Identification and Estimation of Causal Intensive Margin Effects by Difference-in-Difference Methods

Markus Hersche and Elias Moor

IRME Retreat

April 14, 2020

Introduction

Introduction

- Causal effect of binary treatment on outcome with corner solutions at zero.
- Examples: Working hours, health expenditures, trade volumes.
- Decomposition of treatment effect into **extensive** and **intensive margin effect**.

Introduction

Why decomposition?
Example: effect of introduction of partial retirement policy

- Status quo: individuals must withdraw full pension at given age.
- Partial retirement policy: individuals choose between partial and full pension.
- Total effect on labor supply might be zero $\Rightarrow$ ineffective policy?
- Maybe positive extensive margin effect was offset by negative intensive margin effect:
    - Older workers who would have retired in absence of partial retirement policy, now decide to stay in labor market.
    - Individuals who would have worked full-time in absence of partial retirement policy, now work part-time.

Introduction

- Estimating intensive margin effect is challenging, even if treatment is randomly assigned.
- Problem: conditioning on sample with positive outcome induces selection problem
  $\Rightarrow$ mean comparison of treatment and control groups does not identify causal intensive margin effect (without additional assumptions).

Introduction

- Sample of individuals with positive outcome consists of three groups:
    1. Group of individuals with positive outcome irrespective whether they are treated or not.
    2. Group of individuals with positive outcome only because they are treated (who would have zero outcome if they were not treated).
    3. Group of individuals with positive outcome only because they are not treated (who would have zero outcome if they were treated).
- For causal intensive margin effect, only interested in first group.
- Unobserved characteristics of groups are likely different.
- Working hours example: individuals in first group might be more motivated $\Rightarrow$ higher average hours.
- Conditional on positive working hours, difference in means of treated and untreated could be difference in two groups, and not because of causal effect $\Rightarrow$ selection problem.

Introduction

- In general setting without random treatment $\Rightarrow$ two selection problems.
    1. Standard selection problem in observational studies due to confounding variables.
    2. Selection problem due to conditioning on positive outcomes.
- Difference-in-difference (DiD) methods developed to deal with first selection problem.
- Idea: extend difference-in-difference to include second selection problem as well.

Introduction

- This paper: use difference-in-difference to estimate causal intensive margin effect.
- Compared to standard difference-in-difference: condition sample on individuals with positive outcomes.
- Derive sufficient conditions under which the causal intensive margin effect is identified.
- Main difference to standard DiD: monotonicity assumptions are additionally required to identify the causal intensive margin effect.

Outline

1. Introduction
2. Literature review
3. Notation and Decomposition of a Treatment Effect
4. Identification
5. Estimation and Inference
6. Empirical Application: Causal Effect of Reaching the Full Retirement Age on Working Hours
7. Conclusion

Literature review

Literature review

- Models for outcomes with corner solutions:
    - Tobit (McDonald & Moffitt, 1980; Tobin, 1958)
    - Two-part models (Cragg, 1971; Duan, Manning, Morris, & Newhouse, 1983)
    - Selection models (Heckman, 1979)
- Literature employing *principal stratification* (Frangakis & Rubin, 2002) to study causal extensive and intensive margin treatment effects for variables with non-negative outcomes (Lee, 2012, 2017; Staub, 2014).
- Paper often follows Lechner (2010): survey on (standard) difference-in-differences methods from a potential outcomes perspective.

Notation and Decomposition of a Treatment Effect

Notation

- Standard potential outcomes framework with non-negative outcome Y and binary treatment D, extended to two periods.
- Observe individuals in pre-treatment period $t-1$ and in post-treatment period t.
- Each individual i has the following potential outcomes:
    - $Y^1_{i,t}$: potential outcome in period t in case of treatment ($D_i = 1$).
    - $Y^0_{i,t}$: potential outcome in period t in case of no treatment ($D_i = 0$).
    - $Y^1_{i,t-1}$: potential outcome in period $t-1$ in case of treatment.
    - $Y^0_{i,t-1}$: potential outcome in period $t-1$ in case of no treatment.
- Observed outcomes without superscript: $Y_{i,t}$, $Y_{i,t-1}$ (in each period only one of the two potential outcomes observed).
- Vector of observed covariates X_i, constant over time.
- Object of interest: average treatment effect on the treated:

$$ATT_t = E(Y^1_{i,t} - Y^0_{i,t} | D_i = 1)$$

Causal Decomposition

- Following Lee (2012) and Staub (2014), define four exhaustive and mutually exclusive subgroups based on joint distribution of potential outcomes:

