EPE - Lecture 2
Natural Experiments

Sylvain Chabé-Ferret

Toulouse School of Economics, Inra

February 2016

In a nutshell

In this lecture, we are going to study how to estimate the effect of an intervention on an outcome using natural experiments, that is methods that try to correct for selection bias by finding situations analog to randomized experiments but due to the natural course of events.

The methods covered

- Regression Discontinuity Designs (RDD)
- Difference in Differences (DID)
- Instrumental Variables (IV)

Outline

RDD: Basic Intuition

RDD uses situations where there is a discontinuity in the probability of receiving the treatment. If there is also a discontinuity in outcomes, it is interpreted as the effect of the treatment.

Sharp and Fuzzy RDD

We distinguish two RD Designs:

- Sharp Designs (probability transitions from 0 to 1)
- Fuzzy Designs (probability transitions from values strictly between 0 and 1)

Sharp RDD Design: Formal Definition

Assumption (Sharp RDD Design)

There exists a running variable Z_i and a threshold $\bar{z}$ such that:

$$D_i = \mathbb{1}[Z_i \leq \bar{z}].$$

Sharp Design: Illustration

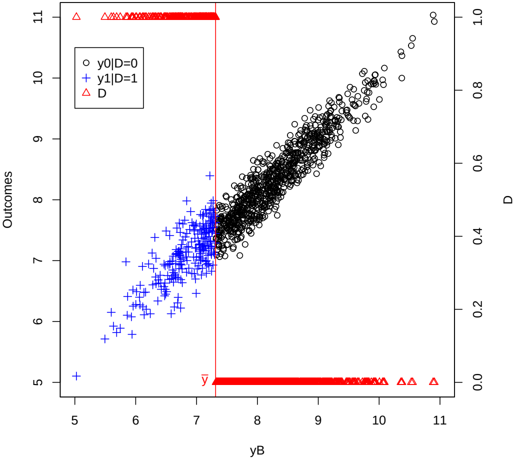

Figure: Sharp RDD Design

Key Assumption: Continuity

Assumption (Continuity of Expected Potential Outcomes)

For $d \in \{0, 1\}$,

$$\lim_{e \to 0^+} \mathbb{E}[Y_i^d | Z_i = \bar{z} - e] = \lim_{e \to 0^+} \mathbb{E}[Y_i^d | Z_i = \bar{z} + e].$$

Continuity: Illustration

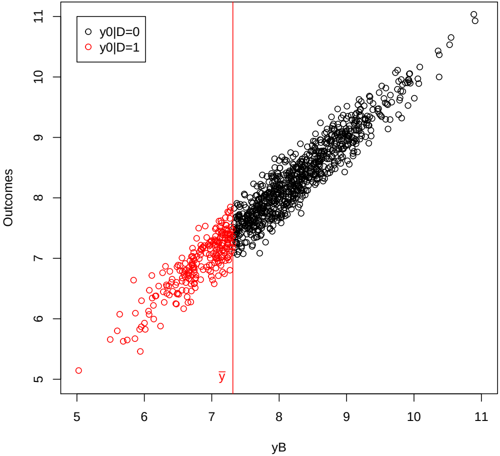

Figure: Continuity of $\mathbb{E}[y_i^0|y_i^B]$

Continuity: Illustration

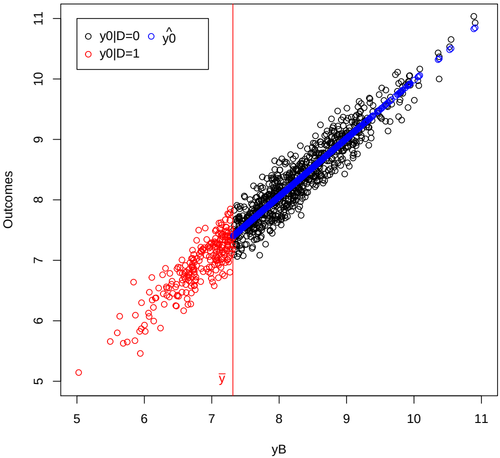

Figure: Continuity of $\mathbb{E}[y_i^0|y_i^B]$

Continuity: Illustration

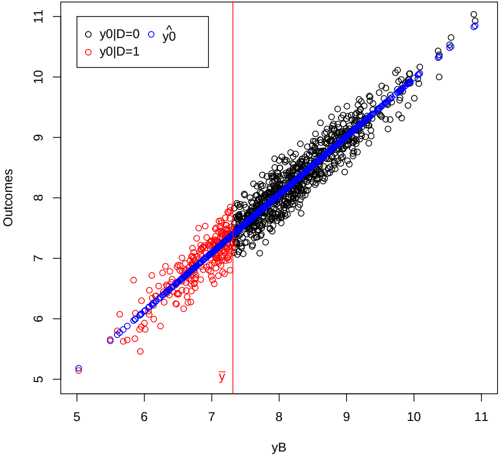

Figure: Continuity of $\mathbb{E}[y_i^0 | y_i^B]$

Continuity: Illustration

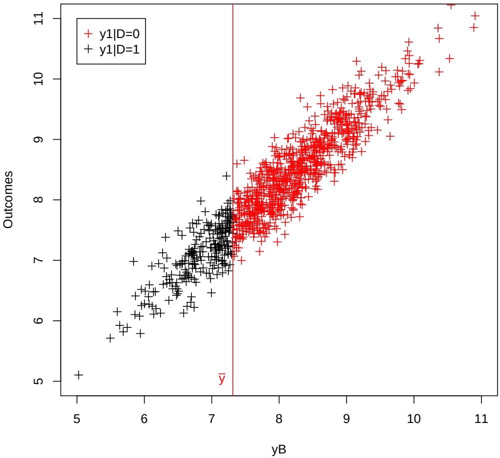

Figure: Continuity of $\mathbb{E}[y_i^1|y_i^B]$

Continuity: Illustration

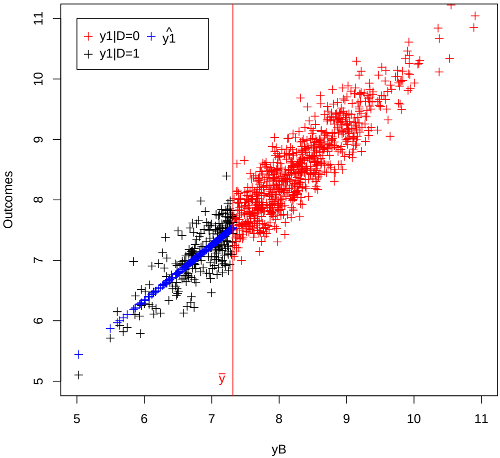

Figure: Continuity of $\mathbb{E}[y_i^1|y_i^B]$

Continuity: Illustration

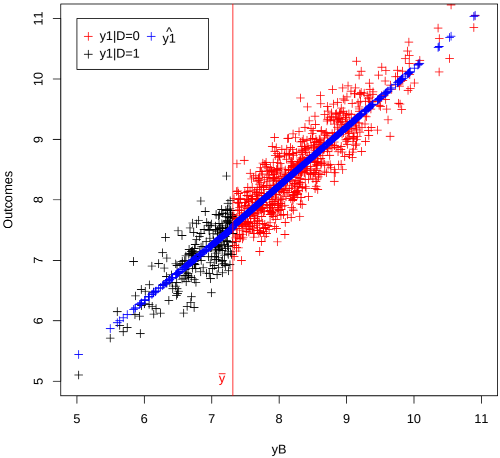

Figure: Continuity of $\mathbb{E}[y_i^1|y_i^B]$

Identification in a Sharp Design

Theorem (Identification in a Sharp RDD Design)

Under Assumptions Sharp RDD Design and Continuity of Expected Potential Outcomes, the Treatment Effect on the Treated is identified at $Z_i = \bar{z}$:

$$\Delta^Y_{TT}(\bar{z}) = \lim_{e \to 0^+} \mathbb{E}[Y_i | Z_i = \bar{z} - e] - \lim_{e \to 0^+} \mathbb{E}[Y_i | Z_i = \bar{z} + e],$$

where $\Delta^Y_{TT}(\bar{z}) = \mathbb{E}[\Delta^Y_i | Z_i = \bar{z}]$.

Proof

$$
\begin{aligned}
\lim_{e\to 0+}\mathbb{E}[Y_i|Z_i=\bar{z}-e]-\lim_{e\to 0+}\mathbb{E}[Y_i|Z_i=\bar{z}+e] &= \lim_{e\to 0+}\mathbb{E}[Y_i^1|Z_i=\bar{z}-e]-\lim_{e\to 0+}\mathbb{E}[Y_i^0|Z_i=\bar{z}+e] \\
&= \mathbb{E}[Y_i^1|Z_i=\bar{z}]-\mathbb{E}[Y_i^0|Z_i=\bar{z}] \\
&= \mathbb{E}[Y_i^1-Y_i^0|Z_i=\bar{z}] \\
&= \Delta^Y_{TT}(\bar{z}),
\end{aligned}
$$

where the first equality uses Sharp RDD Design and the second equality uses Continuity of Expected Potential Outcomes.

Identification in a Sharp Design: Illustration

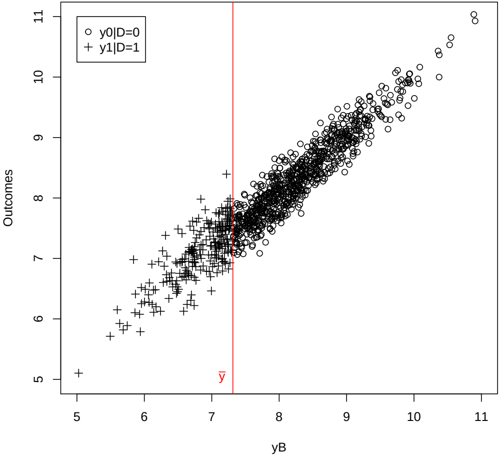

Figure: Identification in a sharp RDD design

Identification in a Sharp Design: Illustration

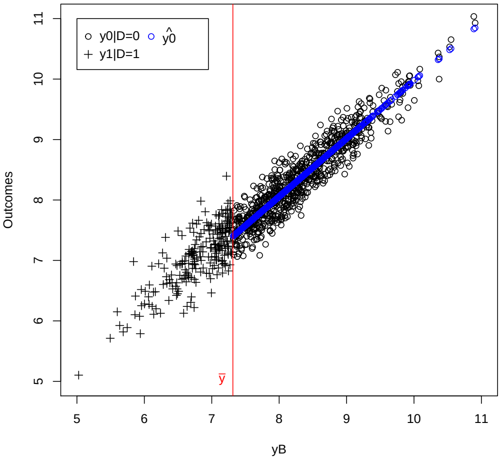

Figure: Identification in a sharp RDD design

Identification in a Sharp Design: Illustration

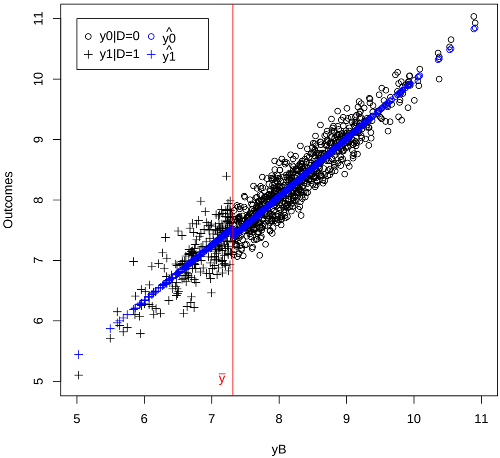

Figure: Identification in a sharp RDD design

Estimation Using OLS

If we assume linearity of the regression functions on each side of the cutoff, we can compute the predicted values at the cutoff using each regression line. The difference between them is the estimated treatment effect.

Estimation Using OLS: Illustration

- $\hat{\mathbb{E}}[y_i^1|y_i^B = \bar{y}] =$7.538
- $\hat{\mathbb{E}}[y_i^0|y_i^B = \bar{y}] =$7.3972

So the estimated TT is $\hat{\Delta^y_{RDDOLS}} =$0.1407.

Value of $\Delta^{y}_{TT}(\bar{z})$ in Our Example

$$\Delta^{y}_{TT}(\bar{z}) = \bar{\alpha} + \theta\bar{\mu} + \theta\frac{\sigma_{\mu}^{2}}{\sigma_{\mu}^{2} + \sigma_{U}^{2}}(\bar{y} - \bar{\mu}).$$

So $TT(z) = 0.1756$

Results of RDD Using OLS

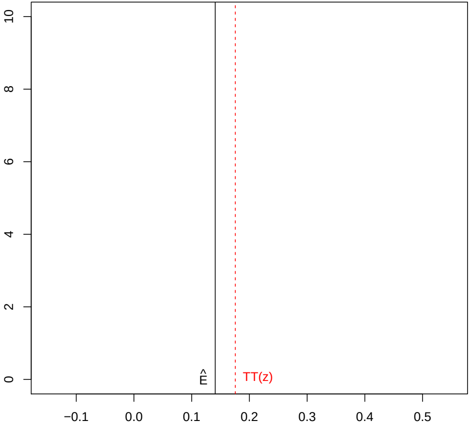

Figure: Sharp RDD Using OLS

Sampling Noise with RDD OLS: Illustration

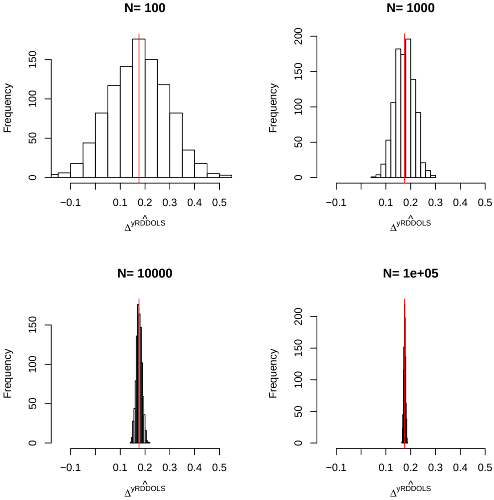

Figure: Distribution of the *RDDOLS* estimator over replications of samples of different sizes

