treatment" decision regarding assignment to MHC.

The Fourth Assumption is the Nonzero first stage:

$$E[D_i^1 - D_i^0] \neq 0$$

where D^1 means the instrument is "on" and D^0 means it is "off".

In context, this means that the propensity to assign MHC of a therapist MUST have a statistically significant effect on the inmate's assignment to MHC, i.e. a therapist who assigns more people to MHC IS statistically effective on the assignment of an inmate to MHC.

The Fifth Assumption is Monotonicity:

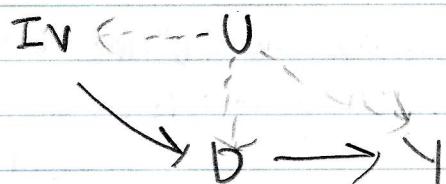
$$\text{Either } \pi_{Ti} \geq 0 \text{ for all } i \text{ or } \pi_{Ti} \leq 0 \text{ for all } i$$

In context, this means that the propensity of a therapist to assign MHC will weakly effect all inmates in the same direction, i.e. some may not be effected, but all those who are will be the same direction (more or less likely to be treated) across all inmates.

C. ONLY ONE ASSUMPTION CAN BE TESTED EMPIRICALLY, SPECIFICALLY THAT OF THE NON-ZERO FIRST STAGE. BECAUSE WE HAVE DATA ON Z AND D, WE CAN MEASURE THE EFFECT OF CHANGING Z ON D TO SEE IF THIS HOLDS. HOWEVER, SUTVA, RANDOM ASSIGNMENT, EXCLUSION RESTRICTION, AND MONOTONICITY ARE NOT VIABLE FOR EMPIRICAL TESTING DUE TO LACK OF DATA (COUNTERFACTUALS) OR A BASIS SOLVING THEORY.

D. SUTVA COULD BE VIOLATED IF ONE PERSON'S ASSIGNMENT TO THE MHC (OR LACK THEREOF) EFFECTED ANOTHER; FOR INSTANCE, LIMITING POSSIBLE ASSIGNMENTS TO MHC TO A SPECIFIC CAP WOULD MEAN EACH ASSIGNMENT LOWERS THE OTHERS' PROBABILITY OF BEING ASSIGNED.

The independence assumption could be violated if the propensity of therapists to assign MHC was not independent of either recidivism or assignment to MHC.



Suppose the unobservable, U, is something like societal attitude towards crime and mental illness. This can affect the propensity to assign MHC as well as reassignment of MHC, and the outcome, recidivism. In this way, U determines IV, Y, and D, so they aren't independent.

The exclusion restriction could be violated if the propensity of a therapist to assign MHC effects recidivism in a way not related to assignment to MHC. If the convict feels unfairly treated by a strict therapist, they could be more prone to recidivism due to a growing hatred for

Aut hority (found in the IV $\leftarrow\rightarrow$ U $\leftarrow\rightarrow$ X on the DAG).

The Nonzero First-Step Could be Violated if the Propensity of a therapist to assign MHC-level Scores does not effect the average probability of assignment to MHC. For instance, if no sampled inmates are mentally ill, different Propensities to assign MHC Should not effect probability of needing to attend MHC.

Monotonicity could be violated if the Propensity of therapists to assign MHC does not operate in the same direction for assigning inmates to MHC. For instance, A therapist may see sexual crimes as a sign of depravity due to a preconceived notion that also makes him more likely to assign people to MHC (e.g. Personal experience, history of mental illness in the family, Personal bitterness, etc.) If the therapist acts on this vendetta, they might send sex criminals to normal court so they don't get the benefits of MHC; thus, their propensity to notice mental illness assigns non-sex, mentally ill patients to MHC while sending sex offending mentally ill patients to normal court.

e. If we use 2SLS, the first stage is:

$$D_i = \gamma + PZ_i + \epsilon_i$$

Z_i = Propensity to assign MHC

The second stage is:

$$Y_i = \beta + S D_i + V_i$$

D_i = Assignment to MHC

V_i = Recidivism

With Z the IV, D_i the treatment, Y_i the outcome.