	$Y^0_{i,t} = 0$	$Y^0_{i,t} > 0$
$Y^1_{i,t} = 0$	Nonparticipants	Switchers 2
$Y^1_{i,t} > 0$	Switchers 1	Participants

- Based on this definition, decompose the average treatment effect on the treated (ATT) at time t as follows:

$$
\begin{aligned}
ATT_t &= E(Y^1_{i,t} - Y^0_{i,t} | D_i = 1) \\
&= E(Y^1_{i,t} | Y^1_{i,t} > 0, Y^0_{i,t} = 0, D_i = 1) P(Y^1_{i,t} > 0, Y^0_{i,t} = 0 | D_i = 1) \\
&- E(Y^0_{i,t} | Y^1_{i,t} = 0, Y^0_{i,t} > 0, D_i = 1) P(Y^1_{i,t} = 0, Y^0_{i,t} > 0 | D_i = 1) \\
&+ E(Y^1_{i,t} - Y^0_{i,t} | Y^1_{i,t} > 0, Y^0_{i,t} > 0, D_i = 1) P(Y^1_{i,t} > 0, Y^0_{i,t} > 0 | D_i = 1)
\end{aligned}
$$

Identification

Identification

- Interested in intensive margin average treatment effect on the treated (IMATT):

$$IMATT_t = E(Y^1_{i,t} - Y^0_{i,t} | Y^1_{i,t} > 0, Y^0_{i,t} > 0, D_i = 1)$$

- Difference-in-difference estimator on positive outcomes:

$$\begin{aligned}\gamma^{DiD}_t(x) = & E(Y_{i,t} - Y_{i,t-1} | Y_{i,t} > 0, Y_{i,t-1} > 0, D_i = 1, X_i = x) \\ & - E(Y_{i,t} - Y_{i,t-1} | Y_{i,t} > 0, Y_{i,t-1} > 0, D_i = 0, X_i = x).\end{aligned}$$

- Goal: derive sufficient conditions under which $\gamma^{DiD}_t(x)$ identifies $IMATT_t$.

Identification

Identifying assumptions

1. SUTVA
2. No pre-treatment effect
3. Common trend in positive outcomes
4. No effect of treatment on covariates
5. Common support
6. Treatment monotonicity at extensive margin
7. Time monotonicity at extensive margin

Identification

SUTVA:

$$Y_{i,t} = (1 - D_i)Y^0_{i,t} + D_i Y^1_{i,t} \quad \forall i, \;\; \text{and}$$
$$Y_{i,t-1} = (1 - D_i)Y^0_{i,t-1} + D_i Y^1_{i,t-1} \quad \forall i.$$

- Ensures that we actually observe potential outcome in the treatment and control group.
- Implies that observed outcome of individual i only depends on potential outcomes and treatment status D_i, but not on treatment status D_j of any other individual j.
- Rules out general equilibrium effects and spill-over effects.

Identification

No pre-treatment effect:

$$\underbrace{E(Y^1_{i,t-1} - Y^0_{i,t-1} | Y^1_{i,t} > 0, Y^0_{i,t} > 0, D_i = 1, X_i = x)}_{\gamma_{t-1}(x)} = 0$$

- Requires that treatment effect in pre-treatment period is zero.
- In expectation, individuals do not change their behavior in period $t-1$ because they will be treated resp. not treated between period $t-1$ and t.