RDD OLS and Nonlinear Outcome-Running Variable Curve

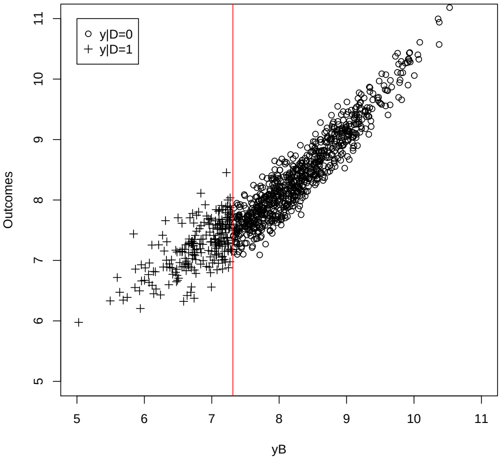

Figure: Non linear

RDD OLS and Nonlinear Outcome-Running Variable Curve

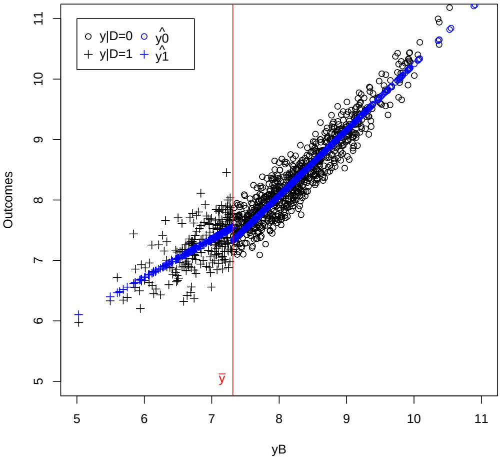

Figure: Non linear

RDD OLS and Nonlinear Outcome-Running Variable Curve

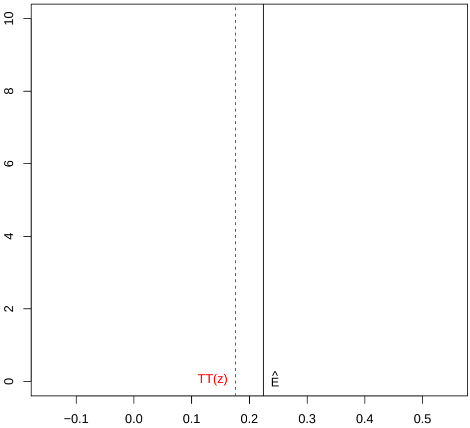

Figure: Bias of RDD OLS When Non Linear Curve

Estimation Using LLR

- Estimate $\hat{\mathbb{E}}[y_i^1 | y_i^B = \bar{y}] =$ using LLR on the left of $\bar{y}$.
- Estimate $\hat{\mathbb{E}}[y_i^0 | y_i^B = \bar{y}] =$ using LLR on the right of $\bar{y}$.
- $\hat{\Delta^y_{RDDLLR}} = \hat{\mathbb{E}}[y_i^1 | y_i^B = \bar{y}] - \hat{\mathbb{E}}[y_i^0 | y_i^B = \bar{y}]$.
- Bandwidth choice: use cross-validation on each side.

Bandwidth choice

- Use cross-validation on each side.
- Use special cros validation (Ludwig and Miller, 2005)
- Use a rule of thumb (Imbens and Kalyanaraman, 2011)

RDD LLR: Bandwidth Choice

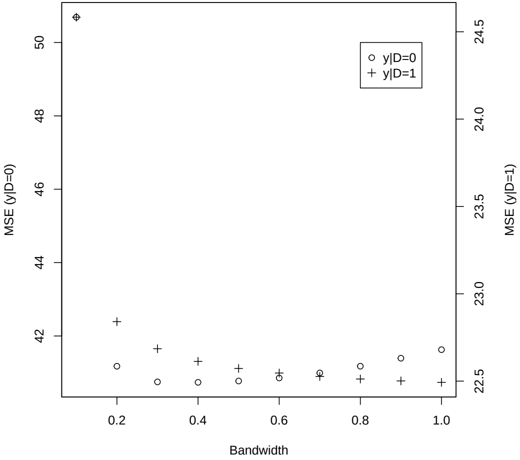

Figure: Cross Validation Results

RDD LLR

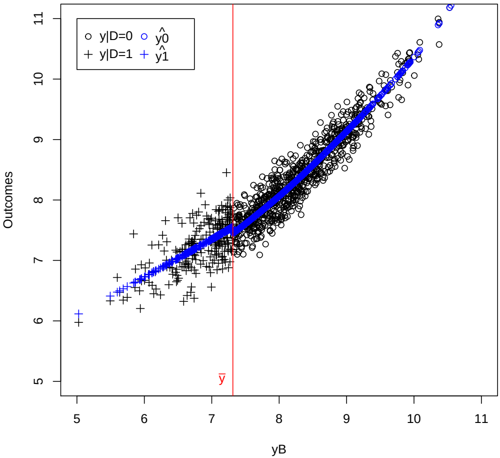

Figure: RDD LLR

RDD LLR and Nonlinear Outcome-Running Variable Curve

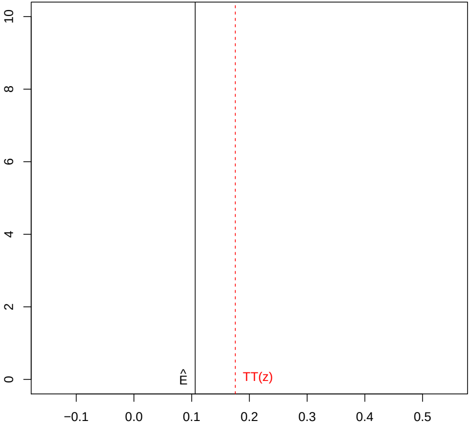

Figure: RDD LLR

Sampling Noise with RDD OLS: Illustration

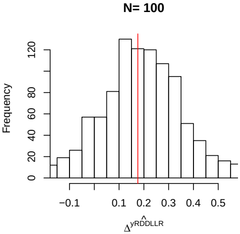

Figure: Distribution of the *RDDLLR* estimator over replications of samples of different sizes

Sampling Noise with RDD LLR: Illustration

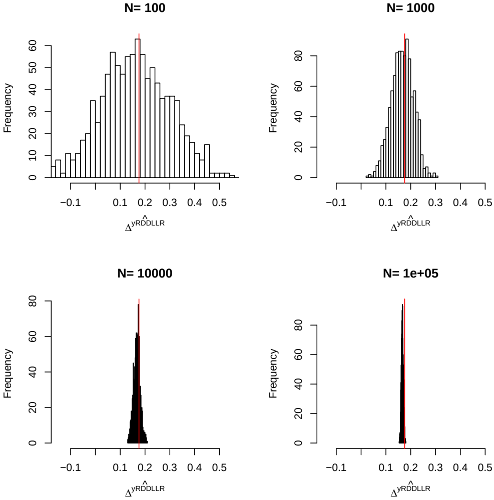

Figure: Distribution of the *RDDLLR* estimator over replications of samples of different sizes

Estimation Using Simplified LLR

Imbens and Lemieux propose the following simplified version of the LLR estimator: $\hat{\delta}$ estimated by OLS on the sample of observations such as $\bar{z} - h \leq Z_i \leq \bar{z} + h$ is an estimate of $TT(z)$:

$$Y_i = \alpha + \beta(Z_i - \bar{z})(1 - D_i) + \gamma(Z_i - \bar{z})D_i + \delta D_i + \epsilon_i$$

It is actually equal to the LLR estimate with uniform kernel and identical bandwidth on each side of the threshold.

Estimating Precision in the Sharp RDD Design

Several approaches:

1. Hahn, Todd and van der Klaauw (2001) derive general CLT results
2. Imbens and Lemieux (2008) simplify the CLT results and propose a plug-in estimator
3. Imbens and Lemieux (2008) propose to use the robust variance OLS estimator
4. Bootstrap should be valid

Asymptotic Variance of the Simplified LLR Estimator

Theorem (Asymptotic Variance of the LLR Estimator)

The variance of the simplified LLR Estimator in a Sharp Design can be approximated by:

$$\mathbb{V}[\hat{\Delta}_{LLRRDD}] \approx \frac{1}{\sqrt{Nh}} \frac{4}{f_Z(\bar{z})} \left(\lim_{e \to 0^+} \mathbb{V}[Y_i | Z_i = \bar{z} - e] + \lim_{e \to 0^+} \mathbb{V}[Y_i | Z_i = \bar{z} + e] \right),$$

with f_Z the density of Z_i.

Proof

Hahn, Todd and van der Klaauw (2001) and Imbens and Lemieux (2008).

Illustration of Imbens and Lemieux Simplified LLR

- The estimated value of TT(z) by simplified LLR is 0.1156
- The estimated 99% sampling noise is: 0.3304.
- The true 99% sampling noise of LLR estimated by Monte Carlo simulations is: 0.3396.

Fuzzy RDD Design: Formal Definition

Assumption (Fuzzy RDD Design)

There exists a running variable Z_i and a threshold $\bar{z}$ such that:

$$\lim_{e \to 0^+} \Pr(D_i = 1 | Z_i = \bar{z} - e) \neq \lim_{e \to 0^+} \Pr(D_i = 1 | Z_i = \bar{z} + e).$$

Fuzzy RDD: Illustration

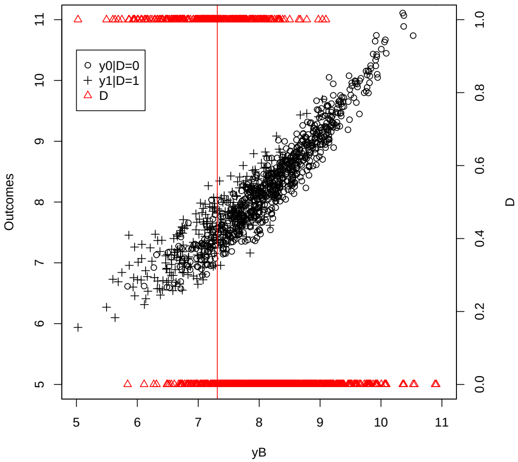

Figure: Fuzzy RDD

Fuzzy RDD: Illustration

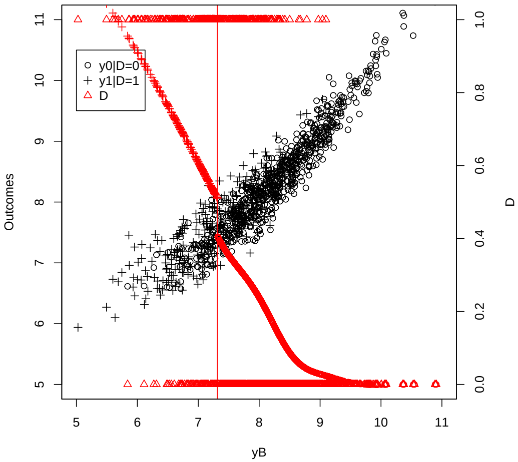

Figure: Fuzzy RDD

Fuzzy RDD: Illustration

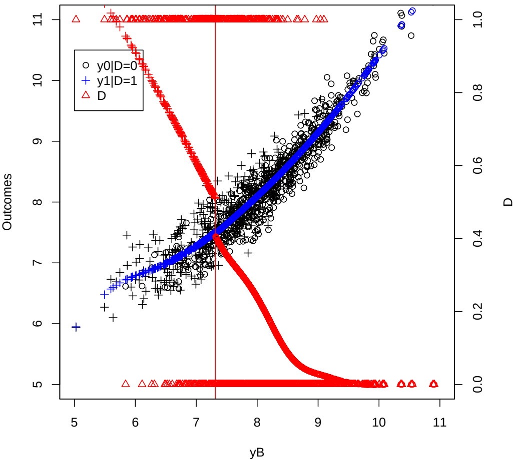

Figure: Fuzzy RDD

Wald Estimator

$$\hat{\Delta}^{Y}_{RDDWALD} = \frac{\lim_{e \to 0^+} \hat{\mathbb{E}}[Y_i | Z_i = \bar{z} - e] - \lim_{e \to 0^+} \hat{\mathbb{E}}[Y_i | Z_i = \bar{z} + e]}{\lim_{e \to 0^+} \hat{\Pr}(D_i = 1 | Z_i = \bar{z} - e) - \lim_{e \to 0^+} \hat{\Pr}(D_i = 1 | Z_i = \bar{z} + e)}$$

Wald Estimator: Illustration

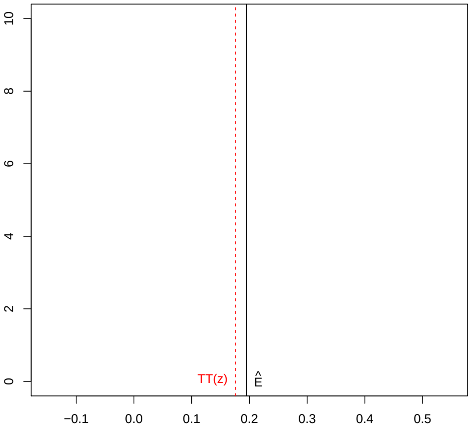

Figure: Wald RDD LLR

Assumption: Independent Treatment Effects

Assumption (Independent Treatment Effects)

$$\Delta_i^Y \perp\!\!\!\perp D_i | Z_i = \bar{z}.$$

Identification Under Independent Treatment Effects

Theorem (Identification of TT(z) in Fuzzy RDD)

Under Continuity of Expected Potential Outcomes and Independent Treatment Effects, we have:

$$\Delta^Y_{RDDWALD} = \Delta^Y_{TT}(z),$$

with

$$\Delta^Y_{RDDWALD} = \frac{\lim_{e\to 0^+}\mathbb{E}[Y_i|Z_i=\bar{z}-e] - \lim_{e\to 0^+}\mathbb{E}[Y_i|Z_i=\bar{z}+e]}{\lim_{e\to 0^+}\Pr(D_i=1|Z_i=\bar{z}-e) - \lim_{e\to 0^+}\Pr(D_i=1|Z_i=\bar{z}+e)}$$

Proof

$$\begin{aligned}
\lim_{z \to \bar{z}^+} \mathbb{E}[Y_i | Z_i = z] &= \lim_{z \to \bar{z}^+} \Big(\mathbb{E}[Y_i^1 | Z_i = z, D_i = 1] \Pr(D_i = 1 | Z_i = z) \\
&\quad + \mathbb{E}[Y_i^0 | Z_i = z, D_i = 0](1 - \Pr(D_i = 1 | Z_i = z)) \Big) \\
&= \lim_{z \to \bar{z}^+} \Big(\Big(\mathbb{E}[\Delta_i^Y + Y_i^0 | Z_i = z, D_i = 1] \Big) \Pr(D_i = 1 | Z_i = z) \\
&\quad + \mathbb{E}[Y_i^0 | Z_i = z, D_i = 0](1 - \Pr(D_i = 1 | Z_i = z)) \Big) \\
&= \Delta_{TT}^Y(\bar{z}) \lim_{z \to \bar{z}^+} \Pr(D_i = 1 | Z_i = z) \\
&\quad + \lim_{z \to \bar{z}^+} \mathbb{E}[Y_i^0 | Z_i = z] \\
&= \Delta_{TT}^Y(\bar{z}) \lim_{z \to \bar{z}^+} \Pr(D_i = 1 | Z_i = z) + \mathbb{E}[Y_i^0 | Z_i = \bar{z}] \\
\lim_{z \to \bar{z}^-} \mathbb{E}[Y_i | Z_i = z] &= \Delta_{TT}^Y(\bar{z}) \lim_{z \to \bar{z}^-} \Pr(D_i = 1 | Z_i = z) + \mathbb{E}[Y_i^0 | Z_i = \bar{z}].
\end{aligned}$$

Correlated Treatment Effects

- $D_i(z)$ potential outcome of individual i when $Z_i = z$
- Always takers ($T_i^{\bar{z}} = a$): $\lim_{z \to \bar{z}^+} D_i(z) = \lim_{z \to \bar{z}^-} D_i(z) = 1$
- Never takers ($T_i^{\bar{z}} = n$): $\lim_{z \to \bar{z}^+} D_i(z) = \lim_{z \to \bar{z}^-} D_i(z) = 0$
- Compliers ($T_i^{\bar{z}} = c$): $\lim_{z \to \bar{z}^+} D_i(z) - \lim_{z \to \bar{z}^-} D_i(z) = 1$
- Defiers ($T_i^{\bar{z}} = d$): $\lim_{z \to \bar{z}^+} D_i(z) - \lim_{z \to \bar{z}^-} D_i(z) = -1$

Assumptions Under Correlated Treatment Effects

Assumption (Monotonicity)

$D_i(z)$ is non-decreasing at $z = \bar{z}$ (or $\Pr(T_i^{\bar{z}} = d | Z_i = \bar{z}) = 0$).

