1. STUDENT ID NUMBER: 10875563
2. COURSE UNIT CODE: ECON60052
3. ECON COURSE UNIT TITLE: Microeconometrics

Contents

1 Question-1 **2**

2 Question-2 **3**

3 Question-3 **5**

4 Question-4 **8**

5 Question-5 **10**

1 Question-1

(a) 2,4; for small sample validity, we would need extra assumption of normality of u.

(b) 1,3

(c) 1,2

- There are variables such as family culture towards health and living standard that impact both sugar consumption and dental health.
- Measurement error is obvious as person is self reporting the sugar consumption. If person just randomly reports her sugar consumption, then we may find no impact while in reality there might be an impact.
- This seems unlikely as your dental health may not determine your sugar consumption.

(d) 1,2

(e) 2,3

(f) 1,2

(g) 1,2,3

(h) 3

(i) 1

(j) 1

2 Question-2

The model is given by

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \tag{1}$$

where $E[u|x_1, x_2] = 0$. We also know that x_1, x_2 are related by

$$x_1 = \delta_0 + \delta_1 x_2 + \epsilon \tag{2}$$

and $E[\epsilon|x_2] = 0$ and $Cov(x_1, x_2) \neq 0$.

a

Since $\epsilon = x_1 - \delta_0 - \delta_1 x_2$,

$$Cov(\epsilon, y) = Cov(x_1 - \delta_0 - \delta_1 x_2, y) = Cov(x_1, y) + Cov(-\delta_0, y) + Cov(-\delta_1 x_2, y)$$

Since δ_0, δ_1 are constants, $Cov(-\delta_0, y) = 0$ and $Cov(-\delta_1 x_2, y) = -\delta_1 Cov(x_2, y)$. Therefore,

$$Cov(\epsilon, y) = Cov(x_1, y) - \delta_1 Cov(x_2, y)$$

From (2), we know that δ_1 is population parameter for regression of x_1 on x_2. Therefore,

$$\delta_1 = \frac{Cov(x_1, x_2)}{Var(x_2)}$$

Using this expression,

$$Cov(\epsilon, y) = Cov(x_1, y) - \frac{Cov(x_1, x_2)}{Var(x_2)} Cov(x_2, y) = \frac{Cov(x_1, y)Var(x_2) - Cov(x_1, x_2)Cov(x_2, y)}{Var(x_2)}$$

Again using (2),

$$Var(\epsilon) = Var(x_1 - \delta_0 - \delta_1 x_2) = Var(x_1 - \delta_1 x_2) + Var(-\delta_0) + 2Cov(x_1 - \delta_0 - \delta_1 x_2, -\delta_0)$$

Since,δ_0 is a constant,

$$Var(\epsilon) = Var(x_1 - \delta_1 x_2) + 0 + 2 * 0$$

$$Var(\epsilon) = Var(x_1 - \delta_1 x_2) = Var(x_1) + Var(-\delta_1 x_2) + 2Cov(x_1, -\delta_1 x_2)$$

$$\implies Var(\epsilon) = Var(x_1) + \delta_1^2 Var(x_2) - 2\delta_1 Cov(x_1, x_2)$$

Again using expression for δ_1,

$$\implies Var(\epsilon) = Var(x_1) + \frac{Cov(x_1, x_2)^2}{Var(x_2)^2} Var(x_2) - 2\frac{Cov(x_1, x_2)}{Var(x_2)} Cov(x_1, x_2)$$

$$\implies Var(\epsilon) = Var(x_1) - \frac{Cov(x_1, x_2)^2}{Var(x_2)} = \frac{Var(x_1)Var(x_2) - Cov(x_1, x_2)^2}{Var(x_2)}$$