Identification

Common trend in positive outcomes:

$$E(Y^0_{i,t} - Y^0_{i,t-1} | Y_{i,t} > 0, Y_{i,t-1} > 0, D_i = 1, X_i = x)$$
$$= E(Y^0_{i,t} - Y^0_{i,t-1} | Y_{i,t} > 0, Y_{i,t-1} > 0, D_i = 0, X_i = x)$$

- Key assumption in difference-in-difference.
- Closely related to standard common trend assumption, except that common trend has to hold in subgroup of observations with positive outcome in period t and $t-1$.
- Requires that treated and control group would experience same time trend in case of no treatment.

Identification

No effect of treatment on covariates:

$$X_i^1 = X_i^0 = X_i \quad \forall i$$

- Ensures that conditioning on X does not condition away parts of causal effect we are interested in.

Identification

Common support:

$$P(D_i = 1 | Y_{i,t} > 0, Y_{i,t-1} > 0, X_i = x) < 1$$

- Requires that in each covariate cell, there are not only treated individuals in subsample with positive outcomes in period t and $t-1$.

Identification

Treatment monotonicity at extensive margin:

$$\begin{aligned} Y^1_{i,t} > 0 \quad &\Rightarrow \quad Y^0_{i,t} > 0 \quad \forall i, \quad \text{or} \\ Y^0_{i,t} > 0 \quad &\Rightarrow \quad Y^1_{i,t} > 0 \quad \forall i. \end{aligned}$$

- Positive outcome in case of treatment implies positive outcome in case of no treatment (or vice versa).
- Treatment response monotone with respect to extensive margin decision.
- Only restricts sign of extensive margin effect.

Identification

Time monotonicity at extensive margin:

$$\begin{aligned} Y^0_{i,t} > 0 \quad &\Rightarrow \quad Y^0_{i,t-1} > 0 \quad \forall i, \quad \text{and} \\ Y^1_{i,t} > 0 \quad &\Rightarrow \quad Y^1_{i,t-1} > 0 \quad \forall i. \end{aligned}$$

- Positive outcome in period t implies positive outcome in period $t-1$ (both in the case of treatment and no treatment).
- No individuals with positive outcome in period t who have a zero outcome in period $t-1$.
- Only restricts sign of extensive margin effect.

Estimation and Inference

Estimation

- Difference-in-difference on positive outcomes requires estimating conditional expectations:

$$\begin{aligned}\gamma_t^{DiD}(x) =& E(Y_{i,t} - Y_{i,t-1} | Y_{i,t} > 0, Y_{i,t-1} > 0, D_i = 1, X_i = x) \\ & - E(Y_{i,t} - Y_{i,t-1} | Y_{i,t} > 0, Y_{i,t-1} > 0, D_i = 0, X_i = x).\end{aligned}$$

- Define $\Delta Y_{i,t} = Y_{i,t} - Y_{i,t-1}$.
- First estimate:

$$m_1(x) = E(\Delta Y_{i,t} | Y_{i,t} > 0, Y_{i,t-1} > 0, D_i = 1, X_i = x) \text{ , and}$$
$$m_0(x) = E(\Delta Y_{i,t} | Y_{i,t} > 0, Y_{i,t-1} > 0, D_i = 0, X_i = x) \text{ ,}$$

using ordinary least squares.

Estimation

- Using fitted functions $\widehat{m_1}(x)$ and $\widehat{m_0}(x)$, calculate fitted values $\widehat{m_1}(X_i)$ and $\widehat{m_0}(X_i)$.
- Intensive margin average treatment effect on the treated:

$$\widehat{IMATT}_t^{DID} = \frac{1}{N_T} \sum_{\substack{i:Y_{i,t}>0, \\ Y_{i,t-1}>0, \\ D_i=1}} \left[\widehat{m_1}(X_i) - \widehat{m_0}(X_i) \right],$$

where N_T is the number of treated observations with positive outcome in period t and $t-1$.

Inference

- Nonparametric quantile bootstrap (Efron & Tibshirani, 1993).
- From sample of observations with positive outcomes in period t and $t-1$, repeatedly draw bootstrap sample of same sample size.
- In bootstrap sample, estimate IMATT as described.
- Yields distribution of bootstrap estimated IMATT's: $\widehat{IMATT}_t^1, \ldots, \widehat{IMATT}_t^B$, where B is number of bootstrap replications.
- Construct bootstrap estimated confidence interval as

$$[q^*_{\alpha/2}, q^*_{1-\alpha/2}] \,,$$

where $q^*_{1-\alpha/2}$ is the $(1-\alpha/2)$-percentile of distribution of bootstrap estimated IMATT's.