Assumption (Continuity of Expected Potential Outcomes Conditional on Types)

$\mathbb{E}[Y_i^1 | Z_i = z, T_i^z]$ and $\mathbb{E}[Y_i^0 | Z_i = z, T_i^z]$ are continuous (at $z = \bar{z}$).

Wald Identifies LATE

Theorem (Wald Identifies LATE)

Under Monotonicity and Continuity of Expected Potential Outcomes Conditional on Type:

$$\Delta^Y_{RDDWALD} = \Delta^Y_{LATE}(z),$$

with

$$\Delta^Y_{LATE}(z) = \mathbb{E}[\Delta^Y_i \,|\, T^{\bar{z}}_i = c, Z_i = \bar{z}].$$

proof

$$
\begin{aligned}
\lim_{z \to \bar{z}^+} \mathbb{E}[Y_i | Z_i = z] = \mathbb{E}[Y_i^1 | Z_i = \bar{z}, T_i^{\bar{z}} = a] \Pr(T_i = a | Z_i = \bar{z}) \\
+ \mathbb{E}[Y_i^1 | Z_i = \bar{z}, T_i^{\bar{z}} = c] \Pr(T_i = c | Z_i = \bar{z}) \\
+ \mathbb{E}[Y_i^0 | Z_i = \bar{z}, T_i^{\bar{z}} = d] \Pr(T_i = d | Z_i = \bar{z}) \\
+ \mathbb{E}[Y_i^0 | Z_i = \bar{z}, T_i^{\bar{z}} = n] \Pr(T_i = n | Z_i = \bar{z}) \\
\lim_{z \to \bar{z}^-} \mathbb{E}[Y_i | Z_i = z] = \mathbb{E}[Y_i^1 | Z_i = \bar{z}, T_i^{\bar{z}} = a] \Pr(T_i = a | Z_i = \bar{z}) \\
+ \mathbb{E}[Y_i^0 | Z_i = \bar{z}, T_i^{\bar{z}} = c] \Pr(T_i = c | Z_i = \bar{z}) \\
+ \mathbb{E}[Y_i^1 | Z_i = \bar{z}, T_i^{\bar{z}} = d] \Pr(T_i = d | Z_i = \bar{z}) \\
+ \mathbb{E}[Y_i^0 | Z_i = \bar{z}, T_i^{\bar{z}} = n] \Pr(T_i = n | Z_i = \bar{z}) \\
N_{\bar{z}} = \mathbb{E}[Y_i^1 - Y_i^0 | Z_i = \bar{z}, T_i^{\bar{z}} = c] \Pr(T_i = c | Z_i = \bar{z}) \\
- \mathbb{E}[Y_i^1 - Y_i^0 | Z_i = \bar{z}, T_i^{\bar{z}} = d] \Pr(T_i = d | Z_i = \bar{z})
\end{aligned}
$$

Proof (cont.)

$$
\begin{aligned}
D_{\bar{z}} &= \lim_{z \to \bar{z}^+} \Pr(D_i = 1 | Z_i = z) - \lim_{z \to \bar{z}^-} \Pr(D_i = 1 | Z_i = z) \\
&= \lim_{z \to \bar{z}^+} \left(\Pr(T_i^z = a | Z_i = z) + \Pr(T_i^z = c | Z_i = z) \right) \\
&\quad - \lim_{z \to \bar{z}^-} \left(\Pr(T_i^z = a | Z_i = z) + \Pr(T_i^z = d | Z_i = z) \right) \\
&= \Pr(T_i^{\bar{z}} = c | Z_i = \bar{z}) - \Pr(T_i^{\bar{z}} = d | Z_i = \bar{z})
\end{aligned}
$$

Under Monotonicity, we have: $\Pr(T_i^{\bar{z}} = d | Z_i = \bar{z}) = 0$, which proves the result.

Estimation Using Simplified LLR

Imbens and Lemieux propose the following simplified version of the WALD LLR estimator: $\hat{\delta}$ estimated by 2SLS using $\mathbb{1}[Z_i \leq \bar{z}]$ as an instrument on the sample of observations such as $\bar{z} - h \leq Z_i \leq \bar{z} + h$ is an estimate of $LATE(z)$:

$$Y_i = \alpha + \beta(Z_i - \bar{z})\mathbb{1}[Z_i \leq \bar{z}] + \gamma(Z_i - \bar{z})\mathbb{1}[Z_i > \bar{z}] + \delta D_i + \epsilon_i$$

It is actually equal to the Wald LLR estimate with uniform kernel and identical bandwidth on each side of the threshold.

Bandwidth Choice

- Same techniques as for Sharp RDD
- 4 bandwidths to be selected
- For simplicity, you can use the minimum of the four bandwidths.

Simplified Wald Estimator: Illustration

Bandwidth=0.3

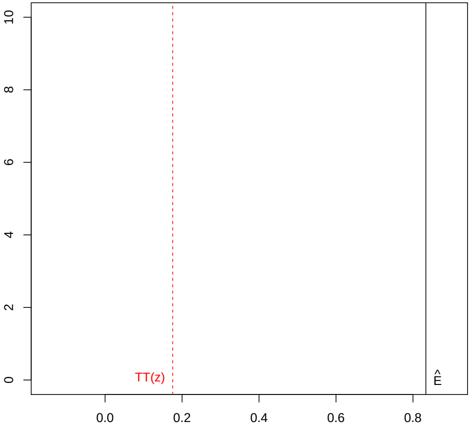

Figure: Simplified Wald RDD LLR

Simplified Wald Estimator: Illustration

Bandwidth=0.4

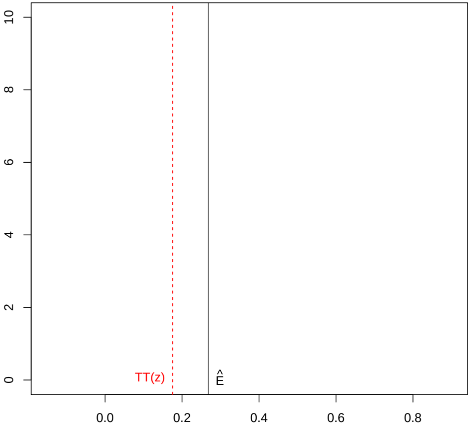

Figure: Simplified Wald RDD LLR

Sampling Noise with RDD LLR IV: Illustration

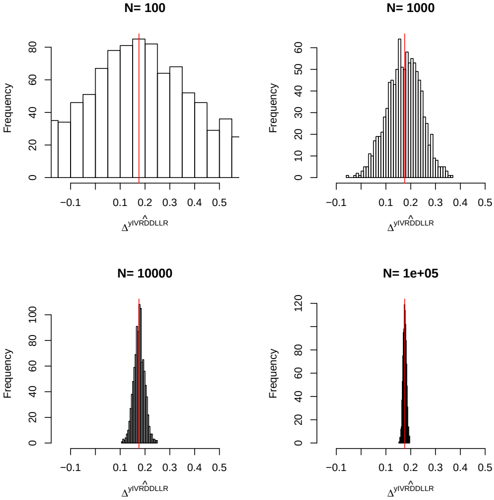

Figure: Distribution of the *IVRDDLLR* estimator over replications of samples of different sizes

Estimating Precision in the Sharp RDD Design

Several approaches:

1. Hahn, Todd and van der Klaauw (2001) derive general CLT results
2. Imbens and Lemieux (2008) simplify the CLT results and propose a plug-in estimator
3. Imbens and Lemieux (2008) propose to use the robust variance of the 2SLS estimator
4. Bootstrap should be valid

Asymptotic Variance of the Simplified LLR-IV Estimator

Theorem (Asymptotic Variance of the LLR-IV Estimator)

The variance of the simplified LLR-IV Estimator in a Fuzzy Design can be approximated by:

$$\mathbb{V}[\hat{\Delta}_{LLRRDDIV}] \approx \frac{1}{\sqrt{Nh}} \left(\frac{1}{\tau_D^2 V_{\tau_Y}} + \frac{\tau_Y^2}{\tau_D^4} V_{\tau_D} - 2\frac{\tau_Y}{\tau_D^3} C_{\tau_Y,\tau_D} \right),$$

with

$$\tau_D = \lim_{e\to 0^+} \mathbb{E}[D_i|Z_i = \bar{z} + e] - \lim_{e\to 0^+} \mathbb{E}[D_i|Z_i = \bar{z} - e]$$

$$V_{\tau_Y} = \frac{4}{f_Z(\bar{z})} \left(\sigma^2_{Y^r} + \sigma^2_{Y^l}\right) \qquad V_{\tau_D} = \frac{4}{f_Z(\bar{z})} \left(\sigma^2_{D^r} + \sigma^2_{D^l}\right)$$

$$C_{\tau_Y,\tau_D} = \frac{4}{f_Z(\bar{z})} \left(C_{YD^r} + C_{YD^l}\right) \qquad \sigma^2_{Y^r} = \lim_{e\to 0^+} \mathbb{V}[Y_i|Z_i = \bar{z} + e]$$

$$C_{YD^r} = \lim_{e\to 0^+} \mathbb{C}[Y_i, D_i|Z_i = \bar{z} + e]$$

Proof

Hahn, Todd and van der Klaauw (2001) and Imbens and Lemieux (2008).

Illustration of Imbens and Lemieux Simplified LLR-IV

- The estimated value of LATE(z) by simplified LLR IV is 0.2678
- The estimated 99% sampling noise is: 2.7953.
- The true 99% sampling noise of LLR IV estimated by Monte Carlo simulations is: 0.3374.

Outline

Regression Discontinuity Designs (RDD)

Instrumental Variables (IV)

Difference in Differences (DID)

DID-IV

IV: Basic Intuition

There is a variable that influences who receives the treatment and that is not correlated with potential outcomes. Any correlation between this variable and the outcomes is interpreted as an effect of the treatment.

IV: Examples

- Distance to school
- Random draft lottery number
- Randomized encouragement to participate

IV: Illustration

$$
\begin{aligned}
D_i &= \mathbb{1}[y_i^B + \kappa_i Z_i + V_i \leq \bar{y}] \\
Z_i &\sim \mathcal{B}(p_Z) \\
Z_i &\perp\!\!\!\perp (Y_i^0, Y_i^1, V_i) \\
\kappa_i &= \begin{cases} \bar{\kappa} & \text{if } \xi = 1 \\ \underline{\kappa} & \text{if } \xi = 0 \end{cases} \\
\xi &\sim \mathcal{B}(p_\xi) \\
\xi &\perp\!\!\!\perp (Y_i^0, Y_i^1, V_i, Z_i)
\end{aligned}
$$

IV Assumptions: First Stage Rank Condition

Assumption (First Stage Full Rank)

We assume that

$$\Pr(D_i = 1|Z_i = 1) \neq \Pr(D_i = 1|Z_i = 0).$$

The instrument Z_i has a direct effect on treatment participation

IV: Illustration

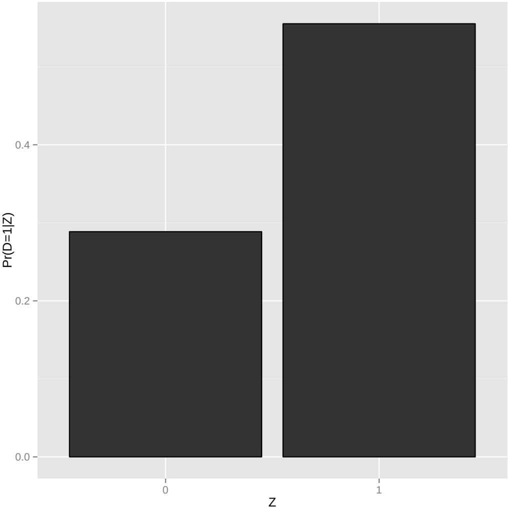

Figure: Illustration of the IV assumptions: First Stage

IV Assumptions: Exclusion Restriction

Assumption (Exclusion Restriction)

We assume that

$$\forall d, z \in \{0, 1\}, \ Y_i^{d,z} = Y_i^d.$$

There is no direct effect of Z_i on outcomes.

IV Assumptions: Independence

Assumption (Independence)

We assume that

$$(Y_i^1, Y_i^0, D_i^1, D_i^0) \perp\!\!\!\perp Z_i.$$

Z_i is not correlated with the other determinants of y_i and D_i.

IV: Illustration

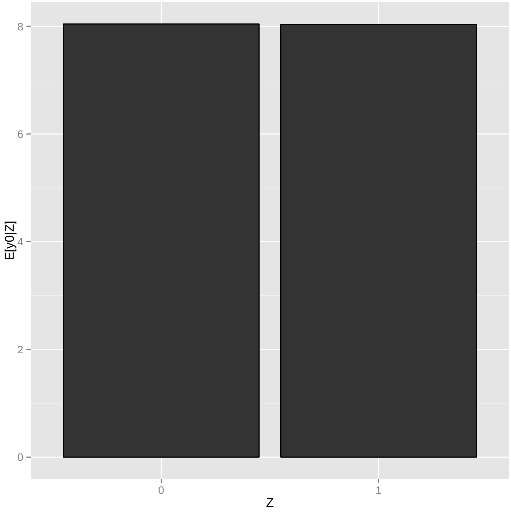

Figure: Illustration of the IV assumptions: Independence

IV: Illustration

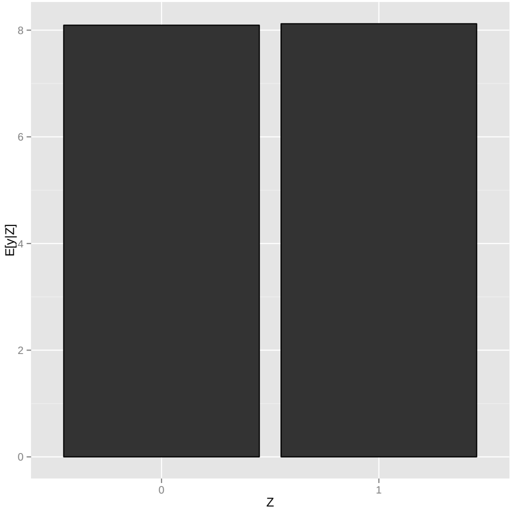

Figure: Illustration of the IV assumptions: Reduced Form

IV Assumptions: Independent Treatment Effect

Assumption (Independent Treatment Effect)

We assume that

$$\Delta_i^Y \perp\!\!\!\perp Z_i | D_i.$$

The treatment effect is not correlated with participation in the treatment.