Inserting these expressions back into main expression, we get

$$\frac{Cov(\epsilon, y)}{Var(\epsilon)} = \frac{\frac{Cov(x_1,y)Var(x_2) - Cov(x_1,x_2)Cov(x_2,y)}{Var(x_2)}}{\frac{Var(x_1)Var(x_2) - Cov(x_1,x_2)^2}{Var(x_2)}}$$

$$= \frac{Cov(x_1, y)Var(x_2) - Cov(x_1, x_2)Cov(x_2, y)}{Var(x_2)} \frac{Var(x_2)}{Var(x_1)Var(x_2) - Cov(x_1, x_2)^2}$$

$$\implies \frac{Cov(\epsilon, y)}{Var(\epsilon)} = \frac{Cov(x_1, y)Var(x_2) - Cov(x_1, x_2)Cov(x_2, y)}{Var(x_1)Var(x_2) - Cov(x_1, x_2)^2} \tag{3}$$

b

Using (2) in (1), we get

$$y = \beta_0 + \beta_1\delta_0 + \beta_1\delta_1 x_2 + \beta_1\epsilon + \beta_2 x_2 + u$$

$$y = \beta_0 + \beta_1\delta_0 + (\beta_1\delta_1 + \beta_2)x_2 + \beta_1\epsilon + u$$

We see that (3) gives population form of coefficient on ϵ when we regress y on ϵ. Suppose,

$$y = \beta_0 + \beta_1\delta_0 + \beta_1\epsilon + \eta$$

where $\eta = (\beta_1\delta_1 + \beta_2)x_2 + u$ satsifies usual assumptions because x_2, u are uncorrelated with ϵ.

$$\beta_1 = \frac{Cov(\epsilon, y)}{V(\epsilon)} = \frac{Cov(x_1, y)Var(x_2) - Cov(x_1, x_2)Cov(x_2, y)}{Var(x_1)Var(x_2) - Cov(x_1, x_2)^2}$$

Hence, given the structures we imposed on x_1, x_2, y, u, ϵ, the expression identifies this linear coefficient for relationship between y and ϵ which is same as β_1. This is just Frisch Waugh theorem as we are regressing y on residuals after partialling out x_2.

c

When we know that $\epsilon = 0$, then (2) becomes,

$$x_1 = \delta_0 + \delta_1 x_2$$

and therefore x_1 is a deterministic function of x_2 and our basic model (1) suffers from perfect multicollinearity and we would not be able to estimate all $\beta_j, j = 0, 1, 2$ simultaneously.

d

When $Cov(x_1, x_2) = 0$, we get

$$\frac{Cov(\epsilon, y)}{Var(\epsilon)} = \frac{Cov(x_1, y)}{Var(x_1)}$$

In this case, we get same expression even after partialling out because x_1 and x_2 are not correlated. So, Frisch Waugh formula boils down to usual simple linear regression formula when regressors are not correlated.

3 Question-3

a

The outcome we observe Y_i is given by

$$Y_i = Y_i(0) + [Y_i(1) - Y_i(0)]D_i$$

If $D_i = 1$ i.e user i consumes health professional's information or watches the video, then individual is treated and what we observe is $Y_i(1)$ and when $D_i = 0$, we observe $Y_i(0)$, where $Y_i(1)$ is distance travelled away from morning location relative to February 2020 which is benchmark month for individual i, if individual i was intensely treated i.e looked at video while $Y_i(0)$ is distance travelled away from morning location relative to February 2020 if individual i didn't look at the video.

b

So the observed difference between expected outcome of intensely treated individuals to less-intensely treated individual $E(Y_i|D_i = 1) - E(Y_i|D_i = 0)$ is decomposed as

$$\underbrace{E[Y_i|D_i = 1] - E[Y_i|D_i = 0]}_{(I)}$$

$$= \underbrace{E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]}_{(II)} + \underbrace{E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]}_{(III)}$$

$$= \underbrace{E[Y_i(1) - Y_i(0)]}_{(IV)} + \underbrace{(II) - (IV)}_{\textit{Sorting Gain}} + \underbrace{E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]}_{(III)}$$

1. We can see that we need to add and subtract the $E[Y_i(1) - Y_i(0)]$. So, $(IV) = E[Y_i(1) - Y_i(0)]$.

2. As we see that (I) is observed difference in Y between individuals with high intensity treatment with individuals in low intensity treatment. This is also population raw differential. The expression (II) gives

average causal effect on the treated individuals because

$$(II) = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] = E[Y_i(1) - Y_i(0)|D_i = 1]$$