Empirical Application: Causal Effect of Reaching the Full Retirement Age on Working Hours

Empirical Application

- Apply difference-in-difference methodology to estimate causal intensive margin effect of reaching the full retirement age (FRA) on working hours of women.
- Data from the Swiss Labor Force Survey (SLFS).
- Exploit pension reform in Switzerland taking place in 2004, increasing FRA of women from 63 to 64.
- Women with year of birth 1941 or earlier reach FRA at age 63, women with year of birth 1942 or later reach FRA at age 64.

Empirical Application

- Outcome of interest is working hours, denoted by $Y_{i,t}$.
- Restrict sample to women aged 63
  $\Rightarrow D_i = 1$ for women who have reached FRA (year of birth 1941 or earlier)
  $\Rightarrow D_i = 0$ for women who have not reached FRA (year of birth 1942 or later).
- Since reform affects individuals only based on year of birth, assignment to treatment assumed to be almost random.
- Covariates X: categorical education variable and dummy for being Swiss citizen.

Empirical Application: Discussion of Assumptions

Identifying assumptions

1. SUTVA (untestable)
2. No pre-treatment effect (untestable)
3. Common trend in positive outcomes (untestable)
4. No effect of treatment on covariates (untestable)
5. Common support (testable)
6. Treatment monotonicity at extensive margin (untestable)
7. Time monotonicity at extensive margin (testable)

Empirical Application: Discussion of Assumptions

SUTVA:

$$Y_{i,t} = (1 - D_i)Y^0_{i,t} + D_i Y^1_{i,t} \quad \forall i, \ \text{ and}$$
$$Y_{i,t-1} = (1 - D_i)Y^0_{i,t-1} + D_i Y^1_{i,t-1} \quad \forall i.$$

- Evidence for spillover within couples (Cribb, Emmerson, & Tetlow, 2013; Lalive & Parrotta, 2017; Stancanelli, 2017).
- Potential threat to identification, but assumed to be negligible.

Empirical Application: Discussion of Assumptions

No pre-treatment effect:

$$\underbrace{E(Y^1_{i,t-1} - Y^0_{i,t-1} | Y^1_{i,t} > 0, Y^0_{i,t} > 0, D_i = 1, X_i = x)}_{\gamma_{t-1}(x)} = 0$$

- Assumption rules out that people change their working hours in anticipation of reaching FRA in the next period.
- Compare mean working hours in period $t-1$, conditional on having positive working hours in period t and $t-1$.
- Mean in control group: 23.8 hours, mean in treatment group: 24.6 hours (p-value t-test: 0.54).

Empirical Application: Discussion of Assumptions

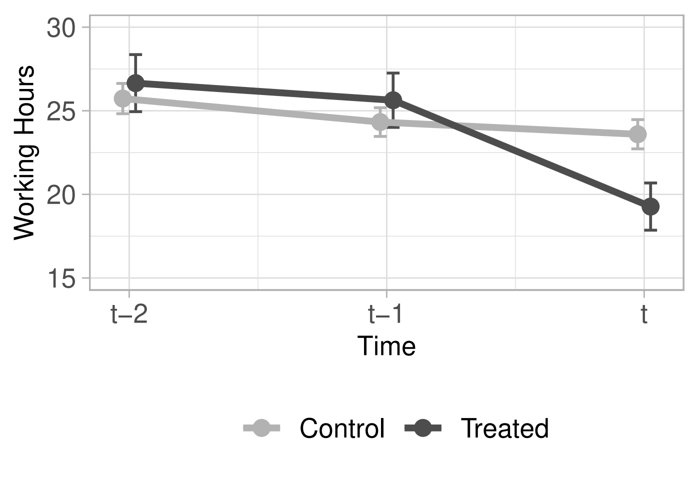
Common trend in positive outcomes:

$$E(Y^0_{i,t} - Y^0_{i,t-1} | Y_{i,t} > 0, Y_{i,t-1} > 0, D_i = 1, X_i = x)$$
$$= E(Y^0_{i,t} - Y^0_{i,t-1} | Y_{i,t} > 0, Y_{i,t-1} > 0, D_i = 0, X_i = x)$$

- Motivate assumption with pre-treatment trends of control and treatment group.

