Sharp Vs Fuzzy DID

Sharp Vs Fuzzy DID

Most of treatment effect methods extend to situations with imperfect compliance without imposing much stronger assumptions

- RCTs with perfect compliance identify the ATE without having to make any assumption.
- RCTs with imperfect compliance identify the LATE under exclusion restriction and monotonicity: often credible assumptions in context of RCTs

Sharp Vs Fuzzy DID

Most of treatment effect methods extend to situations with imperfect compliance without imposing much stronger assumptions

- RCTs with perfect compliance identify the ATE without having to make any assumption.
- RCTs with imperfect compliance identify the LATE under exclusion restriction and monotonicity: often credible assumptions in context of RCTs
- Sharp RDs identify the ATE at the threshold under the assumption that the conditional mean of the outcome is continuous
- Fuzzy RDs identify the LATE at the threshold under the same assumption plus monotonicity
- DID with imperfect compliance rely however on much stronger assumptions than standard DID

Sharp Vs Fuzzy DID

Many natural experiments do not lead to a sharp change in treatment rate for any group defined by a set of observable characteristics, but only to a larger increase of the treatment rate in some groups than in others.

- A good example is Duflo (AER, 2001), who uses a primary school construction program in Indonesia to measure returns to education
- Many primary schools were constructed in districts where there were few schools before the program: **Treatment Group**
- Few primary schools were constructed in districts which already had many schools: **Control Group**
- Primary school completion rate increased more in treatment group (but not from 0 to 100) compared to control group (didn't remain constant)

Sharp Vs Fuzzy DID

Table 1: Share of individuals completing primary school

	Older cohort	Younger cohort
High treatment regions	81.2%	90.0%
Low treatment regions	89.8%	94.3%

- In such fuzzy design, a commonly used estimand is the Wald-DID estimand

$$W_{DID} = \frac{E(Y_{1,1}) - E(Y_{1,0}) - \Big(E(Y_{0,1} - E(Y_{0,0})\Big)}{E(D_{1,1}) - E(D_{1,0}) - \Big(E(D_{0,1}) - E(D_{0,0})\Big)} \qquad (1)$$

- ratio of DID of outcome Y variable and DID of treatment rates D
- What happens to W_{DID} in sharp design?
- In fuzzy designs, Wald-DID estimand relies on stronger assumptions than DID in sharp designs (De Chaisemartin & D'Haultfoeuille, RESTUD 2018)

- Notation: drop i subscript from notation in Session 1

- For any random variable R and for every (d, g, t), let $R_{g,.} \sim R|G = g$, $R_{.,t} \sim R|T = t$, $R_{g,t} \sim R|G = g, T = t$, and $R_{d,g,t} \sim R|D = d, G = g, T = t$

- For any $(g, t) \in \{0, 1\}^2$, let $\Delta^{TR}_{g,t} = E\big(Y(1) - Y(0)|D = 1, G = g, T = t\big)$: ATT in group g and period t

- DID in treatment rates: $DID_D = E(D_{1,1}) - E(D_{1,0}) - \big(E(D_{0,1} - E(D_{0,0})\big)$

- Recall DID assumption: Assumption 1

$$E(Y_{i1}(0)|G_i = 1) - E(Y_{i0}(0)|G_i = 1) = E(Y_{i1}(0)|G_i = 0) - E(Y_{i0}(0)|G_i = 0)$$

- **Theorem 1:** If Assumption 1 is satisfied

$$W_{DID} = \frac{E(D_{1,1})}{DID_D}\Delta^{TR}_{1,1} - \frac{E(D_{1,0})}{DID_D}\Delta^{TR}_{1,0} - \frac{E(D_{0,1})}{DID_D}\Delta^{TR}_{0,1} + \frac{E(D_{0,0})}{DID_D}\Delta^{TR}_{0,0} \quad (2)$$