Identification in the IV framework under Independent Treatment Effect

Theorem (Identification under Independent Treatment Effect)

Under First Stage Full Rank, Exclusion Restriction, Independence and Independent Treatment Effect, the Wald estimator identifies TT:

$$\Delta^Y_{Wald} = \Delta^Y_{TT},$$

where:

$$\Delta^Y_{Wald} = \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\Pr(D_i = 1|Z_i = 1) - \Pr(D_i = 1|Z_i = 0)}.$$

Proof

$$
\begin{aligned}
\mathbb{E}[Y_i|Z_i = 1] &= \mathbb{E}[Y_i^0 + (Y_i^1 - Y_i^0)D_i|Z_i = 1] \\
&= \mathbb{E}[Y_i^0|Z_i = 1] + \mathbb{E}[\Delta_i^Y|D_i = 1, Z_i = 1]\Pr(D_i = 1|Z_i = 1) \\
&= \mathbb{E}[Y_i^0] + \mathbb{E}[\Delta_i^Y|D_i = 1]\Pr(D_i = 1|Z_i = 1) \\
\mathbb{E}[Y_i|Z_i = 0] &= \mathbb{E}[Y_i^0] + \mathbb{E}[\Delta_i^Y|D_i = 1]\Pr(D_i = 1|Z_i = 0).
\end{aligned}
$$

where the first equality uses exclusion restriction, the third equality uses Independence and Independent Treatment Effect. The results follows from the First Stage Full Rank, which implies that the denominator of the Wald estimator is different from zero.

Types

If we do not make the Independent Treatment Effect Assumption, we need to distinguish four types of individuals:

- Always takers ($T_i = a$): $D_i^1 = D_i^0 = 1$
- Never takers ($T_i = n$): $D_i^1 = D_i^0 = 0$
- Compliers ($T_i = c$): $D_i^1 = 1$ and $D_i^0 = 0$
- Defiers ($T_i = d$): $D_i^1 = 0$ and $D_i^0 = 1$

Types: Illustration

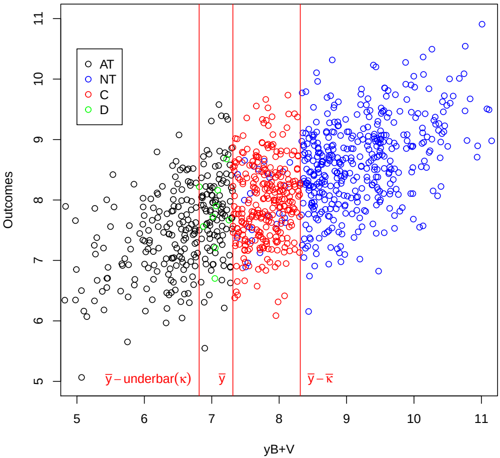

Figure: Types

IV Assumptions: Monotonicity

Assumption (Monotonicity)

We assume that

$$\forall i, \text{ either } D_i^1 \geq D_i^0 \text{ or } D_i^1 \leq D_i^0.$$

The instrument moves everyone in the population in the same direction. Without loss of generality, let's assume that $\forall i$, $D_i^1 \geq D_i^0$. As a consequence, there are no defiers.

IV with Monotonity and Correlated Effects: Illustration

$$
\begin{aligned}
D_i &= \mathbb{1}[y_i^B + \bar{\kappa} Z_i + V_i \leq \bar{y}] \\
V_i &= \zeta(\mu_i - \bar{\mu}) + \lambda_i \\
\lambda_i &\sim \mathcal{N}(0, (1-\zeta^2)\sigma_\mu^2 + \sigma_U^2)
\end{aligned}
$$

Types with Monotonicity: Illustration

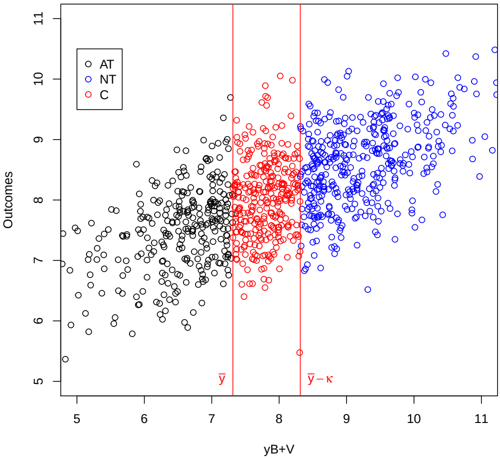

Figure: Types

Types with Monotonicity: Illustration

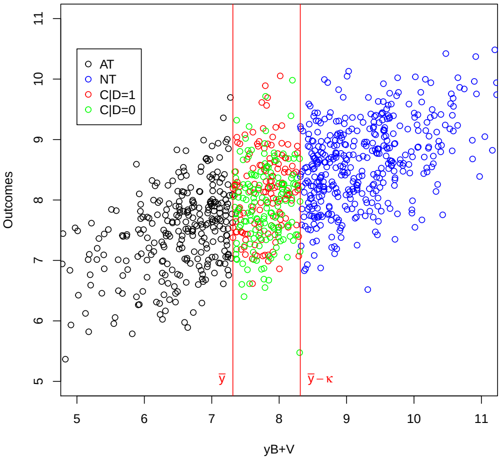

Figure: Types

Identification in the IV framework under Monotonicity

Theorem (Identification under Monotonicity)

Under First Stage Full Rank, Exclusion Restriction, Independence and Monotonicity, the Wald estimator identifies the LATE, i.e. the effect of the treatment on the compliers:

$$\Delta^Y_{Wald} = \Delta^Y_{LATE},$$

where:

$$\Delta^Y_{LATE} = \mathbb{E}[\Delta^Y_i | T_i = c].$$

Proof

$$
\begin{aligned}
\mathbb{E}[Y_i|Z_i=1] &= \mathbb{E}[Y_i^1|T_i=a]\Pr(T_i=a)+\mathbb{E}[Y_i^1|T_i=c]\Pr(T_i=c)\\
&\quad+\mathbb{E}[Y_i^0|T_i=d]\Pr(T_i=d)+\mathbb{E}[Y_i^0|T_i=n]\Pr(T_i=n)\\
\mathbb{E}[Y_i|Z_i=0] &= \mathbb{E}[Y_i^1|T_i=a]\Pr(T_i=a)+\mathbb{E}[Y_i^0|T_i=c]\Pr(T_i=c)\\
&\quad+\mathbb{E}[Y_i^1|T_i=d]\Pr(T_i=d)+\mathbb{E}[Y_i^0|T_i=n]\Pr(T_i=n),
\end{aligned}
$$

where we have used Exclusion Restriction and Independence. Now:

$$
\begin{aligned}
\mathbb{E}[Y_i|Z_i=1]-\mathbb{E}[Y_i|Z_i=0] &= \mathbb{E}[\Delta_i^Y|T_i=c]\Pr(T_i=c)-\mathbb{E}[\Delta_i^Y|T_i=d]\Pr(T_i=d)\\
&= \mathbb{E}[\Delta_i^Y|T_i=c]\Pr(T_i=c),
\end{aligned}
$$

where the second equality uses Monotonicity. We have:

$$
\begin{aligned}
\Pr(D_i=1|Z_i=1) &= \Pr(T_i=a)+\Pr(T_i=c)\\
\Pr(D_i=1|Z_i=0) &= \Pr(T_i=a)+\Pr(T_i=d),
\end{aligned}
$$

where we have used Independence. As a consequence, under Monotonicity, we have:

$$
\Pr(D_i=1|Z_i=1)-\Pr(D_i=1|Z_i=0)=\Pr(T_i=c).
$$

Using First Stage Full Rank proves the result.

Estimation Using the Wald Estimator

We can directly use the empirical equivalent to the Wald Estimator:

$$\hat{\Delta}^Y_{Wald} = \frac{\frac{1}{\sum_{i=1}^N Z_i}\sum_{i=1}^N Z_i Y_i - \frac{1}{\sum_{i=1}^N (1-Z_i)}\sum_{i=1}^N (1-Z_i) Y_i}{\frac{1}{\sum_{i=1}^N Z_i}\sum_{i=1}^N Z_i D_i - \frac{1}{\sum_{i=1}^N (1-Z_i)}\sum_{i=1}^N (1-Z_i) D_i}.$$

In our example, we have $\hat{\Delta}^Y_{Wald} =$0.066.

The Value of LATE in our Illustration

$$\Delta^y_{LATE} = \bar{\alpha} + \theta\bar{\mu} + \theta \frac{(1+\zeta)\sigma^2_\mu}{\sqrt{2(\sigma^2_\mu + \sigma^2_U)}} \frac{\phi\left(\frac{\bar{y}-\bar{\mu}}{\sqrt{2(\sigma^2_\mu+\sigma^2_U)}}\right) - \phi\left(\frac{\bar{y}-\bar{\kappa}-\bar{\mu}}{\sqrt{2(\sigma^2_\mu+\sigma^2_U)}}\right)}{\Phi\left(\frac{\bar{y}-\bar{\kappa}-\bar{\mu}}{\sqrt{2(\sigma^2_\mu+\sigma^2_U)}}\right) - \Phi\left(\frac{\bar{y}-\bar{\mu}}{\sqrt{2(\sigma^2_\mu+\sigma^2_U)}}\right)}.$$

In our example, we have $\Delta^y_{LATE} =$0.1794.

Proof

$$
\begin{aligned}
\Delta^{y}_{LATE} &= \mathbb{E}[\Delta^{Y}_{i} | T_i = c] \\
&= \mathbb{E}[\bar{\alpha} + \theta\mu_i + \eta_i | \bar{y} \leq y_i^B + V_i < \bar{y} - \bar{\kappa}] \\
&= \bar{\alpha} + \theta\mathbb{E}[\mu_i | \bar{y} \leq y_i^B + V_i < \bar{y} - \bar{\kappa}] \\
&= \bar{\alpha} + \theta\bar{\mu} + \theta\left(\frac{\mathbb{C}[\mu_i, y_i^B + V_i]}{\sqrt{\mathbb{V}[y_i^B + V_i]}}\frac{\phi\left(\frac{\bar{y}-\bar{\mu}}{\sqrt{\mathbb{V}[y_i^B+V_i]}}\right) - \phi\left(\frac{\bar{y}-\bar{\kappa}-\bar{\mu}}{\sqrt{\mathbb{V}[y_i^B+V_i]}}\right)}{\Phi\left(\frac{\bar{y}-\bar{\kappa}-\bar{\mu}}{\sqrt{\mathbb{V}[y_i^B+V_i]}}\right) - \Phi\left(\frac{\bar{y}-\bar{\mu}}{\sqrt{\mathbb{V}[y_i^B+V_i]}}\right)}\right),
\end{aligned}
$$

where the fourth equality uses the formula for the expectation of a doubly censored bivariate normal distribution.

We also have that $\mathbb{V}[y_i^B + V_i] = 2(\sigma_\mu^2 + \sigma_U^2)$ and $\mathbb{C}[\mu_i, y_i^B + V_i] = (1 + \zeta)\sigma_\mu^2$, which proves the result.

Wald Estimator and the LATE

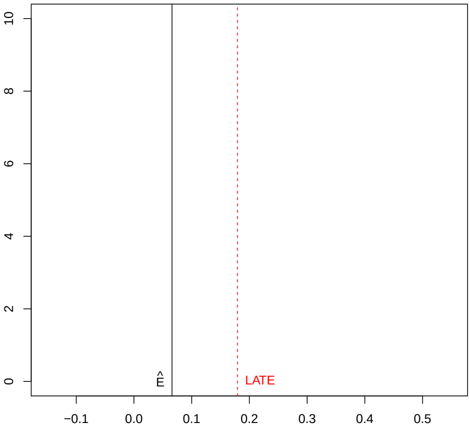

Figure: Wald Estimator and the LATE

Sampling Noise with Wald: Illustration

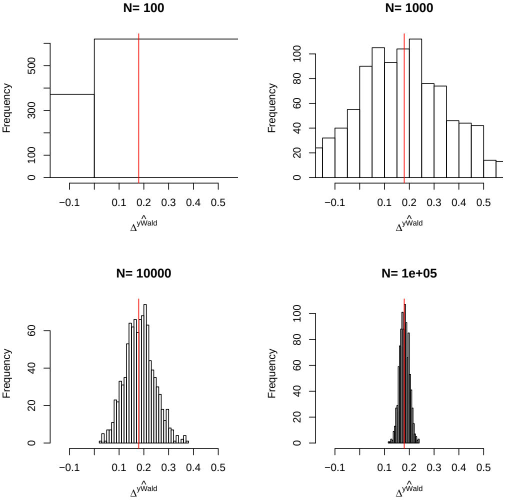

Figure: Distribution of the Wald estimator over replications of samples of different sizes

Sampling Noise with Wald: Illustration

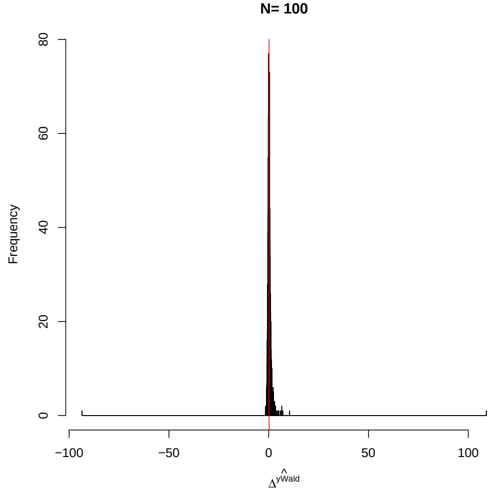

Figure: Distribution of the Wald estimator over replications

Wald is 2SLS

Lemma (Wald is 2SLS)

Under the First Stage Full Rank Assumption, the 2SLS coefficient β in the following regression:

$$Y_i = \alpha + \beta D_i + U_i$$

estimated using Z_i as an instrument for D_i is the Wald estimator:

$$\begin{aligned}
\hat{\beta}_{2SLS} &= \frac{\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \frac{1}{N}\sum_{i=1}^{N} Y_i\right)\left(Z_i - \frac{1}{N}\sum_{i=1}^{N} Z_i\right)}{\frac{1}{N}\sum_{i=1}^{N}\left(D_i - \frac{1}{N}\sum_{i=1}^{N} D_i\right)\left(Z_i - \frac{1}{N}\sum_{i=1}^{N} Z_i\right)} \\
&= \hat{\Delta}^{Y}_{Wald}.
\end{aligned}$$

Proof

In matrix notation, we have:

$$\underbrace{\begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix}}_{Y} = \begin{pmatrix} 1 & D_1 \\ \vdots & \vdots \\ 1 & D_N \end{pmatrix} \underbrace{\begin{pmatrix} \alpha \\ \beta \end{pmatrix}}_{\Theta} + \underbrace{\begin{pmatrix} U_1 \\ \vdots \\ U_N \end{pmatrix}}_{U}$$

$$\underbrace{\begin{pmatrix} D_1 \\ \vdots \\ D_N \end{pmatrix}}_{D} = \underbrace{\begin{pmatrix} 1 & Z_1 \\ \vdots & \vdots \\ 1 & Z_N \end{pmatrix}}_{Z} \underbrace{\begin{pmatrix} \gamma \\ \delta \end{pmatrix}}_{\Xi} + \underbrace{\begin{pmatrix} \omega_1 \\ \vdots \\ \omega_N \end{pmatrix}}_{\Omega}$$

The 2SLS estimator is:

$$\hat{\Theta}_{2SLS} = (Z'D)^{-1}Z'Y$$

To be completed.