Empirical Application: Discussion of Assumptions

Common trend in positive outcomes:



Empirical Application: Discussion of Assumptions

No effect of treatment on covariates:

$$X_i^1 = X_i^0 = X_i \quad \forall i$$

- Covariates X: categorical education variable and dummy for being a Swiss citizen.
- Unlikely that reaching FRA has an (non-negligible) effect on these variables.

Empirical Application: Discussion of Assumptions

Common support:

$$P(D_i = 1 | Y_{i,t} > 0, Y_{i,t-1} > 0, X_i = x) < 1$$

- Can be tested: in each covariate cell, check fraction of treated observations.

Empirical Application: Discussion of Assumptions

Common support:

$$P(D_i = 1 | Y_{i,t} > 0, Y_{i,t-1} > 0, X_i = x) < 1$$

Secondary Education	Higher Education	Swiss citizen	Fraction Treated
0	0	0	0.091
0	0	1	0.308
0	1	0	0.350
0	1	1	0.262
1	0	0	0.154
1	0	1	0.298

Table 1: Analysis of Common Support

Empirical Application: Discussion of Assumptions

Treatment monotonicity at extensive margin:

$$\begin{aligned} Y^1_{i,t} > 0 \;\; &\Rightarrow \;\; Y^0_{i,t} > 0 \quad \forall i, \;\; \text{or} \\ Y^0_{i,t} > 0 \;\; &\Rightarrow \;\; Y^1_{i,t} > 0 \quad \forall i. \end{aligned}$$

- Rules out that people start to work because they reach FRA.
- Some incentives exist to take up a job after reaching FRA (part of earnings are exempted from social security contributions $\Rightarrow$ increases net wage).
- More plausible that reaching FRA will either have no effect or drive people out of the labor market.

Empirical Application: Discussion of Assumptions

Time monotonicity at extensive margin:

$$Y^0_{i,t} > 0 \quad \Rightarrow \quad Y^0_{i,t-1} > 0 \quad \forall i, \quad \text{and}$$
$$Y^1_{i,t} > 0 \quad \Rightarrow \quad Y^1_{i,t-1} > 0 \quad \forall i.$$

- Can be tested: in treated and control subsample, calculate fraction of individuals with positive working hours in period t, conditional on not working in period $t-1$.
- In estimation sample, 5% of the treated and 7.4% of the control sample returned to work after having not worked in the period before.
- Potential threat to identification.
- Overall, people aged 63 rather leave labor force as they become older.

Empirical Application: Estimation Results

	YoB 1941-42	YoB 1939-46
IMATT FRA reached	-5.04	-4.354
95% C.I.	[-8.57, -1.63]	[-6.73, -2.07]
Obs. (treat/cont)	63/87	156/405

Table 2: Results difference-in-difference on positive outcomes

Conclusion

Conclusion

- Paper extends literature on identification and estimation of causal intensive margin effects to difference-in-difference.
- Intensive margin effects interesting when total effect masks relevant sub-effects, e.g. when extensive and intensive margin effects have different sign.
- Derive sufficient conditions under which the difference-in-differences estimator on positive outcomes identifies the causal intensive margin effect.
- Compared to standard difference-in-difference estimator, time and treatment monotonicity at extensive margin are additionally required.

Thank you!

References

Cragg, J. (1971). Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica*, *39*(5), 829–844.

Cribb, J., Emmerson, C., & Tetlow, G. (2013). Incentives, shocks or signals: labour supply effects of increasing the female state pension age in the UK. *IFS Working Paper*, *W13/03*.

Duan, N., Manning, W. G., Morris, C. N., & Newhouse, J. P. (1983). A comparison of alternative models for the demand for medical care. *Journal of Business & Economic Statistics*, 115–126.

Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall/CRC.

Frangakis, C. E., & Rubin, D. B. (2002). Principal Stratification in Causal Inference. *Biometrics*, *58*(1), 21–29.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, *47*(1), 153–161.

Lalive, R., & Parrotta, P. (2017, 6). How does pension eligibility affect labor supply in couples? *Labour Economics*, *46*, 177–188.

Lechner, M. (2010). The Estimation of Causal Effects by Difference-in-Difference Methods. *Foundations and Trends in Econometrics*, 4(3), 165–224.

Lee, M.-j. (2012). Treatment Effects in Sample Selection Models and Their Nonparametric Estimation. *Journal of Econometrics*, *167*(2), 317–329.

Lee, M.-j. (2017). Extensive and intensive margin effects in sample selection models : racial effects on wages. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *180*, 817–839.

McDonald, J. F., & Moffitt, R. A. (1980). The Uses of Tobit Analysis. *The Review of Economics and Statistics*, *62*(2), 318–321.

Stancanelli, E. (2017). Couples' retirement under individual pension design: A regression discontinuity study for France. *Labour Economics*, *49*, 14–26.

Staub, K. (2014). A causal interpretation of extensive and intensive margin effects in generalized tobit models. *The Review of Economics and Statistics*, *96*(May), 371–375.

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica*, *26*(1), 24–36.

Appendix

Conventional decomposition

Because treatment is random, ATT at time t is identified by the difference in mean outcomes of treated and untreated:

$$
\begin{aligned}
ATT_t &= E(Y^1_{i,t}|D_i = 1) - E(Y^0_{i,t}|D_i = 1) \\
&= E(Y_{i,t}|D_i = 1) - E(Y_{i,t}|D_i = 0)
\end{aligned}
$$

A non-negative outcome (with a point mass at zero) is often decomposed into an extensive and an intensive part, i.e. $E(Y_{i,t}) = E(Y_{i,t}|Y_{i,t} > 0)P(Y_{i,t} > 0)$. Similar to Staub (2014), the difference in mean outcomes can then be rewritten as

$$
\begin{aligned}
ATT_t =& E(Y_{i,t}|D_i = 1) - E(Y_{i,t}|D_i = 0) \\
=& E(Y_{i,t}|Y_{i,t} > 0, D_i = 1)P(Y_{i,t} > 0|D_i = 1) \\
&- E(Y_{i,t}|Y_{i,t} > 0, D_i = 0)P(Y_{i,t} > 0|D_i = 0) \\
=& \left[P(Y_{i,t} > 0|D_i = 1) - P(Y_{i,t} > 0|D_i = 0)\right] E(Y_{i,t}|Y_{i,t} > 0, D_i = 1) \\
&+ \left[E(Y_{i,t}|Y_{i,t} > 0, D_i = 1) - E(Y_{i,t}|Y_{i,t} > 0, D_i = 0)\right] P(Y_{i,t} > 0|D_i = 0) \,.
\end{aligned}
$$

Conventional decomposition

Under random treatment, the observed difference can be rewritten in terms of potential outcomes,

$$
\begin{aligned}
ATT_t = & \big[P(Y^1_{i,t} > 0) - P(Y^0_{i,t} > 0)\big] E(Y^1_{i,t} | Y^1_{i,t} > 0) \\
& + \big[E(Y^1_{i,t} | Y^1_{i,t} > 0) - E(Y^0_{i,t} | Y^0_{i,t} > 0)\big] P(Y^0_{i,t} > 0) \, .
\end{aligned}
$$

Identification

- By the law of iterated expectations, $IMATT_t$ can be rewritten as:

$$IMATT_t = E(Y_{i,t}^1 - Y_{i,t}^0 | Y_{i,t}^1 > 0, Y_{i,t}^0 > 0, D_i = 1)$$
$$= E\left[\underbrace{E(Y_{i,t}^1 - Y_{i,t}^0 | Y_{i,t}^1 > 0, Y_{i,t}^0 > 0, D_i = 1, X_i = x)}_{\gamma_t(x)} \middle| Y_{i,t}^1 > 0, Y_{i,t}^0 > 0, D_i = 1\right]$$

- First step: derive sufficient conditions under which conditional-on-X IMATT $\gamma_t(x)$ is identified.
- Second step: state conditions under which conditional-on-X IMATT can be aggregated to IMATT, i.e. to $E(Y_{i,t}^1 - Y_{i,t}^0 | Y_{i,t}^1 > 0, Y_{i,t}^0 > 0, D_i = 1)$.