(III) on the other hand is selection bias as this gives difference between average non-treated outcome of treated and average non-treated outcome of non-treated.
(IV) is average causal effect of treatment in the population as this gives expected difference between $Y(1)$ and $Y(0)$ in population.
The Sorting Gain is defined as

$$SortingGain = (II) - (IV) = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] - E[Y_i(1) - Y_i(0)]$$

$$\implies SortingGain = \{E[Y_i(1)|D_i = 1] - E[Y_i(1)]\} - \{E[Y_i(0)|D_i = 1] - E[Y_i(0)\}$$

So Sorting Gain captures difference between the difference of treated outcome of treated individuals with treated outcome of population as a whole and difference between non-treated outcome of treated individuals minus non-treated outcome of population as a whole. If the treated people self-select into treatment i.e they are ones looking for expert advice to actively listen to experts compared to average population, then $E[Y_i(1)|D_i = 1] - E[Y_i(1)]$ be negative as they would tend to stay home and respond more to treatment while $E[Y_i(0)|D_i = 1] - E[Y_i(0)$ might be considered positive as they without expert advice, they tend to go out more often and as result, entire sorting gain becomes negative and this may over estimate the true effect of encouragement campaign or over estimate impact of experts.

3. We expect the selection bias term (III) to be negative as treated individuals care about COVID and expert advice, so even without treatment, they tend to be careful and stay more home compared to non-treated individuals. In the context of US, we expect the term (II) which is average causal effect for treated to be negative as they self-select into it and value expert advice and will tend to go out less compared to without treatment. The average causal effect for the population as a whole (IV) might be 0 or negative as we would not expect expert advice to increase movement rather than decreasing. The sorting we argued would be negative.

c

In general we are interested in average causal effect for the population as a whole or average causal effect for the treated. If we randomize, we can get rid of selection and sorting gain term. Suppose D_i is randomly assigned then

$$(III) = E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0] = E[Y_i(0)] - E[Y_i(0)] = 0$$

as D_i is independent of Y_i. Similarly, for Sorting Gain

$$SortingGain = \{E[Y_i(1)|D_i = 1] - E[Y_i(1)]\} - \{E[Y_i(0)|D_i = 1] - E[Y_i(0)\}$$

$$= \{E[Y_i(1) - E[Y_i(1)\} - \{E[Y_i(0)] - E[Y_i(0)]\} = 0 - 0 = 0$$

Therefore, the observed difference (I) is equal to

$$(I) = E[Y_i(1) - Y_i(0)] = E[Y_i(1) - Y_i(0)|D_i = 1] = E[Y_i(1) - Y_i(0)|D_i = 0]$$

Through randomization, we are able to identify average treatment effect, average treatment effect on the treated and average treatment effect on the non-treated.

d

1. Authors report a table of characteristics at baseline in order to test and show whether randomization actually worked or not. We know if treatment was actually random, then difference between baseline outcomes must be zero and there should not be any selection or sorting.
2. Another better way is to regress these baseline variables on treatment and test if the coefficient on treatment is significantly different from 0 or not. The table only reports the means and standard errors for the mean, while regression is a lot more flexible and we can report p-values for significance which tell us if there is difference between means of two groups.
3. We observe that samples are reasonably balanced as most of outcomes take similar average values in both samples. Sample size is same for both groups in Thanks Giving while observe that some variables take higher values in high intensity sample. Broadly, our sample is balanced and we can do causal analysis.

e

Since assignment was randomly carried out, we can interpret this difference as causal different. We observe that right before Thanks Giving, 0 is not in whiskers and entire interval is below 0 which means that average effect of campaign is negative and significant. A negative impact means that those who listen to expert advice videos tend to stay home more compared to those who didn't when both are compared with their average movement in February 2020.

f

Authors are interested in average causal effect of expert advice on the population. Imperfect compliance is when individuals do not actually follow the treatment assignment and are able to move across treatment

arms. If we have imperfect compliance, our randomization no longer works and we may again get selection or sorting bias depending on how non-compliers behaved. If certain assumptions are satisfied, then we can use LATE estimation by using assignment as a instrument for the treatment which identifies the causal effect on compliers i.e those who followed the treatment.