Proof Theorem 1

$$
\begin{aligned}
W_{DID} &= \frac{DID}{DID_D} \\
DID &= E(Y_{1,1}) - E(Y_{1,0}) - E(Y_{0,1}) + E(Y_{0,0}) \\
&= E\big(Y_{1,1}(1)|D=1\big)E(D_{1,1}) + E\big(Y_{1,1}(0)|D=0\big)\big(1 - E(D_{1,1})\big) \\
&- E\big(Y_{1,0}(1)|D=1\big)E(D_{1,0}) - E\big(Y_{1,0}(0)|D=0\big)\big(1 - E(D_{1,0})\big) \\
&- E\big(Y_{0,1}(1)|D=1\big)E(D_{0,1}) - E\big(Y_{0,1}(0)|D=0\big)\big(1 - E(D_{0,1})\big) \\
&+ E\big(Y_{0,0}(1)|D=1\big)E(D_{0,0}) + E\big(Y_{0,0}(0)|D=0\big)\big(1 - E(D_{0,0})\big) \\
&= E\big(Y_{1,1}(1) - Y_{1,1}(0)|D=1\big)E(D_{1,1}) + E\big(Y_{1,1}(0)\big) \\
&- E\big(Y_{1,0}(1) - Y_{1,0}(0)|D=1\big)E(D_{1,0}) - E\big(Y_{1,0}(0)\big) \\
&- E\big(Y_{0,1}(1) - Y_{0,1}(0)|D=1\big)E(D_{0,1}) - E\big(Y_{0,1}(0)\big) \\
&+ E\big(Y_{0,0}(1) - Y_{0,0}(0)|D=1\big)E(D_{0,0}) + E\big(Y_{0,0}(0)\big) \\
&= E(D_{1,1})\Delta^{TR}_{1,1} - E(D_{1,0})\Delta^{TR}_{1,0} - E(D_{0,1})\Delta^{TR}_{0,1} + E(D_{0,0})\Delta^{TR}_{0,0}
\end{aligned}
$$

Sharp Vs Fuzzy DID

- Under common trends alone, W_{DID} identifies a weighted sum of the 4 ATTs in each group and time period
  - 2 of the ATTs enter with a negative weights
  - Weights can be estimated from data
- Intuition:
  - In a sharp DID, the only departure from the scenario where nobody is treated is that units in group 1 and period 1 receive the treatment
  - any discrepancy between the trends of the mean outcome in the two groups must come from the effect of the treatment in group 1 and period 1: DID identifies ATT
  - In a fuzzy design, there are potentially four departures from the scenario where nobody is treated
  - the discrepancy between the trends of the mean outcome in the two groups can come from the treatment effect in any group and time period
  - 2 ATTs enter with a negative sign because of first difference inherent to DID
- Why is this an issue?

Wald- DID

- Why is this an issue?
- W_{DID} may not even have the same sign as any of the $\Delta_{g,t}^{TR}$ if they are heteregeneous
- Under common trends assumption alone, W_{DID} does not identify anything meaningful: **link to structural models**
- We can impose more assumptions to make W_{DID} identify something meaningful

Wald - DID: More assumptions

Assumption 2 (*Treatment monotonicity*): $\exists$ 2 random variables $D(0), D(1)$ for period 0 and 1:

1. $D = D(T)$
2. $\big(D(0), D(1)\big) \perp T|G$
3. $P\big(D(1) \geq D(0)|G\big) = 1$ or $P\big(D(1) \leq D(0)|G\big) = 1$

Assumption 3 (*Stable treatment effect*): $\forall g \in \{0, 1\}$

$$E\big(Y(1) - Y(0)|G = g, T = 1, D(0) = 1\big) = E\big(Y(1) - Y(0)|G = g, T = 0, D(0) = 1\big)$$

- Assumption 2.2: Distribution of D stable across period in each group
- Assumption 2.3: between each pair of consecutive periods, in a given group there cannot be both units whose treatment increases and units whose treatment decreases

Wald - DID: More assumptions

- Assumption 3: in every group, the average treatment effect among units treated in period 0 does not change between 0 and 1.
- This restricts treatment effect heterogeneity over time, but not between groups
- $S = \{D(1) \neq D(0), T = 1\}$: units whose treatment status switches between T = 0 and T = 1
- $\forall g \in \{0,1\}, \Delta^S_{g,1} = E\big(Y(1) - Y(0) | S, G = g, T = 1\big)$: LATE of switchers in group g and at period 1