2SLS Estimator and the LATE

In our illustration, the 2SLS estimator is 0.066.

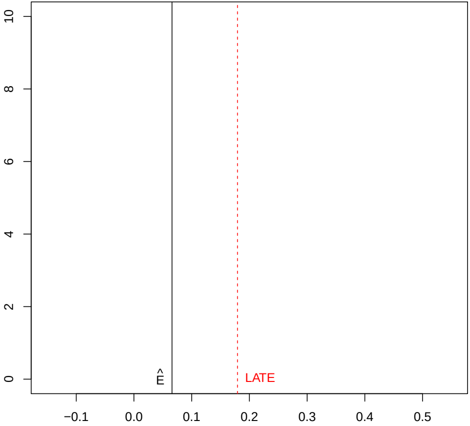

Figure: 2SLS Estimator and the LATE

More Complex Instruments

1. Ordered multivalued instrument: the 2SLS estimator is a weighted average of LATEs, with nonnegative weights summing to one (Imbens and Angrist, 1994).
2. Continuous instrument: the 2SLS estimator is a weighted average of LATEs with nonnegative weights summing to one (Heckman and Vytlacil (2005)).
3. Several instruments (Heckman, Urzua and Vytlacil, 2006):
   - Each instrument estimates a different LATE
   - Overidentification tests might fail whereas both instruments are OK
   - If the two instruments are correlated, you have to use them jointly

Conditioning on Additional Covariates

You might want to combine your Instrumental Variable with conditioning on some covariates. There are at least two possible reasons for that:

1. You feel that your instrument is only valid conditional on other covariates
2. You want to soak up variation in the outcomes to decrease sampling noise

Conditioning on Additional Covariates: Parametric Case

The first approach is parametric: estimate

$$Y_i = \alpha + \beta D_i + \gamma' X_i + U_i,$$

using 2SLS with Z_i as an instrument for D_i.

- You might want to center the X's at the average value of the covariates for the compliers to recover the LATE (to be shown)
- Maybe the regression line $\mathbb{E}[Y_i^0|X_i]$ is not linear: specification bias

2SLS Conditioning Parametrically on Covariates

In our illustration, 2SLS conditioning parametrically on y_i^B yields an estimate for the LATE of 0.1101.

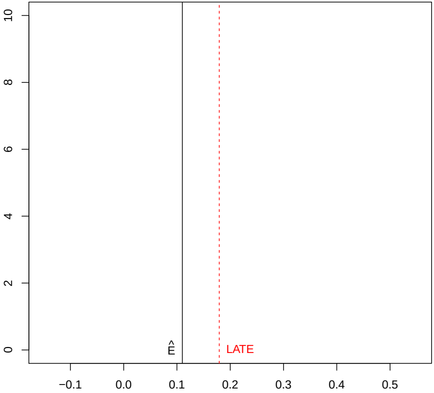

Figure: 2SLS(X) and LATE

Sampling Noise with 2SLS(X): Illustration

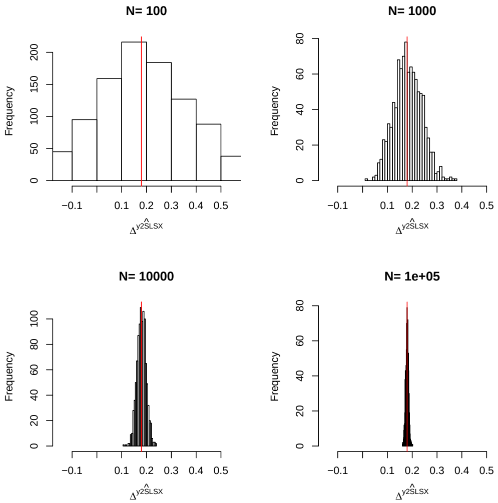

Figure: Distribution of the 2SLS(X) estimator over replications of samples of different sizes

Conditioning on Additional Covariates: Nonlinear Case

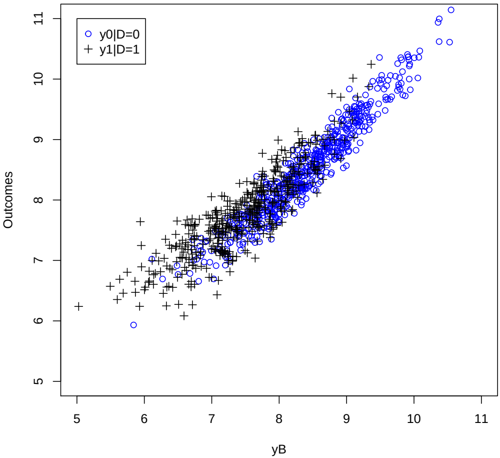

Figure: Nonlinear Regression Curve

Specification Bias

In our illustration, 2SLS conditioning parametrically on y_i^B yields an estimate for the LATE of 0.1279.

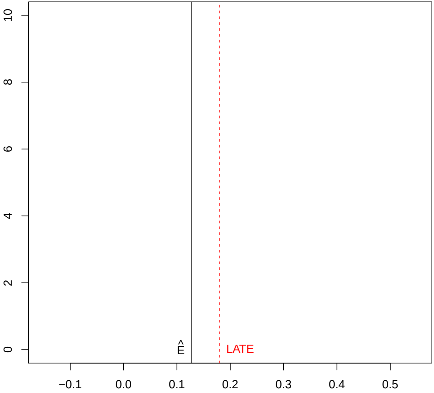

Figure: 2SLS(X) and LATE

Specification Bias with 2SLS(X): Illustration

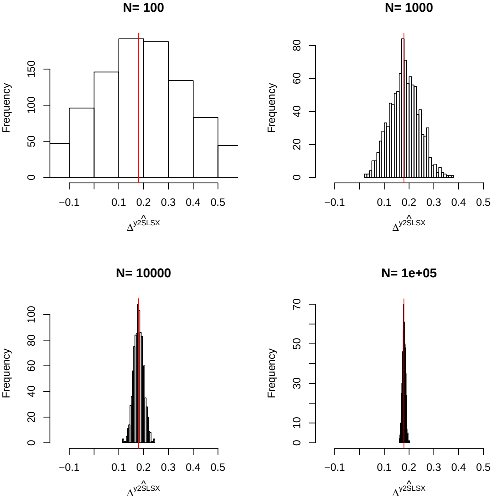

Figure: Distribution of the 2SLS(X) estimator over replications of samples of different sizes

Nonparametric Wald Estimator

Frolich (2008) proposes the following nonparametric Wald estimator:

$$\hat{\Delta}^{Y}_{NPWald} = \frac{\sum_{i:Z_i=1}(Y_i - \hat{m}_0(X_i)) + \sum_{i:Z_i=0}(\hat{m}_1(X_i) - Y_i)}{\sum_{i:Z_i=1}(D_i - \hat{\mu}_0(X_i)) + \sum_{i:Z_i=0}(\hat{\mu}_1(X_i) - D_i)},$$

where $\hat{m}_z$ and $\hat{\mu}_z$ are nonparametric regression estimators of $\mathbb{E}[Y_i|X_i, Z_i = z]$ and $\mathbb{E}[D_i|X_i, Z_i = z]$ respectively.

Nonparametric Wald Estimator

- This is simply the ratio of two average treatment effects: Z on Y and Z on D.
- In practice, you can use the LLR Matching estimator on the propensity score (with trimming) to recover the numerator and the denominator.

Nonparametric Wald

In our illustration, NPWald conditioning nonparametrically on y_i^B yields an estimate for the LATE of 0.1079.

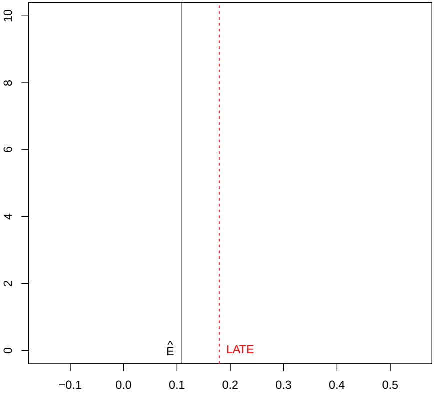

Figure: NPWald and LATE

ExtrapoLATEing

It is frustrating to recover only the LATE. One would like to infer something about TT using the LATE.

- With a continuous instrument and strong support conditions, Heckman and Vytlacil (2005) show how to reweight the MTE to recover the TT.
- Angrist and Fernandez-Val (2013) assume that all treatment effect heterogeneity is due to observed covariates and propose a reweighting estimator.

Inference with IV estimators

- Linear 2SLS estimator: use the heteroskedasticity-robust 2SLS standard errors derived from the CLT
- Nonparametric Wald Estimator: Frolich derives the efficiency bound and shows that the estimator reaches it. Bootstrap should work.

Inference with IV estimators: Illustration

Without controls:

- True 99% sampling noise (from the simulations) is 1.0712
- 99% Sampling noise estimated using the heteroskedasticity-robust 2SLS standard errors is 0.8209

With controls:

- True 99% sampling noise (from the simulations) is 0.2708
- 99% sampling noise estimated using the heteroskedasticity-robust 2SLS standard errors is 0.2602

Outline

DID: Basic Intuition

The difference between treated and untreated before the treatment approximates selection bias. Correcting the With/Without comparison after treatment by the With/Without comparison before treatment recovers TT. Hence the name Differences in Differences (DID).

DID: Illustration

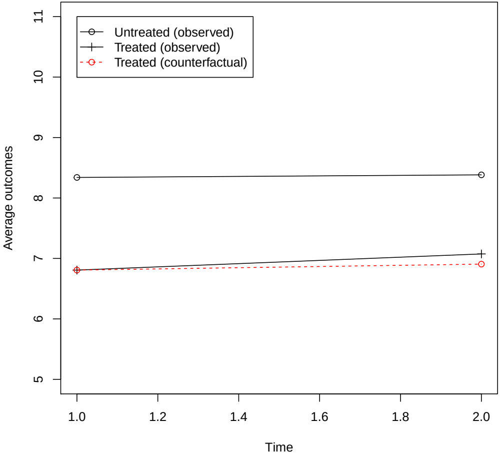

Figure: Evolution of average outcomes in the treated and control group

DID: Illustration

- The With/Without comparison After the treatment is: Δ^y_{WWA} =-1.3082
- The With/Without comparison Before the treatment is: Δ^y_{WWB} =-1.5326
- The DID estimator is: $\Delta^y_{DID} = \Delta^y_{WWA} - \Delta^y_{WWB}$ =-1.3082+1.5326=0.2244

DID: illustration

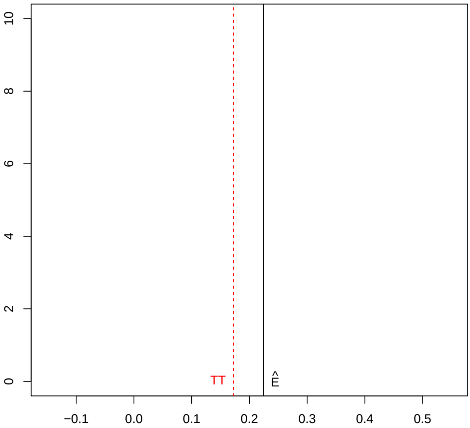

Figure: DID

Key Assumption: Parallel Trends

We have two periods, Before and After, denoted $\{A, B\}$.