4 Question-4

a

The first estimator in option:A is fixed effects or within estimator as we are time demeaning data while the second estimator in option:B is Least Squares Dummy Variables Estimator (LSDV) as it introduces a dummy for each time period. The model to be estimated is

$$y_{i,t} = \beta_0 + \beta_1 x_{i,t} + \alpha_i + u_{i,t}$$

Taking average with respect to time, we get

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_i + \bar{u}_i$$

Then taking the difference of these two equations we get specification for the model in option:A

$$\underbrace{\ddot{y}_{i,t}}_{y_{i,t}-\bar{y}_i} = \beta_1 \underbrace{\ddot{x}_{i,t}}_{x_{i,t}-\bar{x}_i} + \underbrace{\ddot{u}_{i,t}}_{u_{i,t}-\bar{u}_i} \tag{A}$$

While specification for option:B is

$$y_{i,t} = \beta_1 x_{i,t} + \sum_{j=1}^{5} \gamma_j d_j + u_{i,t} \tag{B}$$

where d_j is dummy for for individual j. The estimator from (A) is $\hat{\beta}_1^A$ and from (B) $\hat{\beta}_1^B$. We see numerical estimate of β_1 is same from both specifications which is -0.8883 while SE from (A) is 0.5366 and from (B) SE is 0.6692. Both of these specifications do not account for serial correlation and hetroskadasticity in the error term. (A) seems to have a lower standard error but it is an underestimation as this does not account for degrees of freedom lost while calculating means. Since we have 5 individuals, the correct degrees of freedom would be 9. So, the correct estimate for $\hat{\sigma^2}$ would be $\frac{0.7588}{9}$ rather than $\frac{0.7588}{14}$. So, correct SE is $0.5366 \times \frac{\sqrt{14}}{\sqrt{9}} = 0.6692$ which is same as from LSDV regression. Hence, if we are not using R's FE estimator, then we need to be careful about degrees of freedom correction.

b

$$y_{it} = \beta_0 + \beta_1 x_{it} + \alpha_i + u_{it}$$

Let $\bar{y_i} = \frac{\sum_{t=1}^{T} y_{it}}{T}$, $\bar{x_i} = \frac{\sum_{t=1}^{T} x_{it}}{T}$ & $\bar{u_i} = \frac{\sum_{t=1}^{T} u_{it}}{T}$

Then, based on specification in (A) the expression for $\hat{\beta_1^A}$ will be:

$$\hat{\beta_1^A} = \frac{\sum_{t=1}^{2}\sum_{i=1}^{2}(x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)}{\sum_{t=1}^{2}\sum_{i=1}^{2}(x_{it} - \bar{x}_i)^2}$$

The expression for $\hat{\beta_1^B}$ can be derived by solving the following minimization problem:

$$\sum_{i=1}^{2}\sum_{t=1}^{2}(y_{it} - \alpha_i - \beta_1^B x_{it})^2$$

where α_i are individual intercepts or individual fixed effects on the dummies, including β_0. Let $\hat{\beta_1^B}$ and $\hat{\alpha}_i$ be the solution to the minimization problem. These have to satisfy the first order conditions, which are:

$$-2\sum_{t=1}^{2}(y_{it} - \hat{\alpha}_i - \hat{\beta_1^B} x_{it}) = 0$$

$\implies \hat{\alpha}_i = \bar{y_i} - \bar{x_i}\hat{\beta_1^B}$ for each α_i where $\bar{x}_i$,$\bar{y}_i$ is as defined above, and

$$-2\sum_{i=1}^{2}\sum_{t=1}^{2}(y_{it} - \hat{\alpha}_i - \hat{\beta_1^B} x_{it})(x_{it}) = 0$$

for β_1^B. Then by substituting for $\hat{\alpha}_i$, we get

$$\hat{\beta_1^B} = \frac{\sum_{t=1}^{2}\sum_{i=1}^{2}(y_{it} - \bar{y}_i)(x_{it})}{\sum_{t=1}^{2}\sum_{i=1}^{2}(x_{it} - \bar{x}_i)(x_{it})}$$