Identification simple difference estimator on positive outcomes

The *simple difference estimator on positive outcomes* is given by

$$\gamma_t^D(x) = E(Y_{i,t}|Y_{i,t} > 0, D_i = 1, X_i = x) - E(Y_{i,t}|Y_{i,t} > 0, D_i = 0, X_i = x).$$

Identifying assumptions

1. *SUTVA*,
2. *no effect of treatment on covariates*,
3. *common support*,
4. *unconfoundedness*, and

5. *no switchers*,

Or

5. *conditional mean independence.*

Identification simple difference estimator on positive outcomes

Unconfoundedness:

$$(Y^1_{i,t}, Y^0_{i,t}) \perp\!\!\!\perp D_i \mid X_i \ .$$

No Switchers:

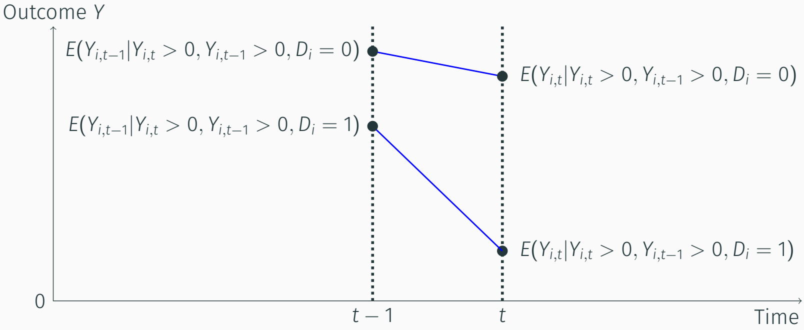
$$Y^1_{i,t} > 0 \quad \Leftrightarrow \quad Y^0_{i,t} > 0 \quad \forall i.$$

Conditional Mean Independence:

$$E(Y^1_{i,t}|Y^1_{i,t} > 0, Y^0_{i,t} = 0, D_i = 1, X_i = x) = E(Y^1_{i,t}|Y^1_{i,t} > 0, Y^0_{i,t} > 0, D_i = 1, X_i = x),$$
$$E(Y^0_{i,t}|Y^1_{i,t} = 0, Y^0_{i,t} > 0, D_i = 1, X_i = x) = E(Y^0_{i,t}|Y^1_{i,t} > 0, Y^0_{i,t} > 0, D_i = 1, X_i = x).$$

Graphical derivation

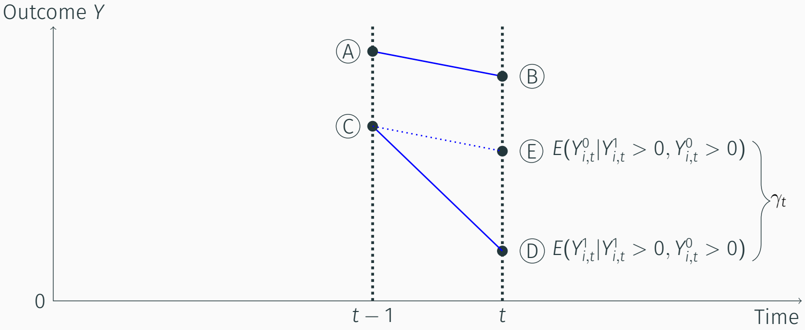
Figure 1: Graphical derivation of the *difference-in-difference estimator on positive outcomes*



Note: Observed quantities are shown in the upper graph. Causal intensive margin effect is shown in the lower graph.

Graphical derivation

Figure 2: Graphical derivation of the *difference-in-difference estimator on positive outcomes*



Note: Observed quantities are shown in the upper graph. Causal intensive margin effect is shown in the lower graph.