Assumption (Parallel Trends)

We assume that:

$$\mathbb{E}[Y^0_{i,A}|D_i = 1] - \mathbb{E}[Y^0_{i,B}|D_i = 1] = \mathbb{E}[Y^0_{i,A}|D_i = 0] - \mathbb{E}[Y^0_{i,B}|D_i = 0].$$

Parallel Trends Is Constant Selection Bias

$$\mathbb{E}[Y^0_{i,A}|D_i = 1] - \mathbb{E}[Y^0_{i,B}|D_i = 1] = \mathbb{E}[Y^0_{i,A}|D_i = 0] - \mathbb{E}[Y^0_{i,B}|D_i = 0]$$
$$\Leftrightarrow \mathbb{E}[Y^0_{i,A}|D_i = 1] - \mathbb{E}[Y^0_{i,A}|D_i = 0] = \mathbb{E}[Y^0_{i,B}|D_i = 1] - \mathbb{E}[Y^0_{i,B}|D_i = 0].$$

Parallel Trends: Illustration

In our sample:

- $\hat{\mathbb{E}}[y_i^0|D_i = 1] - \hat{\mathbb{E}}[y_i^B|D_i = 1] = 0.099$ and $\hat{\mathbb{E}}[y_i^0|D_i = 0] - \hat{\mathbb{E}}[y_i^B|D_i = 0] = 0.0429$
- $\hat{\mathbb{E}}[y_i^0|D_i = 1] - \hat{\mathbb{E}}[y_i^0|D_i = 0] =$ -1.4765 and $\hat{\mathbb{E}}[y_i^B|D_i = 1] - \hat{\mathbb{E}}[y_i^B|D_i = 0] =$ -1.5326

Parallel Trends Does Not Hold in Our Illustration

$$
\begin{aligned}
&\mathbb{E}[y_i^0|D_i = 1] - \mathbb{E}[y_i^B|D_i = 1] \\
&\qquad = \mathbb{E}[\mu_i + \delta + U_i^0|D_i = 1] - \mathbb{E}[\mu_i + U_i^B|D_i = 1] \\
&\qquad = \mathbb{E}[\mu_i|D_i = 1] + \delta + \mathbb{E}[U_i^0|D_i = 1] \\
&\qquad \quad - \mathbb{E}[\mu_i|D_i = 1] - \mathbb{E}[U_i^B|D_i = 1] \\
&\qquad = \delta + \mathbb{E}[\rho U_i^B + \epsilon_i|D_i = 1] - \mathbb{E}[U_i^B|D_i = 1] \\
&\qquad = \delta - (1-\rho)\mathbb{E}[U_i^B|D_i = 1]
\end{aligned}
$$

$$
\mathbb{E}[y_i^0|D_i = 0] - \mathbb{E}[y_i^B|D_i = 0] = \delta - (1-\rho)\mathbb{E}[U_i^B|D_i = 0]
$$

$$
\begin{aligned}
&\mathbb{E}[y_i^0|D_i = 1] - \mathbb{E}[y_i^B|D_i = 1] - (\mathbb{E}[y_i^0|D_i = 0] - \mathbb{E}[y_i^B|D_i = 0]) \\
&\qquad = -(1-\rho)(\mathbb{E}[U_i^B|D_i = 1] - \mathbb{E}[U_i^B|D_i = 0]) \\
&\qquad = -(1-\rho)(\mathbb{E}[U_i^B|\mu_i + U_i^B \le \bar{y}] - \mathbb{E}[U_i^B|\mu_i + U_i^B > \bar{y}])
\end{aligned}
$$

Parallel Trends in Our Illustration

Simply set $\rho = 1$.

Parallel Trends: Illustration

In our new sample:

- $\hat{\mathbb{E}}[y_i^0|D_i = 1] - \hat{\mathbb{E}}[y_i^B|D_i = 1] = 0.0552$ and $\hat{\mathbb{E}}[y_i^0|D_i = 0] - \hat{\mathbb{E}}[y_i^B|D_i = 0] = 0.0568$
- $\hat{\mathbb{E}}[y_i^0|D_i = 1] - \hat{\mathbb{E}}[y_i^0|D_i = 0] =$ -1.5342 and $\hat{\mathbb{E}}[y_i^B|D_i = 1] - \hat{\mathbb{E}}[y_i^B|D_i = 0] =$ -1.5326

DID: Illustration with Parallel Trends

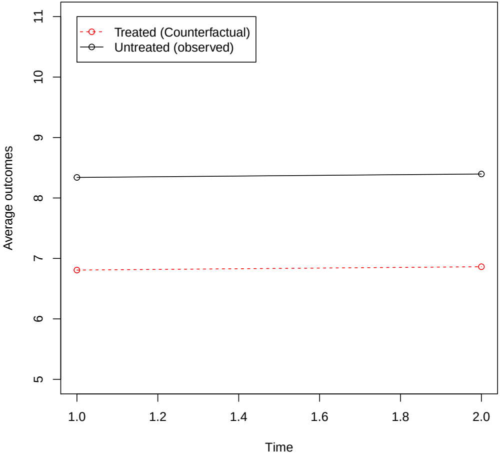

Figure: Evolution of average potential outcomes in the treated and control group

Identification of TT Under Parallel Trends

Theorem (Identification of TT with DID)

Under the Parallel Trends Assumption, TT is identified by DID:

$$\Delta^Y_{DID} = \Delta^{Y_A}_{TT},$$

with:

$$\begin{aligned}\Delta^Y_{DID} = {} & \mathbb{E}[Y_{i,A}|D_i = 1] - \mathbb{E}[Y_{i,A}|D_i = 0] \\ & - (\mathbb{E}[Y_{i,B}|D_i = 1] - \mathbb{E}[Y_{i,B}|D_i = 0]).\end{aligned}$$

Proof

$$\begin{aligned}\Delta^{Y}_{DID} &= \mathbb{E}[Y_{i,A}|D_i = 1] - \mathbb{E}[Y_{i,A}|D_i = 0] - (\mathbb{E}[Y_{i,B}|D_i = 1] - \mathbb{E}[Y_{i,B}|D_i = 0]) \\ &= \mathbb{E}[Y^1_{i,A}|D_i = 1] - \mathbb{E}[Y^0_{i,A}|D_i = 0] - (\mathbb{E}[Y^0_{i,B}|D_i = 1] - \mathbb{E}[Y^0_{i,B}|D_i = 0]).\end{aligned}$$

Under Parallel Trends, we have:

$$\mathbb{E}[Y^0_{i,A}|D_i = 1] = \mathbb{E}[Y^0_{i,A}|D_i = 0] + (\mathbb{E}[Y^0_{i,B}|D_i = 1] - \mathbb{E}[Y^0_{i,B}|D_i = 0])$$

As a consequence, we have:

$$\begin{aligned}\Delta^{Y}_{DID} &= \mathbb{E}[Y^1_{i,A}|D_i = 1] - \mathbb{E}[Y^0_{i,A}|D_i = 1] \\ &= \mathbb{E}[Y^1_{i,A} - Y^0_{i,A}|D_i = 1] \\ &= \Delta^{Y_A}_{TT}.\end{aligned}$$

DID Estimators

There are (at least) four different DID estimators:

1. Direct
2. Pooled OLS
3. Fixed Effects
4. First Difference

- With a panel of two periods, they are all numerically identical
- The last two are infeasible with repeated cross sections

The Direct DID Estimator

$$\hat{\Delta}^{Y}_{DID} = \frac{1}{\sum_{i=1}^{N} D_i} \sum_{i=1}^{N} Y_{i,A} D_i - \frac{1}{\sum_{i=1}^{N}(1 - D_i)} \sum_{i=1}^{N} Y_{i,A}(1 - D_i)$$
$$- \left(\frac{1}{\sum_{i=1}^{N} D_i} \sum_{i=1}^{N} Y_{i,B} D_i - \frac{1}{\sum_{i=1}^{N}(1 - D_i)} \sum_{i=1}^{N} Y_{i,B}(1 - D_i) \right).$$

The Direct DID Estimator: Illustration

- The With/Without comparison After the treatment is: Δ^y_{WWA} =-1.366
- The With/Without comparison Before the treatment is: Δ^y_{WWB} =-1.5326
- The DID estimator is: $\Delta^y_{DID} = \Delta^y_{WWA} - \Delta^y_{WWB}$ =-1.366+1.5326=0.1666

The Direct DID Estimator: Illustration

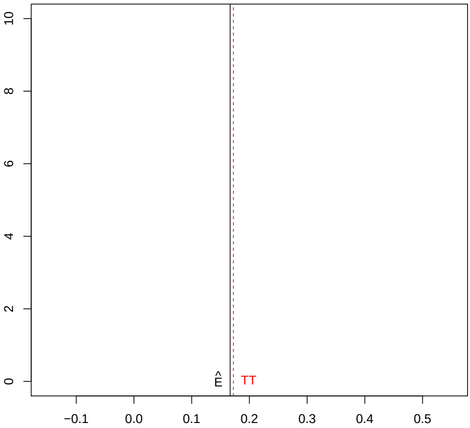

Figure: DID

The Pooled OLS DID Estimator

Let's pool all the observations from Before and After. Now, we have $2N$ observations. Let's t_i denote time: $t_i = 1$ when the outcome of observation i is observed After the treatment and $t_i = 0$ when the outcome of observation i is observed Before the treatment.

$$Y_i = \alpha + \delta t_i + \gamma D_i + \beta t_i D_i + U_i$$

$\hat{\beta}_{OLS}$ in the previous regression is the Pooled OLS DID estimator.

```
y.pool <- c(y, yB)
Ds.pool <- c(Ds, Ds)
t <- c(rep(1, N), rep(0, N))
t.D <- t * Ds.pool
```

The Pooled OLS DID Estimator: Illustration

```
reg.did.pooled.ols <- lm(y.pool   t + Ds.pool + t.D)
```

In our illustration, $\hat{\beta}_{OLS} =$0.1666.

The Fixed Effects DID Estimator

$$Y_{i,t} = \mu_i + \delta_t + \beta t_i D_i + U_i$$

$\hat{\beta}_{FE}$ in the previous regression is the Fixed Effects DID estimator.

```
data.panel <- cbind(c(seq(1, N), seq(1, N)), t, y.pool, Ds
colnames(data.panel) <- c("Individual", "time", "y", "Ds",
data.panel <- as.data.frame(data.panel)
```

The Fixed Effects DID Estimator in Our Illustration

You have to load the library plm.

```
reg.did.fe <- plm(y   time + t.D, data = data.panel, index
= c("Individual", "time"), model = "within")
```

In our illustration, $\hat{\beta}_{FE} =$0.1666.

The First Difference DID Estimator

$$Y_{i,A} - Y_{i,B} = \delta + \beta D_i + U_i$$

$\hat{\beta}_{FD}$ in the previous regression estimated by OLS is the First Difference DID estimator.

The First Difference DID Estimator in Our Illustration

You have to load the library plm.

```
reg.did.fd <- plm(y   time + t.D, data = data.panel, index
= c("Individual", "time"), model = "fd")
```

In our illustration, $\hat{\beta}_{FD} =$0.1666.

All Four DID Estimators are Equivalent

Theorem (Equivalence of DID EStimators)

With panel data and two periods of observation, we have:

$$\hat{\Delta}^{Y}_{DID} = \hat{\beta}_{OLS} = \hat{\beta}_{FE} = \hat{\beta}_{FD}.$$

Proof

To do

Sampling Noise with DID in Panels: Illustration

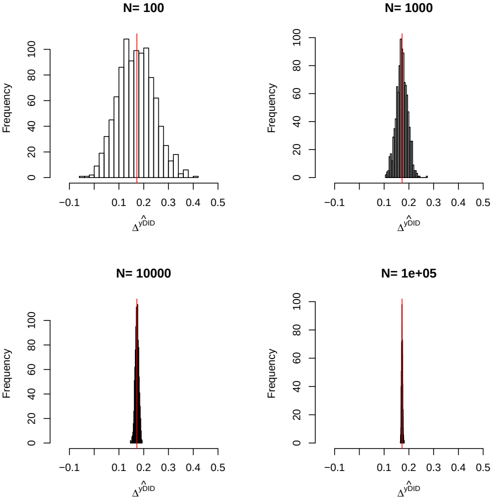

Figure: Distribution of the DID estimator over replications of panels of different sizes

Sampling Noise with DID in Cross Sections: Illustration

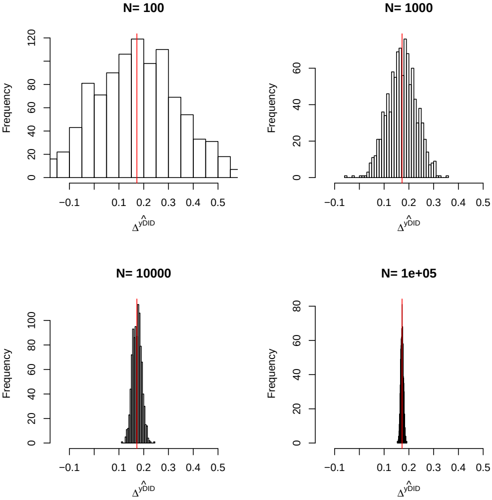

Figure: Distribution of the DID estimator over replications of repeated cross sections of different sizes

Inference with DID In Panel Data

- True 99% sampling noise (from the simulations) is 0.1153
- 99% sampling noise estimated using default FE standard errors is 0.0954
- 99% sampling noise estimated using heteroskedasticity robust FE standard errors is 0.1124

Inference with DID In Repeated Cross Sections

- True 99% sampling noise (from the simulations) is 0.2769
- 99% sampling noise estimated using default OLS standard errors is 0.17
- 99% sampling noise estimated using heteroskedasticity robust OLS standard errors is 0.1818

```
reg.did.pooling <- plm(y   time + Ds + t.D, data =
data.panel, index = c("Individual", "time"), model
= "pooling")
```

- 99% sampling noise estimated using corrected OLS standard errors is 0.17
- 99% sampling noise estimated using heteroskedasticity robust corrected OLS standard errors is 0.0417

Conditioning on Observed Covariates in DID

There are again two reasons why you might want to do that

- You suspect Parallel Trends to hold only conditionnally on some covariates
- You want to soak up the variance due to the covariates and thus increase precision

Conditioning on Observed Covariates in DID

There are four ways to condition on covariates in DID

- Adding covariates parametrically to the pooled OLS specification
- Adding covariates parametrically to the FE specification
- Adding covariates parametrically to the first difference specification
- Adding covariates nonparametrically to the first difference specification
- Adding covariates nonparametrically in repeated cross sections

Adding Covariates Parametrically to Pooled OLS

$$Y_i = \alpha + \delta t_i + \gamma D_i + \beta t_i D_i + \gamma' X_i + U_i.$$

Fine, but only to soak up variance.

Adding Covariates Parametrically to FE

$$Y_i = \mu_i + \delta_t + \beta t_i D_i + \gamma' X_i + U_i.$$

Useless: dropped since colinear with fixed effect.

Adding Covariates Parametrically to FD

$$Y_{i,A} - Y_{i,B} = \delta + \beta D_i + \gamma' X_i + U_i$$

- Useful for both purposes (soak up variance and allow for different trends).
- Careful: cannot use plm, and have to correct OLS standard errors (to do)

Adding Covariates Nonparametrically to FD

$$\Delta^Y_{DIDNPX} = \mathbb{E}[Y_{i,A} - Y_{i,B}|D_i = 1] - \mathbb{E}[\mathbb{E}[Y_{i,A} - Y_{i,B}|D_i = 0, X_i]|D_i = 1]$$

- Simply use Matching with $Y_{i,A} - Y_{i,B}$ (Heckman, Ichimura, Smith and Todd, 1998)
- Careful: you cannot put pre-treatment outcomes in the X vector (Chabé-Ferret, 2015)

Adding Covariates Nonparametrically in Repeated Cross Sections

Abadie (2005) derives a reweighting estimator for this case.

Outline

DID-IV: Basic Intuition

There are two groups of people ($Z_i = 1$ and $Z_i = 0$): in the first group, more people receive the treatment than in the second. Z_i has no direct effect on outcomes in the absence of the treatment and the parallel trends assumptions holds for this variable. If the trends on outcomes between the two groups delineated by Z_i are not parallel, then it must be because the treatment has an effect.

DID-IV: Example

- There are 50 states and the treatment is only available in 25 of them.
- Duflo (2001): a government program builds schools in some parts of Indonesia (where initial education levels are lower).
  - What is the impact of the program on education?
  - What is the impact of education on wages?