Now, for $N = 2$ $T = 2$ we are going to show that $\hat{\beta_1^A} = \hat{\beta_1^B}$. To show the algebraic equivalence, we need to show that $\sum_{t=1}^{2}\sum_{i=1}^{2}(y_{it} - \bar{y}_i)(x_{it}) = \sum_{t=1}^{2}\sum_{i=1}^{2}(x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)$ &
$\sum_{t=1}^{2}\sum_{i=1}^{2}(x_{it} - \bar{x}_i)^2 = \sum_{t=1}^{2}\sum_{i=1}^{2}(x_{it} - \bar{x}_i)x_{it}$ i.e. denominator and numerator of both fractions are equal.

Now considering the numerator:

$$\sum_{t=1}^{2}\sum_{i=1}^{2}(x_{it}-\bar{x}_i)(y_{it}-\bar{y}_i) = (y_{11}-\bar{y}_1)(x_{11}-\bar{x}_1)+(y_{12}-\bar{y}_1)(x_{12}-\bar{x}_1)+(y_{21}-\bar{y}_2)(x_{21}-\bar{x}_2)+(y_{22}-\bar{y}_2)(x_{22}-\bar{x}_2)$$

$$= (y_{11}-\bar{y}_1)(x_{11})-(y_{11}-\bar{y}_1)(\bar{x}_1)+(y_{12}-\bar{y}_1)(x_{12})-(y_{12}-\bar{y}_1)(\bar{x}_1)+(y_{21}-\bar{y}_2)(x_{21})-(y_{21}-\bar{y}_2)(\bar{x}_2)+(y_{22}-\bar{y}_2)(x_{22})-(y_{22}-\bar{y}_2)(\bar{x}_2)$$

where $-(y_{12} - \bar{y}_1)(\bar{x}_1) = -(y_{12} - \frac{y_{11}+y_{12}}{2})(\bar{x}_1) = (y_{11} - \bar{y}_1)(\bar{x}_1)$
$\implies \sum_{t=1}^{2}\sum_{i=1}^{2}(x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) = (y_{11} - \bar{y}_1)(x_{11}) + (y_{12} - \bar{y}_1)(x_{12}) + (y_{21} - \bar{y}_2)(x_{21}) + (y_{22} - \bar{y}_2)(x_{22})$
$= \sum_{t=1}^{2}\sum_{i=1}^{2}(y_{it} - \bar{y}_i)(x_{it})$

Similarly, we can show that $\sum_{t=1}^{2}\sum_{i=1}^{2}(x_{it} - \bar{x}_i)^2 = \sum_{t=1}^{2}\sum_{i=1}^{2}(x_{it} - \bar{x}_i)x_{it}$

$\implies \hat{\beta_1^A} = \hat{\beta_1^B}$

5 Question-5

a

We certainly know that treated and non-treated sample differs on one important dimension which is their land holding. In general land holding is also correlated with many other important family characteristics and wealth such as land-holding by male members and role of women in household, therefore treated individuals differ from non-treated and we have selection bias without randomization. Program specifically targets the poor women with little or no land, and therefore their expenditure is less than the non-treated.

b

This clearly is a design for regression discontinuity as we have a deterministic variable for assignment. This is sharp design as everyone below the cut off of 0.5 is treated. Define X as land holding in acres and D_i the treatment status of woman i. Since we observe X_i without measurement error and people are precisely sorted based on this i.e no manipulation, the treatment is a discontinuous function of X_i.

$$D_i = \begin{cases} 1 & X_i \leq 0.5 \\ 0 & X_i > 0.5 \end{cases}$$

Let us denote $Y_i(0), Y_i(1)$ as woman i's consumption without credit and with credit respectively. As always, we can write the observed Y_i

$$Y_i = Y_i(0) + [Y_i(1) - Y_i(0)]D_i$$

$$Y_i = \begin{cases} Y_i(1) & X_i \leq 0.5 \\ Y_i(0) & X_i > 0.5 \end{cases}$$

This set up satisfies basic RDD assumptions i.e the land-holding X_i is a running variable, we have a precise cut-off at 0.5 that determines treatment. This is reverse RD design, however the specifications are same and we only interpret those below 0.5 as treated rather than those above it. The following assumption for the sharp design is satisfied,

$$\lim_{\epsilon \to 0} E[D|X = 0.5 - \epsilon] - \lim_{\epsilon \to 0} E[D|X = 0.5 + \epsilon] = 1 - 0 = 1$$

which shows discontinuity of treatment at 0.5.