Wald - DID: More assumptions

- **Theorem 2:** If Assumption 1 to 3 are satisfied,

$$W_{DID} = \frac{E(D_{1,1}) - E(D_{1,0})}{DID_D}\Delta^S_{1,1} - \frac{E(D_{0,1}) - E(D_{0,0})}{DID_D}\Delta^S_{0,1} \tag{3}$$

- W_{DID} identifies in this case a weighted sum of LATEs of switchers in both groups
- Weights can be estimated: we can check whether this estimator is identifying something meaningful or not
- **Proof**

$$\begin{aligned} E(D_{1,1}) - E(D_{1,0}) &= P(D(1) = 1|G = 1, T = 1) - P(D(0) = 1|G = 1, T = 0) \\ &= P(D(1) = 1|G = 1, T = 1) - P(D(0) = 1|G = 1, T = 1) \\ &= P(D(1) = 1, D(0) = 0|G = 1, T = 1) \\ &= P(S|G = 1, T = 1). \qquad (1.22) \end{aligned}$$

Moreover,

$$
\begin{aligned}
& E(Y_{1,1}) - E(Y_{1,0}) \\
= \; & E(Y|G=1,T=1) - E(Y|G=1,T=0) \\
= \; & E\left(Y(1)|G=1,T=1,D(0)=1\right) P(D(0)=1|G=1,T=1) \\
+ \; & E\left(Y(1)|G=1,T=1,S\right) P(S|G=1,T=1) \\
+ \; & E(Y(0)|G=1,T=1,D(1)=0) P(D(1)=0|G=1,T=1) \\
- \; & E\left(Y(1)|G=1,T=0,D(0)=1\right) P(D(0)=1|G=1,T=0) \\
- \; & E(Y(0)|G=1,T=0,D(0)=0) P(D(0)=0|G=1,T=0) \\
= \; & E\left(Y(1)-Y(0)|G=1,T=1,D(0)=1\right) P(D(0)=1|G=1,T=1) \\
+ \; & E\left(Y(1)-Y(0)|G=1,T=1,S\right) P(S|G=1,T=1) \\
+ \; & E(Y(0)|G=1,T=1) \\
- \; & E\left(Y(1)-Y(0)|G=1,T=0,D(0)=1\right) P(D(0)=1|G=1,T=0) \\
- \; & E(Y(0)|G=1,T=0) \\
= \; & E\left(Y(1)-Y(0)|G=1,T=1,D(0)=1\right) P(D(0)=1|G=1,T=1) \\
+ \; & E\left(Y(1)-Y(0)|G=1,T=1,S\right) P(S|G=1,T=1) \\
+ \; & E(Y(0)|G=1,T=1) \\
- \; & E\left(Y(1)-Y(0)|G=1,T=0,D(0)=1\right) P(D(0)=1|G=1,T=1) \\
- \; & E(Y(0)|G=1,T=0) \\
= \; & E\left(Y(1)-Y(0)|G=1,T=1,S\right) P(S|G=1,T=1) + E(Y(0)|G=1,T=1) - E(Y(0)|G=1,T=0) \\
= \; & \Delta^S_{1,1} P(S|G=1,T=1) + E(Y_{1,1}(0)) - E(Y_{1,0}(0)). \qquad (1.23)
\end{aligned}
$$

Similarly, one can show that

$$E(D_{0,1}) - E(D_{0,0}) = P(S|G=0, T=1) \quad (1.24)$$

$$E(Y_{0,1}) - E(Y_{0,0}) = \Delta^S_{0,1} P(S|G=0, T=1) + E(Y_{0,1}(0)) - E(Y_{0,0}(0)). \quad (1.25)$$

Taking the difference between Equations (1.23) and (1.25), and using Assumption 1, we obtain

$$DID = \Delta^S_{1,1} P(S|G=1, T=1) - \Delta^S_{0,1} P(S|G=0, T=1).$$

Dividing each side by DID_D and using Equations (1.22) and (1.24) yields the result.