  In this paper, cohorts act as time periods.

DID-IV In Our Illustration

$$
\begin{aligned}
\mu_i &= \mu_i^S + \mu_i^U + \bar{\mu} \\
\mu_i^S &\sim \mathcal{N}(0, \frac{1}{3}\sigma_\mu^2) \\
\mu_i^U &\sim \mathcal{N}(0, \frac{2}{3}\sigma_\mu^2) \\
Z_i &= \begin{cases} 1 & \text{if } \mu_i^S \leq 0 \\ 0 & \text{if } \mu_i^S > 0 \end{cases} \\
D_i &= \mathbb{1}[y_i^B \leq \bar{y} \wedge Z_i = 1].
\end{aligned}
$$

DID-IV: Illustration

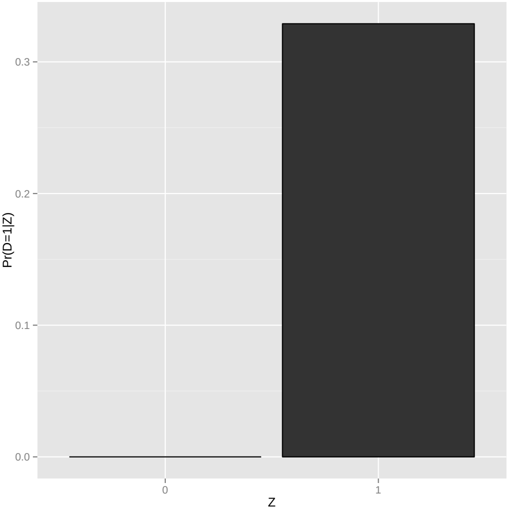

Figure: Illustration of the DID-IV assumptions: First Stage

DID-IV Assumptions: First Stage Rank Condition

Assumption (Strong DID First Stage Full Rank)

We have two periods, Before and After, denoted $\{A, B\}$. We assume that the treatment is not available in period B. We also have:

$$\Pr(D_i = 1 | Z_i = 1) \neq \Pr(D_i = 1 | Z_i = 0) = 0.$$

This assumption is not valid in Duflo (2001).

DID-IV: Illustration with Parallel Trends

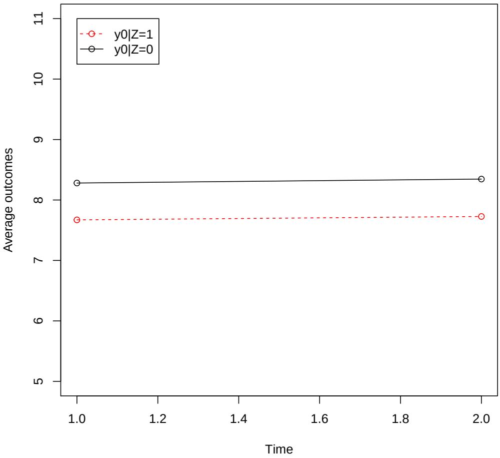

Figure: Evolution of average potential outcomes over time

DID-IV: Illustration with Parallel Trends

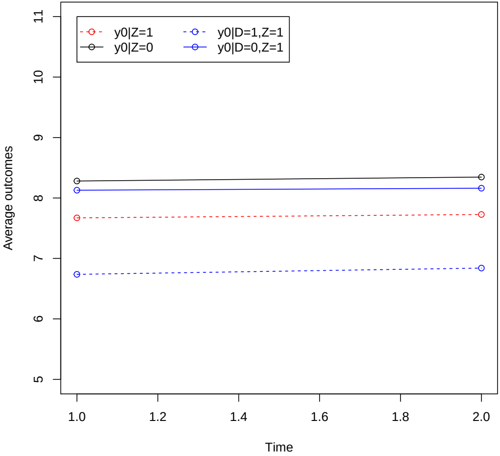

Figure: Evolution of average potential outcomes over time

Parallel Trends: Illustration

In our sample:

- $\hat{\mathbb{E}}[y_i^0|Z_i = 1] - \hat{\mathbb{E}}[y_i^B|Z_i = 1] = 0.0563$ and $\hat{\mathbb{E}}[y_i^0|Z_i = 0] - \hat{\mathbb{E}}[y_i^B|Z_i = 0] = 0.0654$
- $\hat{\mathbb{E}}[y_i^0|Z_i = 1] - \hat{\mathbb{E}}[y_i^0|Z_i = 0] =$ -0.6188 and $\hat{\mathbb{E}}[y_i^B|Z_i = 1] - \hat{\mathbb{E}}[y_i^B|Z_i = 0] =$ -0.6097
- $\hat{\mathbb{E}}[y_i^0|D_i = 1, Z_i = 1] - \hat{\mathbb{E}}[y_i^B|D_i = 1, Z_i = 1] = 0.1037$ and $\hat{\mathbb{E}}[y_i^0|D_i = 0, Z_i = 1] - \hat{\mathbb{E}}[y_i^B|D_i = 0, Z_i = 1] = 0.0331$
- $\hat{\mathbb{E}}[y_i^0|D_i = 1, Z_i = 1] - \hat{\mathbb{E}}[y_i^0|D_i = 0, Z_i = 1] =$ -1.3212 and $\hat{\mathbb{E}}[y_i^B|D_i = 1, Z_i = 1] - \hat{\mathbb{E}}[y_i^B|D_i = 0, Z_i = 1] =$ -1.3918
- $\hat{\mathbb{E}}[y_i^0|D_i = 1] - \hat{\mathbb{E}}[y_i^B|D_i = 1] = 0.1037$ and $\hat{\mathbb{E}}[y_i^0|D_i = 0] - \hat{\mathbb{E}}[y_i^B|D_i = 0] = 0.0434$
- $\hat{\mathbb{E}}[y_i^0|D_i = 1] - \hat{\mathbb{E}}[y_i^0|D_i = 0] =$ -1.3801 and $\hat{\mathbb{E}}[y_i^B|D_i = 1] - \hat{\mathbb{E}}[y_i^B|D_i = 0] =$ -1.4404

DID-IV Assumptions: Parallel Trends for the Instrument

Assumption (IV Parallel Trends)

We assume that:

$$\mathbb{E}[Y^0_{i,A}|Z_i = 1] - \mathbb{E}[Y^0_{i,B}|Z_i = 1] = \mathbb{E}[Y^0_{i,A}|Z_i = 0] - \mathbb{E}[Y^0_{i,B}|Z_i = 0].$$

Identification of TT with DID-IV

Theorem (Identification of TT with DID-IV)

Under Strong DID First Stage Full Rank and IV Parallel Trends, TT is identified by the Wald-DID estimator:

$$\Delta^Y_{WaldDID} = \Delta^{Y_A}_{TT},$$

with:

$$\Delta^Y_{WaldDID} = \frac{\mathbb{E}[Y_{i,A}|Z_i=1] - \mathbb{E}[Y_{i,A}|Z_i=0] - (\mathbb{E}[Y_{i,B}|Z_i=1] - \mathbb{E}[Y_{i,B}|Z_i=0])}{\Pr(D_{i,A}=1|Z_i=1) - \Pr(D_{i,A}=1|Z_i=0) - (\Pr(D_{i,B}=1|Z_i=1) - \Pr(D_{i,B}=1|Z_i=0))}.$$

Proof

Under IV Parallel Trends, we have:

$$\mathbb{E}[Y^0_{i,A}|Z_i = 1] = \mathbb{E}[Y^0_{i,A}|Z_i = 0] + (\mathbb{E}[Y^0_{i,B}|Z_i = 1] - \mathbb{E}[Y^0_{i,B}|Z_i = 0])$$

As a consequence, the numerator of the Wald-DID estimator is:

$$\begin{aligned}
\mathbb{E}[Y_{i,A}|Z_i = 1] - \mathbb{E}[Y_{i,A}|Z_i = 0] - (\mathbb{E}[Y_{i,B}|Z_i = 1] - \mathbb{E}[Y_{i,B}|Z_i = 0])& \\
= \mathbb{E}[Y_{i,A}|Z_i = 1] - \mathbb{E}[Y^0_{i,A}|Z_i = 1]& \\
= \mathbb{E}[Y^1_{i,A}|D_i = 1, Z_i = 1]\Pr(D_i = 1|Z_i = 1) + \mathbb{E}[Y^0_{i,A}|D_i = 0, Z_i = 1]\Pr(D_i = 0|Z_i = 1)& \\
- \mathbb{E}[Y^0_{i,A}|D_i = 1, Z_i = 1]\Pr(D_i = 1|Z_i = 1) + \mathbb{E}[Y^0_{i,A}|D_i = 0, Z_i = 1]\Pr(D_i = 0|Z_i = 1)& \\
= \left(\mathbb{E}[Y^1_{i,A}|D_i = 1, Z_i = 1] - \mathbb{E}[Y^0_{i,A}|D_i = 1, Z_i = 1]\right)\Pr(D_i = 1|Z_i = 1)& \\
= \Delta^{Y_A}_{TT}\Pr(D_i = 1|Z_i = 1),&
\end{aligned}$$

where the last equality uses the fact that $D_i = 1$ is a subset of $Z_i = 1$. Using Strong DID First Stage Full Rank proves that the denominator of the Wald-DID estimator is equal to $\Pr(D_i = 1|Z_i = 1)$, which proves the result.

The Direct Wald-DID Estimator

$$\hat{\Delta}^{Y}_{WaldDID} =$$

$$\frac{\frac{1}{\sum_{i=1}^{N} Z_i}\sum_{i=1}^{N} Y_{i,A}Z_i - \frac{1}{\sum_{i=1}^{N}(1-Z_i)}\sum_{i=1}^{N} Y_{i,A}(1-Z_i) - \left(\frac{1}{\sum_{i=1}^{N} Z_i}\sum_{i=1}^{N} Y_{i,B}Z_i - \frac{1}{\sum_{i=1}^{N}(1-Z_i)}\sum_{i=1}^{N} Y_{i,B}(1-Z_i\right.}{\frac{1}{\sum_{i=1}^{N} Z_i}\sum_{i=1}^{N} D_{i,A}Z_i - \frac{1}{\sum_{i=1}^{N}(1-Z_i)}\sum_{i=1}^{N} D_{i,A}(1-Z_i) - \left(\frac{1}{\sum_{i=1}^{N} Z_i}\sum_{i=1}^{N} D_{i,B}Z_i - \frac{1}{\sum_{i=1}^{N}(1-Z_i)}\sum_{i=1}^{N} D_{i,B}(1-Z_i\right.}$$

Under Strong DID First Stage Full Rank, the denominator simplifies to $\frac{1}{\sum_{i=1}^{N} Z_i}\sum_{i=1}^{N} D_iZ_i$.

The Direct Wald-DID Estimator: Illustration

In our sample:

- $\hat{\mathbb{E}}[y_i|Z_i = 1] - \hat{\mathbb{E}}[y_i^B|Z_i = 1] = 0.1104$ and $\hat{\mathbb{E}}[y_i|Z_i = 0] - \hat{\mathbb{E}}[y_i^B|Z_i = 0] = 0.0654$
- The numerator of the Wald-DID estimator is thus: $\hat{\mathbb{E}}[y_i|Z_i = 1] - \hat{\mathbb{E}}[y_i^B|Z_i = 1] - (\hat{\mathbb{E}}[y_i|Z_i = 0] - \hat{\mathbb{E}}[y_i^B|Z_i = 0]) =$0.1104-0.0654=0.045
- $\hat{\mathbb{E}}[y_i|Z_i = 1] - \hat{\mathbb{E}}[y_i|Z_i = 0] =$ -0.5646 and $\hat{\mathbb{E}}[y_i^B|Z_i = 1] - \hat{\mathbb{E}}[y_i^B|Z_i = 0] =$ -0.6097
- The numerator of the Wald-DID estimator is thus: $\hat{\mathbb{E}}[y_i|Z_i = 1] - \hat{\mathbb{E}}[y_i|Z_i = 0] - (\hat{\mathbb{E}}[y_i^B|Z_i = 1] - \hat{\mathbb{E}}[y_i^B|Z_i = 0]) =$-0.5646+0.6097=0.045
- The denominator of the Wald-DID estimator is $\hat{\Pr}(D_i = 1|Z_i = 1) =$0.3289
- The Wald-DID estimator is thus: 0.045÷0.3289=0.1369

The Value of TT in our Illustration

$$\Delta^y_{TT} = \bar{\alpha} + \theta\mathbb{E}[\mu_i|\mu_i + U^B_i \leq \bar{y} \wedge \mu^S_i \leq 0]$$

To compute the expectation of a doubly censored normal, I use the package tmvtnorm.
The value of Δ^y_{TT} in our illustration is: 0.1714.

DID-IV: illustration

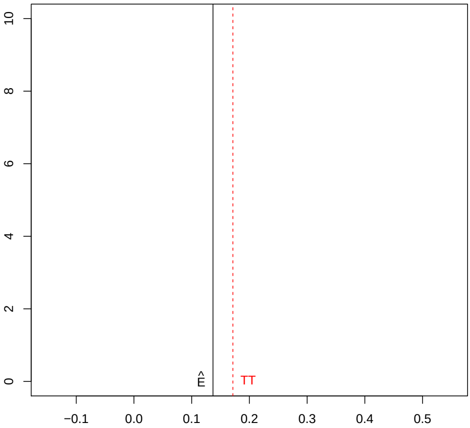

Figure: DID-IV

The Pooled 2SLS DID Estimator

Let's estimate the following regression with $t_i Z_i$ as an instrument for $t_i D_i$ and Z_i as an instrument for D_i:

$$Y_i = \alpha + \delta t_i + \gamma D_i + \beta t_i D_i + U_i$$

$\hat{\beta}_{2SLS}$ in the previous regression is the Pooled 2SLS DID estimator.

```r
y.pool <- c(y, yB)
Ds.pool <- c(Ds, Ds)
Z.pool <- c(Z, Z)
t <- c(rep(1, N), rep(0, N))
t.D <- t * Ds.pool
t.Z <- t * Z.pool
```

The Pooled 2SLS DID-IV Estimator: Illustration

```
reg.iv.did.pooled.2sls <- ivreg(y.pool   t + Ds.pool +
t.D | t + Z.pool + t.Z)
```

In our illustration, $\hat{\beta}_{2SLS} =$0.1369.

The Fixed Effects DID-IV Estimator

Estimating the following equation with $t_i Z_i$ as an instrument for $t_i D_i$:

$$Y_{i,t} = \mu_i + \delta_t + \beta t_i D_i + U_i$$

$\hat{\beta}_{IVFE}$ in the previous regression is the Fixed Effects DID-IV estimator.

```
data.panel <- cbind(c(seq(1, N), seq(1, N)), t, y.pool, Ds
colnames(data.panel) <- c("Individual", "time", "y", "Ds",
data.panel <- as.data.frame(data.panel)
```

The Fixed Effects DID-IV Estimator in Our Illustration

You have to load the library plm.

```
reg.iv.did.fe <- plm(y   time + t.D | time + t.Z, data
= data.panel, index = c("Individual", "time"), model =
"within")
```

In our illustration, $\hat{\beta}_{IVFE} =$0.1369.