We also need to assume that $E[Y_i(d)|X = x]$ is continuous for X near 0.5 and d=0,1. This shows that both potential outcomes move continuously around threshold and any sudden jump is because of treatment. We identify ATE or ATT near 0.5 by

$$\tau(0.5) = E[Y(1) - Y(0)|X = 0.5] = \lim_{\epsilon \to 0} E[Y_i|X = 0.5 - \epsilon] - \lim_{\epsilon \to 0} E[Y_i|X = 0.5 + \epsilon]$$

For estimation, we can use following specification with different slopes on both sides of threshold,

$$Y_i = \beta_0 + \tau D_i + \beta_1(X_i - 0.5) + \beta_2 D_i(X_i - 0.5) + u \tag{RDD-1}$$

This would give

$$E[Y|X \geq 0.5] = \beta_0 + \beta_1(X - 0.5)$$

$$E[Y|X < 0.5] = \beta_0 + \tau + (\beta_1 + \beta_2)(X - 0.5)$$

As we move near cut-off,

$$\lim_{\epsilon \to 0} E[Y|X = 0.5 - \epsilon] = \beta_0 + \tau$$

$$\lim_{\epsilon \to 0} E[Y|X = 0.5 + \epsilon] = \beta_0$$

Because as $\epsilon \to 0$, the deviation of X from 0.5 goes to 0 and that is why that term disappears. Therefore,

$$\tau = \lim_{\epsilon \to 0} E[Y|X = 0.5 - \epsilon] - \lim_{\epsilon \to 0} E[Y|X = 0.5 + \epsilon]$$

Hence, we can use (RDD-1) to estimate causal effect of poverty alleviation program using OLS for those with land-holding close to 0.5.

Advantages vs Disadvantages

RDD is a local design and gives treatment effect near the cutoff of 0.5. So, given assumptions above, it has strong internal validity but we cannot be sure about external validity. Our identification strongly relies on the fact that there is no other continuity at landholding of 0.5 that effects the consumption. Also, the causal interpretation of τ relies strictly on functional form. Therefore, using data we should look for different functional forms with the possibility of quadratic interaction terms or any higher order polynomial. We know that best alternative is an RCT at village level. Moreover, given this data, we can also use DiD framework but it would be difficult to argue for common trends assumption.

c

As discussed in (b), this is still RD design although design is no longer sharp and we have fuzzy RDD. Now, land-holding X determines eligibility rather than actual treatment. Let us denote treatment eligibility by Z, then

$$Z_i = \begin{cases} 1 & X_i \leq 0.5 \\ 0 & X_i > 0.5 \end{cases}$$

Let us define D(x) as treatment status of an individual when their land is exogenously set to x.We need following RDD assumptions to get LATE for compliers.

1. **RDD1**: The discontinuity of probability of treatment is,

$$\lim_{\epsilon \to 0} P[D|X = 0.5 - \epsilon] \neq \lim_{\epsilon \to 0} P[D|X = 0.5 + \epsilon]$$

   Hence, if your land is below 0.5, you are not sure to get treatment but the probability of getting treatment changes discontinuously at 0.5.

2. **RDD2**: The expected potential outcomes in the absence of treatment is continuous therefore,

$$\lim_{\epsilon \to 0} E[Y(0)|X = 0.5 - \epsilon] = \lim_{\epsilon \to 0} E[Y(0)|X = 0.5 + \epsilon]$$

   If there is discontinuity even in the absence of treatment, we would not be able to entangle true effect from treatment and jump in Y(0).