The causal model is then

$$Y_i = \alpha + \delta S_i + N_i$$

with $\text{Cov}(Z, N_i) = 0$, $P \neq 0$

F. If all five assumptions are satisfied, ZV (or, in this case, 2SLS Specification) estimate the Local Average Treatment Effect:

$$\delta_{ZV, LATE} = E[(Y_i^1 - Y_i^0) | D_i^1 - D_i^0 = 1]$$

In other words, this is the ATE of D on Y for inmates whose assignment was changed by the instrument, i.e. whose assignment to MHC was changed by the propensity to assign. In effect, this is the ATE for "compliers".

If there are heterogeneous treatment effects, this will only measure the treatment effect on inmates whose assignment changed as a result of a change in propensity to assign MHC,

while ATT measures treatment effects across all inmates and ATC measures treatment effects for inmates assigned to MHC.

G. If we have a weak instrument, then the bias of 2SLS is:

$$\begin{aligned}\hat{\beta}_{2SLS} - \beta &= (X' P_Z X)^{-1} X' P_Z V \\ &= (X' P_Z X)^{-1} \Pi' ZV + (X' P_Z X)^{-1} \gamma' P_Z V\end{aligned}$$

Taking the expectation and using algebra,

$$E[\hat{\beta}_{2SLS} - \beta] \approx \frac{\partial v_m}{\partial z_m} \left[\frac{E(\Pi' Z^2 \Pi)}{\sigma^2} + 1 \right]^{-1}$$

Replace the interpretation with the F-statistic for joint significance of the first stage:

$$E[\hat{\beta}_{2SLS} - \beta] \approx \frac{\partial v_m}{\partial z_m} \cdot \frac{1}{F+1}$$

Then this is centred around OLS bias of

$$E[\hat{\beta}_{OLS} - \beta] = \frac{\partial v_m}{\partial z_m}$$

If $F \rightarrow 0$, i.e. very weak first stage so instrument barely affects treatment, this is the same bias as OLS. If $F \rightarrow \infty$, i.e. very strong instrument that explains treatment well, this approaches 0.

We test if this bias is problematic by running an F-test on the joint significance of Z in the first stage,

that is, we test if the instruments are jointly zero coefficients.
If F is low, then we are less comfortable rejecting the idea that the instruments are jointly 0, and the bias raises.
If F is high, we are confident that the instruments are not jointly 0, and bias falls. If F is low, bias is likely a large problem. It shrinks as F grows.

3. a. Done. Note there are two tables for this part because I'm unfamiliar with HTML and RTF and couldn't make them play nice. I also don't know what .gitignore vs Exam-2-Empirical.Rproj are, because R studio makes them. The links:

<https://github.com/zachary-chance1/Exam-2-Empirical.git>

b. The OLS Coefficient on $\ln P$ is -0.52; in other words, if price increases by 10%, quantity decreases by 5.2%. This is not causal (not the ATE), but is instead the SDO. Once again, we know SDO suffers from Selection bias and heteroskedastic treatment effect bias. In other words, this number means very little because people choose to come to the market on any given day, which determines price. Thus, this estimate is very biased and is not the causal parameter.

c. When Wave 2 is the IV, the coefficient on $\ln P$ is -0.84. If price increases by 10%, quantity decreases by 8.4%. When Speed 3 is the IV, the coefficient on $\ln P$ is -1.57. If price increases by 10%, quantity decreases by 15.7%.

Under Wave 2 IV, fish appear to be inelastic. Under Speed 3 IV, fish appear to be elastic.

(While I can't interpret the day of week fixed effects, the question is regarding price elasticity, so I opted not to do so for the sake of clarity)

d. If wave 2 is the IV, the coefficient is 0.107.
For every 1 unit increase in 2-day lagged wave height,
price rose by 10.7%. The F-statistic is 27.625.
This leads us to believe 2-day lagged waves is a strong instrument.
If Speed3 is the IV, the coefficient is 0.017.
For every 1 unit increase in wind speed (3-day lagged),
price rose by 1.7%. The F-statistic is 8.474, which
is below the rule of thumb of 10. This leads us to
believe that Speed3 is a weak instrument.
Of these two, I prefer Wave2 because it is not weak.
However, it still has flaws in that it may not
satisfy the exclusion principle (effect of wave height on
type and length of fish, not just quantity) and OLS best
estimates LATE for compliers. That said, weak
instruments provide biased estimates, so Wave2
is likely the best choice.