The First Difference DID-IV Estimator

Estimate the following equation with Z_i as an instrument for D_i:

$$Y_{i,A} - Y_{i,B} = \delta + \beta D_i + U_i$$

$\hat{\beta}_{IVFD}$ in the previous regression estimated by 2SLS is the First Difference DID estimator.

The First Difference DID Estimator in Our Illustration

You have to load the library plm.

```
reg.iv.did.fd <- plm(y   time + t.D | time + t.Z, data
= data.panel, index = c("Individual", "time"), model =
"fd")
```

In our illustration, $\hat{\beta}_{IVFD} =$0.1369.

All Four DID-IV Estimators are Equivalent

Theorem (Equivalence of DID-IV Estimators)

With panel data and two periods of observation, we have:

$$\hat{\Delta}^{Y}_{WaldDID} = \hat{\beta}_{2SLS} = \hat{\beta}_{IVFE} = \hat{\beta}_{IVFD}.$$

Proof

To do

Sampling Noise with DID-IV in Panels: Illustration

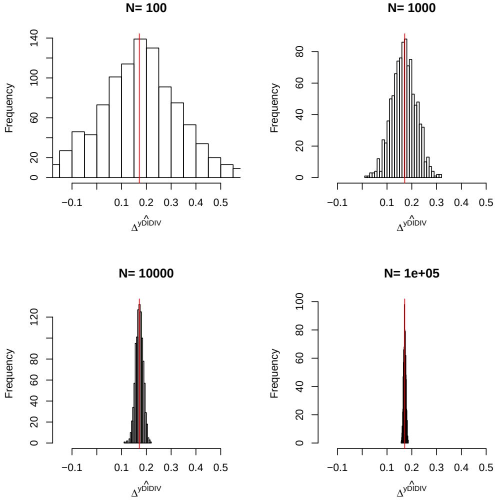

Figure: Distribution of the DID-IV estimator over replications of panels of different sizes

Sampling Noise with DID-IV in Cross Sections: Illustration

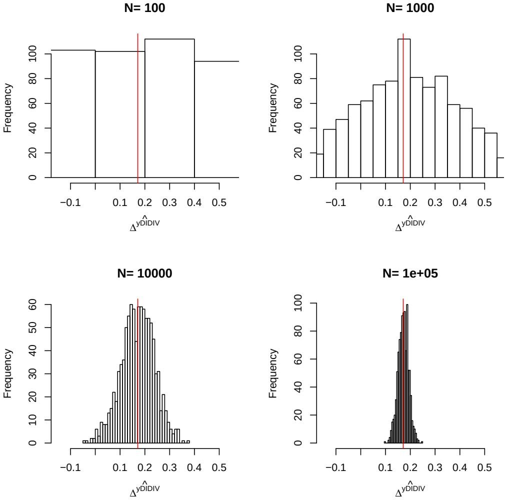

Figure: Distribution of the DID-IV estimator over replications of repeated cross sections of different sizes

Inference with DID-IV In Panel Data

- True 99% sampling noise (from the simulations) is 0.2495
- 99% sampling noise estimated using default FE standard errors is 0.299
- 99% sampling noise estimated using heteroskedasticity robust FE standard errors is 0.1172

Inference with DID-IV In Repeated Cross Sections

- True 99% sampling noise (from the simulations) is 1.1602
- 99% sampling noise estimated using default OLS standard errors is 0.3066
- 99% sampling noise estimated using heteroskedasticity robust OLS standard errors is 0.3656

```
reg.iv.did.pooling <- plm(y   time + Ds + t.D | time
+ Z + t.Z, data = data.panel, index = c("Individual",
"time"), model = "pooling")
```

- 99% sampling noise estimated using corrected OLS standard errors is 0.3066
- 99% sampling noise estimated using heteroskedasticity robust corrected OLS standard errors is 0.0426

DID-IV Assumptions: Weak First Stage Rank Condition

Assumption (Weak DID First Stage Full Rank)

We have two periods, Before and After, denoted $\{A, B\}$*. The treatment is available in both periods, but its takeover increases disproportionately among those with* $Z_i = 1$*:*

$$\begin{aligned}\Pr(D_{i,A} = 1|Z_i = 1) - \Pr(D_{i,B} = 1|Z_i = 1) \\ > \Pr(D_{i,A} = 1|Z_i = 0) - \Pr(D_{i,B} = 1|Z_i = 0).\end{aligned}$$

This assumption is valid in Duflo (2001).

DID-IV In Our Illustration

$$\mu_i = \mu_i^S + \mu_i^d + \mu_i^U + \bar{\mu}$$

$$\mu_i^S \sim \mathcal{N}(0, \frac{1}{3}\sigma_\mu^2)$$

$$\mu_i^d \sim \mathcal{N}(0, \frac{1}{3}\sigma_\mu^2)$$

$$\mu_i^U \sim \mathcal{N}(0, \frac{1}{3}\sigma_\mu^2)$$

$$E_{i,B} = \begin{cases} 1 & \text{if } \mu_i^d \leq -0.5 \wedge Z_i = 1 \\ 1 & \text{if } \mu_i^d \leq 0.25 \wedge Z_i = 0 \end{cases}$$

$$E_{i,A} = \begin{cases} 1 & \text{if } \mu_i^d \leq 0 \wedge Z_i = 1 \\ 1 & \text{if } \mu_i^d \leq 0.85 \wedge Z_i = 0 \end{cases}$$

$$D_{i,t} = \mathbb{1}[y_{i,BB} \leq \bar{y} \wedge E_{i,t} = 1].$$

The Model Used in the Simulations

$$
\begin{aligned}
y_{i,A}^0 &= \mu_i + \delta + U_{i,A}^0 \\
y_{i,A}^1 &= \mu_i(1+\theta) + \bar{\alpha} + \delta + U_{i,A}^0 + \eta_i \\
y_{i,BB} &= \mu_i + U_{i,BB} \\
U_{i,B}^0 &= \rho U_{i,BB} + \epsilon_{i,B} \\
U_{i,A}^0 &= \rho U_{i,B} + \epsilon_{i,A} \\
U_{i,BB} &\sim \mathcal{N}(0, \sigma_U^2) \\
\epsilon_{i,t} &\sim \mathcal{N}(0, \sigma_\epsilon^2) \\
\eta_i &\sim \mathcal{N}(0, \sigma_\eta^2)
\end{aligned}
$$

DID-IV: Illustration

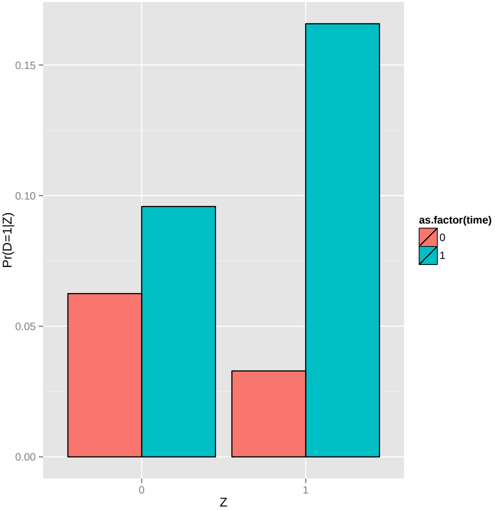

Figure: Illustration of the DID-IV assumptions: Weak First Stage

DID-IV: Illustration with Parallel Trends

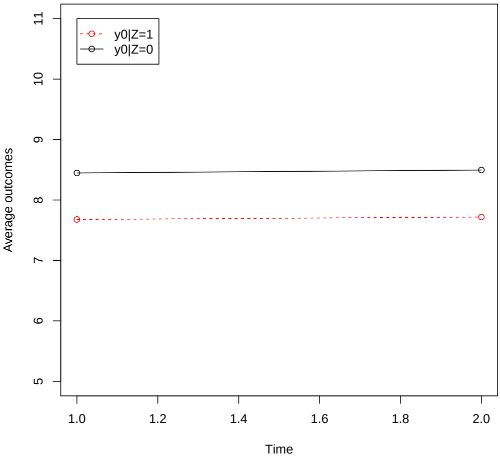

Figure: Evolution of average potential outcomes over time

Parallel Trends: Illustration

In our sample:

- $\hat{\mathbb{E}}[y^0_{i,A}|Z_i = 1] - \hat{\mathbb{E}}[y^0_{i,B}|Z_i = 1] = 0.0407$ and $\hat{\mathbb{E}}[y^0_{i,A}|Z_i = 0] - \hat{\mathbb{E}}[y^0_{i,B}|Z_i = 0] = 0.0491$

The Direct Wald-DID Estimator: Illustration

In our sample:

- $\hat{\mathbb{E}}[y_{i,A}|Z_i = 1] - \hat{\mathbb{E}}[y_{i,B}|Z_i = 1] = 0.0616$ and $\hat{\mathbb{E}}[y_{i,A}|Z_i = 0] - \hat{\mathbb{E}}[y_{i,B}|Z_i = 0] = 0.0559$
- The numerator of the Wald-DID estimator is thus: $\hat{\mathbb{E}}[y_{i,A}|Z_i = 1] - \hat{\mathbb{E}}[y_{i,B}|Z_i = 1] - (\hat{\mathbb{E}}[y_{i,A}|Z_i = 0] - \hat{\mathbb{E}}[y_{i,B}|Z_i = 0])$ =0.0616-0.0559=0.0057
- The denominator of the Wald-DID estimator is $\hat{\Pr}(D_{i,A} = 1|Z_i = 1) - \hat{\Pr}(D_{i,B} = 1|Z_i = 1) - (\hat{\Pr}(D_{i,A} = 1|Z_i = 0) - \hat{\Pr}(D_{i,B} = 1|Z_i = 0))$ $=0.1329-0.0333=0.0996$
- The Wald-DID estimator is thus: $0.0057 \div 0.0996 = 0.0577$

Identification of TT with DID-IV under Constant Effect

Theorem (Identification of TT with DID-IV)

Under Weak DID First Stage Full Rank, IV Parallel Trends and Constant Treatment Effect, TT is identified by the Wald-DID estimator:

$$\Delta^{Y}_{WaldDID} = \Delta^{Y_A}_{TT}.$$

Proof

We have, for $t \in \{A, B\}$ and $d \in \{0, 1\}$:

$$\mathbb{E}[Y_{i,t}|Z_i = d] = \mathbb{E}[Y^0_{i,t}|Z_i = d] + \mathbb{E}[Y^1_{i,t} - Y^0_{i,t}|D_{i,t} = 1, Z_i = d] \Pr(D_{i,t} = 1|Z_i = d)$$

Under Constant Treatment Effect, $\mathbb{E}[Y^1_{i,t} - Y^0_{i,t}|D_{i,t} = 1, Z_i = d] = \Delta^Y_{TT}$. As a consequence, the numerator of the Wald-DID estimator writes as follows:

$$\begin{aligned}
&\mathbb{E}[Y_{i,A}|Z_i = 1] - \mathbb{E}[Y_{i,A}|Z_i = 0] - (\mathbb{E}[Y_{i,B}|Z_i = 1] - \mathbb{E}[Y_{i,B}|Z_i = 0]) \\
&\qquad = \mathbb{E}[Y^0_{i,A} - Y^0_{i,B}|Z_i = 1] + \Delta^Y_{TT} \left(\Pr(D_{i,A} = 1|Z_i = 1) - \Pr(D_{i,B} = 1|Z_i = 1)\right) \\
&\qquad\quad - \mathbb{E}[Y^0_{i,A} - Y^0_{i,B}|Z_i = 0] - \Delta^Y_{TT} \left(\Pr(D_{i,A} = 1|Z_i = 0) - \Pr(D_{i,B} = 1|Z_i = 0)\right).
\end{aligned}$$

Using the Parallel Trend Assumption and dividing by the denominator of the Wald-DID estimator (which is non null under Weak First Stage Full Rank) yields the result.

What Happens When Treatment Effects Are Not Constant?

	Period 0	Period 1
Control Group	Always Treated: Y(1)	Always Treated: Y(1)
	Switchers: Y(0)	Switchers: Y(1)
	Never Treated: Y(0)	Never Treated: Y(0)
Treatment Group	Always Treated: Y(1)	Always Treated: Y(1)
	Switchers: Y(0)	Switchers: Y(1)
	Never Treated: Y(0)	Never Treated: Y(0)

What Happens When Treatment Effects Are Not Constant?

Theorem (Wald-DID with Non Constant Treatment Effect)

Under Weak DID First Stage Full Rank, IV Parallel Trends and Common Effect of Time on Both Potential Outcomes, the Wald-DID estimator identifies a weighted average of treatment effects with possibly negative weights:

$$num(\Delta^Y_{WaldDID}) = \mathbb{E}[\Delta^Y_i|S^1_i = 1]\Pr(S^1_i = 1) - \mathbb{E}[\Delta^Y_i|S^0_i = 1]\Pr(S^0_i = 1)$$

Proof

de Chaisemartin and D'haultfoeuille (2015), theorem 3.1.

What Happens When Treatment Effects Are Not Constant?

- If $\Pr(S_i^0 = 1) = 0$, Wald-DID recovers the effect on switchers under Common Effect of Time on Both Potential Outcomes
- de Chaisemartin and D'haultfoeuille (2015) propose a time-corrected Wald-DID estimator that estimates the effect on switchers when $\Pr(S_i^0 = 1) = 0$ under a much weaker assumption of Common Trends Within Treatment Status.
- They also propose bounds when $\Pr(S_i^0 = 1) \neq 0$

Sampling Noise with DID-IV in Panels: Illustration

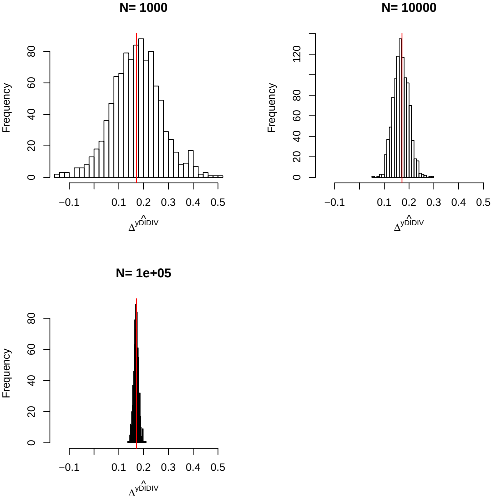

Figure: Distribution of the DID-IV estimator over replications of panels of different sizes