- In which cases is W_{DID} estimate meaningful or not under these assumptions?
- When the share of treated units does not change over time in the control group, W_{DID} identifies the treatment effect among treatment group switchers
  - under: common trends + treatment monotonicity + stable treatment effect

- In which cases is W_{DID} estimate meaningful or not under these assumptions?
- When the share of treated units does not change over time in the control group, W_{DID} identifies the treatment effect among treatment group switchers
  - under: common trends + treatment monotonicity + stable treatment effect
- W_{DID} does not identify treatment effect when treatment rates increase in both treatment and control groups
  - Need to assume treatment effect homogeneity across groups: $\Delta^S_{1,1} = \Delta^S_{0,1}$

To sum up

- Sharp design: DID estimand only relies on a common trends assumption
- Fuzzy design: Wald-DID estimand identifies some interpretable measure of the treatment effect only if the treatment effect is homogeneous, at least over time,
- Sometimes it is necessary to impose treatment effect homogeneity both over time and between groups
- Have positive weights in Wald-DID only in very specific cases
- De Chaisemartin & D'Haultfoeuille (RESTUD 2018) propose an alternative estimand that relies on weaker assumptions

Time-Corrected Wald Estimand (De Chaisemartin & D'Haultfoeuille, RESTUD 2018)

- **Assumption 4**: Conditional common trend $E\big(Y(0)|G, T = 1, D(0) = 0\big) - E\big(Y(0)|G, T = 0, D(0) = 0\big)$ and $E\big(Y(1)|G, T = 1, D(0) = 1\big) - E\big(Y(1)|G, T = 0, D(0) = 1\big)$ do not depend on G

- Mean of $Y(0)$ (resp. Y (1)) follows the same evolution over time among treatment and control group units that were untreated (resp. treated) at $T = 0$

- Define $\delta_d = E(Y_{d,0,1}) - E(Y_{d,0,0})$: change in the mean outcome between period 0 and 1 for control group units with treatment status d

$$W_{TC} = \frac{E(Y_{1,1}) - E(Y_{1,0} + \delta_{D_{1,0}})}{E(D_{1,1}) - E(D_{1,0})} \tag{4}$$

Time-Corrected Wald Estimand

$$
\begin{aligned}
W_{TC} &= \frac{E(Y_{1,1}) - E(Y_{1,0} + \delta_{D_{1,0}})}{E(D_{1,1}) - E(D_{1,0})} \\
&= \frac{E(Y_{1,1}) - E\big(Y_{1,0} + (1 - D_{1,0})\delta_0 + D_{1,0}\delta_1\big)}{E(D_{1,1}) - E(D_{1,0})} \\
&= \frac{E(Y_{1,1}) - E(Y_{1,0}) - (1 - P(D_{1,0} = 1))\delta_0 - P(D_{1,0} = 1)\delta_1}{E(D_{1,1}) - E(D_{1,0})}
\end{aligned}
$$

- **Theorem 3:** If Assumption (2) and (4) are satisfied and if $E(D_{0,1}) = E(D_{0,0})$, then $W_{TC} = \Delta^S_{1,1}$
- Proof:

Following the same steps as those used to derive Equation (1.23), we obtain

$$
\begin{aligned}
& E(Y_{1,1}) - E(Y_{1,0}) \\
= \quad & E(Y(1) - Y(0)|S, G = 1, T = 1)P(S|G = 1) \\
+ \quad & (E\,(Y(1)|D(0) = 1, G = 1, T = 1) - E\,(Y(1)|D(0) = 1, G = 1, T = 0))\,P(D(0) = 1|G = 1) \\
+ \quad & (E\,(Y(0)|D(0) = 0, G = 1, T = 1) - E\,(Y(0)|D(0) = 0, G = 1, T = 0))\,P(D(0) = 0|G = 1). \quad (1.27)
\end{aligned}
$$