3. **RDD3**: Just like usual LATE framework, we need to assume that near 0.5 potential outcomes and treatment is uncorrelated with X_i

$$\{Y_i(1), Y_i(0), D_i(x)\} \perp X_i$$

   and there is neighbourhood of 0.5 i.e $e > 0$ s.t

$$\forall\ \epsilon\ s.t\ 0 < \epsilon < e,\ D_i(0.5 - \epsilon) \geq D_i(0.5 + \epsilon)$$

   This is a local monotonicity assumption needed for LATE so that there are no definers. If there are defiers, we would not be able to estimate causal effect in this context. Local compliers are those individuals for whom $D(0.5 - \epsilon) > D(0.5 + \epsilon)$.

In this setting, we are able to identify causal effect on the compliers i.e those who follow assignment rule. Hence, LATE for compliers near the cutoff is

$$\tau(0.5) = \lim_{\epsilon \to 0} E[Y(1) - Y(0)|D(0.5 - \epsilon) > D(0.5 + \epsilon), X = 0.5]$$

We know that using these assumptions, we identify local average treatment effect on compliers using Wild type population condition because we would use Z as IV for D and both are binary.

$$\tau(0.5) = \frac{\lim_{\epsilon \to 0} E[Y|X = 0.5 - \epsilon] - \lim_{\epsilon \to 0} E[Y|X = 0.5 + \epsilon]}{\lim_{\epsilon \to 0} E[D|X = 0.5 - \epsilon] - \lim_{\epsilon \to 0} E[D|X = 0.5 + \epsilon]} \tag{4}$$

Proof

$$(i) = \lim_{\epsilon \to 0} E[Y|X = 0.5 - \epsilon] - \lim_{\epsilon \to 0} E[Y|X = 0.5 + \epsilon]$$

$$= \lim_{\epsilon \to 0} E[Y(0) + (Y(1) - Y(0)D(X))|X = 0.5 - \epsilon] - \lim_{\epsilon \to 0} E[Y(0) - (Y(1) - Y(0))D(X)|X = 0.5 + \epsilon]$$

The $\lim_{\epsilon\to 0} E[Y(0)|X = 0.5 - \epsilon]$ gets cancelled with $\lim_{\epsilon\to 0} E[Y(0)|X = 0.5 + \epsilon]$ becasue of continuity assumption RDD2.

$$\implies (i) = \lim_{\epsilon\to 0} E[(Y(1) - Y(0)D(0.5 - \epsilon))|X = 0.5 - \epsilon] - \lim_{\epsilon\to 0} E[(Y(1) - Y(0))D(0.5 + \epsilon)|X = 0.5 + \epsilon]$$

Using Imbens and Angrist (1995)[1] with RDD1 and RDD3[2], we can write this as

$$\implies (i) = \lim_{\epsilon\to 0} E[(Y(1)-Y(0)|(D(0.5-\epsilon)-D(0.5+\epsilon)) = 1](E[D(0.5-\epsilon)|X = 0.5-\epsilon]-E[D(0.5+\epsilon)|X = 0.5+\epsilon])$$

where I replaced probabilities in case of Imbens and Angrist (1995) with expectation because D is binary treatment. We know limit of product is product of limit,

$$(i) = \lim_{\epsilon\to 0} E[(Y(1)-Y(0)|(D(0.5-\epsilon)-D(0.5+\epsilon)) = 1|X = 0.5-\epsilon] \lim_{\epsilon\to 0}(E[D(0.5-\epsilon)|X = 0.5-\epsilon]-E[D(0.5+\epsilon)|X = 0.5+\epsilon])$$

Dividing (i) with $\lim_{\epsilon\to 0}(E[D(0.5 - \epsilon)|X = 0.5 - \epsilon] - E[D(0.5 + \epsilon)|X = 0.5 + \epsilon])$,

$$\lim_{\epsilon\to 0} E[(Y(1)-Y(0)|(D(0.5-\epsilon)-D(0.5+\epsilon)) = 1|X = 0.5-\epsilon] = \lim_{\epsilon\to 0} E[(Y(1)-Y(0)|(D(0.5-\epsilon)-D(0.5+\epsilon)) = 1|X = 0.5]$$

which is average treatment effect on the compliers near the cutoff i.e $\tau(0.5)$. We can use either $0.5 - \epsilon$ or $0.5 + \epsilon$ when conditioning for average treatment effect. The condition $(D(0.5 - \epsilon) - D(0.5 + \epsilon)) = 1$ shows that this is causal effect for compliers.