Then,

$$
\begin{aligned}
\delta_1 &= E(Y_{1,0,1}) - E(Y_{1,0,0}) \\
&= E(Y(1)|D(1) = 1, G = 0, T = 1) - E(Y(1)|D(0) = 1, G = 0, T = 0) \\
&= E(Y(1)|D(0) = 1, G = 0, T = 1) - E(Y(1)|D(0) = 1, G = 0, T = 0). \quad (1.28)
\end{aligned}
$$

- Last equality follows from $E(D_{01}) = E(D_{0,0})$ + A2 (treatment monotonicity) $\implies \{G = 0, D(1) = 1\} = \{G = 0, D(0) = 1\}$

Similarly

$$\delta_0 = E(Y(0)|D(0) = 0, G = 0, T = 1) - E(Y(0)|D(0) = 0, G = 0, T = 0). \qquad (1.29)$$

- the result follows combining Equations 1.27, 1.28, 1.29 and Assumption 4 (conditionnal common trend) once we note that:
  $P(D(0) = 1|G = 1) = P(D = 1|G = 1, T = 0)$ and
  $P(S|G = 1) = P(D = 1|G = 1, T = 1) - P(D = 1|G = 1, T = 0)$

Time-Corrected Wald Estimand: Intuition

$$W_{TC} = \frac{E(Y|G=1,T=1) - E\big(Y + (1-D)\delta_0 + D\delta_1 | G=1, T=0\big)}{E(D|G=1,T=1) - E(D|G=1,T=0)}$$

- This is almost the Wald ratio in the treatment group with time as the instrument
- except that we have $Y + (1-D)\delta_0 + D\delta_1$ instead of Y in 2nd term
- This difference arises because time is not a standard instrument: it can directly affect the outcome
- When the treatment rate is stable in the control group, we can identify the trends on $Y(0)$ and $Y(1)$ by looking at how the mean outcome of untreated and treated units changes over time in this group
- Under Assumption 4, these trends are the same in the 2 groups

Time-Corrected Wald Estimand: Intuition

- We can add these changes to the outcome of untreated and treated units in the treatment group in period 0,
- and recover the mean outcome we would have observed in this group in period 1 if switchers had not changed their treatment between the 2 periods
- This is what $Y + (1 - D)\delta_0 + D\delta_1$ does
- the numerator of W_{TC} compares the mean outcome in the treatment group in period 1 to the counterfactual mean we would have observed if switchers had remained untreated
- This gives LATE after a normalization

Time-Corrected Wald Estimand: extentions

- Theorem 3 assumes that $E(D_{0,1}) = E(D_{0,0})$
- When the share of treated units is not stable over time in the control group, $\Delta^S_{1,1}$ is only partially identified if Y is bounded
- One can therefore estimate the bounds of $\Delta^S_{1,1}$ (See De Chaisemartin & D'Haultfoeuille (RESTUD, 2018))
- Partial identification is still very useful for applied economist, specially when the identified set is compact
- For a scalar parameter, it makes a lot of sense, when you think of how it relates to confidence bands in the fully identified setting

Change-in-Changes Estimands in Fuzzy DID

Fuzzy CIC

- This is an alternative estimator to the Wald estimators in Fuzzy DID setting
- **Assumption 5:** Monotonicity and time invariance of unobservables $Y(d) = h_d(U_d, T)$ with $U_d \in \mathbb{R}$ and $h_d(u,t)$ strictly increasing in u for all $(d,t) \in \mathcal{S}((D,T))$. Moreover, $U_d \perp T|G, D(0)$
- Assuption 2 (treatment monotonicity) and 5 generalize the CIC model in Athey $\&$ Imbens (2006)
- They both imply $U_d \perp T|G$: they require that at each period, both potential outcomes are strictly increasing functions of a scalar unobserved heterogeneity term
- whose distribution is stationary over time, as in Athey $\&$ Imbens (2006)
- Assumption 5 also imposes that $U_d \perp T|G, D(0)$: the distribution of U_d must be stationary within subgroup of units sharing the same treatment status at $T = 0$

Fuzzy CIC

Assumption 6: Data restrictions (testable)