Estimation for LATE

We can use 2SLS to estimate LATE by using Z_i as IV for D_i. Then, first and second stage are

$$D_i = \gamma + \delta Z_i + g(X - 0.5) + v \qquad \text{(First Stage)}$$

$$Y_i = \beta_0 + \tau D_i + f(X - 0.5) + u \qquad \text{(Second Stage)}$$

Let us define $\tilde{X} = X - 0.5$ and use quadratic polynomials for f and g.

$$D_i = \gamma + \delta Z_i + \gamma_1\tilde{X}_i + \gamma_2\tilde{X}_i^{\,2} + \gamma_3 Z_i\tilde{X}_i + \gamma_4 Z_i X_i^2 + v \qquad \text{(First Stage)}$$

$$Y_i = \beta_0 + \tau D_i + \beta_1\tilde{X}_i + \beta_2\tilde{X}_i^{\,2} + \beta_3 Z_i\tilde{X}_i + \beta_4 Z_i\tilde{X}_i^2 + u \qquad \text{(Second Stage)}$$

Here we are instrumenting D_i with Z_i, $D_i\tilde{X}_i$ with $Z_i\tilde{X}_i$ and same for quadratic term. Again here, τ would be our LATE estimate and it would be average treatment effect for the compliers with land-holding close to

[1] Angrist, Joshua, and Guido Imbens. "Identification and estimation of local average treatment effects." (1995).

[2] Note that conditioning Y(1)-Y(0) on 0.5-ϵ or $0.5 + \epsilon$ does not matter in the neighborhood as we know both potential outcomes are continuous near the cutoff.

0.5 acres. Using first,

$$E[D_i|0.5-\epsilon<X_i\leq 0.5]-E[D_i|0.5+\epsilon>X_i>0.5]=\gamma+\delta+(\gamma_1+\gamma_3)\tilde{X}_i+(\gamma_2+\gamma_3)\tilde{X_i}^2-\gamma-\gamma_1\tilde{X}_i-\gamma_2\tilde{X_i}^2$$

Given definition of X_i, when ϵ goes to 0, $\tilde{X}_i$ also goes to 0. So,

$$\lim_{\epsilon\to 0}E[D_i|0.5-\epsilon<X_i\leq 0.5]-\lim_{\epsilon\to 0}E[D_i|0.5+\epsilon>X_i>0.5]=\delta$$

Similarly, using first stage in second stage equation,

$$Y_i=\alpha_0+\tau\delta Z_i+\alpha_1\tilde{X}_i+\alpha_2\tilde{X}_i^2+\alpha_3 Z_i\tilde{X}_i+\alpha_4 Z_i\tilde{X}_i^2+\psi_i$$

then, in reduced form

$$\lim_{\epsilon\to 0}E[Y_i|0.5-\epsilon<X\leq 0.5]-\lim_{\epsilon\to 0}E[Y_i|0.5+\epsilon>X>0.5]=\lim_{\epsilon\to 0}\{\alpha_0+\tau\delta+\alpha_3\tilde{X}_i+\alpha_4\tilde{X}_i^2-\alpha_0\}=\tau\delta$$

taking the rations, we see that τ is LATE,

$$\frac{\lim_{\epsilon\to 0}E[Y_i|0.5-\epsilon<X\leq 0.5]-\lim_{\epsilon\to 0}E[Y_i|0.5+\epsilon>X>0.5]}{\lim_{\epsilon\to 0}E[D_i|0.5-\epsilon<X_i\leq 0.5]-\lim_{\epsilon\to 0}E[D_i|0.5+\epsilon>X_i>0.5]}=\frac{\tau\delta}{\delta}=\tau$$

This is clearly an expression for Wald estimator but there is a limit which means that this is true only locally in the neighbourhood of 0.5. Therefore, we have shown that τ identifies LATE near the cutoff.