1. $\mathcal{S}(Y_{d,g,t}) = \mathcal{S}(Y)$ for $(d,g,t) \in \mathcal{S}((D,G,T))$ and $\mathcal{S}(Y)$ is a closed interval of $\mathbb{R}$
2. $F_{Y_{d,g,t}}$ is continuous on $\mathbb{R}$ and strictly increasing on $\mathcal{S}(Y)$ for $(d,g,t) \in \mathcal{S}((D,G,T))$

- 1st condition requires that the outcome have the same support in each of the eight treatment $\times$ group $\times$ period cells as in Athey & Imbens (2006)
- The 2nd condition requires that the distribution of Y be continuous with positive density in each of the eight cells
- Consider $Q_d(y) = F^{-1}_{Y_{d,0,1}} \circ F_{Y_{d,0,0}}(y)$ the quantile-quantile transform of Y from period 0 to 1 in control group conditional on $D = d$
- This transform maps y at rank q in period 0 into the corresponding y' at rank q in period 1

Fuzzy CIC

- Let's also define

$$F_{CIC,d}(y) = \frac{P(D_{1,1} = d)F_{Y_{d,1,1}}(y) - P(D_{1,0} = d)F_{Q_d(Y_{d,1,0})}(y)}{P(D_{1,1} = d) - P(D_{1,0} = d)}$$

$$W_{CIC} = \frac{E(Y_{1,1}) - E(Q_{D_{1,0}}(Y_{1,0}))}{E(D_{1,1}) - E(D_{1,0})}$$

- **Theorem 4:** If assumptions 5 and 6 are satisfied, and if $E(D_{0,1}) = E(D_{0,0})$, then $W_{CIC} = \Delta^S_{1,1}$ and $F^{-1}_{CIC,1}(q) - F^{-1}_{CIC,0}(q) = \tau_q$
- This result combines ideas from Imbens & Rubin (1997) and Athey & Imbens (2006)

Fuzzy CIC: Intuition

- We seek to recover the distribution of, say, $Y(1)$ among switchers in the treatment group $\times$ period 1 cell
- Let's start from the distribution of Y among all treated observations of this cell: both switchers and units already treated at T=0
- We need to withdraw from this distribution, that of units treated at $T = 0$, but it's not observed
- To construct it, we can apply the quantile-quantile transform from period 0 to 1 among treated observations in the control group to the distribution of Y among treated units in the treatment group in period 0.
- In a way, this quantile-quantile transform uses a double-matching to reconstruct the unobserved distribution

Fuzzy CIC: Intuition

- Consider a treated unit in the treatment group $\times$ period 0 cell. She is 1rst matched to a treated unit in the control group $\times$ period 0 cell with same y
- Those two units are observed at the same period of time and are both treated.
- Under Assumption 5 they must have the same u_1
- Second, the control $\times$ period 0 unit is matched to her rank counterpart among treated units of the control group $\times$ period 1 cell
- Denote by y^* the outcome of this last observation
- $U_1 \perp T | G, D(0) = 1 \implies$ those two observations must also have the same u_1
- So $y^* = h_1(u_1, 1)$: outcome that treatment $\times$ period 0 cell unit would have obtained in period 1

W_{CIC} can be re-written as

$$W_{CIC} = \frac{E(Y|G=1,T=1) - E((1-D)Q_0(Y) + DQ_1(Y)|G=1,T=0)}{E(D|G=1,T=1) - E(D|G=1,T=0)}$$

- Here again, W_{CIC} is almost the standard Wald ratio in the treatment group with T as the instrument
- Except that we have $(1-D)Q_0(Y) + DQ_1(Y)$ instead of Y in second term
- it accounts for the fact that time directly affects the outcome, just as $(1-D)\delta_0 + D\delta_1$ does in the W_{TC} estimand
- Under Assumption 4, the trends affecting the outcome are identified by additive shifts
- Under Assumption 5 and 6, they are identified by possibly non-linear quantile-quantile transforms

Next session

- DID with Variation in Treatment Timing
- Read: Goodman-Bacon, Andrew. "Difference-in-differences with variation in treatment timing." Journal of Econometrics (2021)
- optional: De Chaisemartin, Clement and Xavier D'Haultfoeuille. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects". AER(2020)