Heterogeneity, Endogeneity and Causal Effect Estimation

Kevin Sheppard
`https://www.kevinsheppard.com`

Oxford MFE

This version: March 11, 2020

March 2020

Causal Effect Estimation

- Potential Outcomes
- Challenges in Effect Estimation
- Experimental and Quasi-Experimenal Data
  - Randomized Controlled Experiments and ATE
  - Imperfect Compliance and LATE
- Observational Data
  - Regression Discontinuity
  - Difference-in-Difference
  - Panel Models

Potential Outcomes Framework

- Observed outcome for individual or firm i

$$Y_i$$

- D_i is the treatment status variable for individual i

$$D_i = \begin{cases} 0 & \text{if untreated} \\ 1 & \text{treated} \end{cases}$$

- Outcome variable is determined by

$$Y_i = \beta_{0i} + \beta_{1i} D_i$$

- β_{1i} is a heterogeneous treatment effect for individual i
- Also known as the *potential outcomes* model
- Two outcomes

$$Y_i(0) = \beta_{0i} \text{ and } Y_i(1) = \beta_{0i} + \beta_{1i}$$

Key Measures

Definition (Average Treatment Effect (ATE))

The Average Treatment Effect measures the average effect of treatment across the entire population

$$ATE = \mathrm{E}[\beta_{1i}] = \mathrm{E}[Y_i(1)] - \mathrm{E}[Y_i(0)]$$

Definition (Average Treatment Effect on the Treated (TOT))

The Average Treatment Effect on the Treated measures the effect of treatment on the treated

$$TOT = \mathrm{E}\left[\beta_{1i}|D=1\right] = \mathrm{E}\left[Y_i(1)|D=1\right] - \mathrm{E}\left[Y_i(0)\,|D=1\right]$$

ATE and *TOT*

- *ATE* is a weighted average

$$ATE = \omega TOT + (1 - \omega)\, TUT$$

- Average Treatment Effect on the Untreated (*TUT*)

$$TUT = \mathrm{E}\left[\beta_{1i} | D = 0\right] = \mathrm{E}\left[\overset{Y_i(1)}{Y_{1i}} | D = 0\right] - \mathrm{E}\left[\overset{Y_i(0)}{Y_{0i}} | D = 0\right]$$

- $\omega = \Pr[D = 1]$ if the probability treated
- Should we measure ATE or TOT?
  - TOT makes sense when treatment is non-compulsory
    - Individuals who do not undertake treatment are not relevant for cost-benefit calculation
  - ATE is more sensible for mandatory programs
    - Measures the effect on both those who would like to participate and those who would not

Naïve estimation

- Estimate the regression on observed data

No i

$$Y_i = b_0 + b_1 D_i + \epsilon_i$$

$Y_i(0) \mid D=1$

  - $\hat{b}_0 \xrightarrow{p} \mathrm{E}\left[Y_i | D=0\right]$
  - $\hat{b}_1 \xrightarrow{p} \mathrm{E}\left[Y_i | D=1\right] - \mathrm{E}\left[Y_i | D=0\right]$.

B_{0i} B_{1i}

- Leads to selection bias

$$\underbrace{\mathrm{E}\left[Y_i | D=1\right] - \mathrm{E}\left[Y_i | D=0\right]}_{\text{Observed Effect}} = \underbrace{\mathrm{E}\left[Y_i(1) | D=1\right] - \mathrm{E}\left[Y_i(0) | D=1\right]}_{\text{Avg. Treatment Effect on the Treated (TOT)}}$$

$$+ \underbrace{\mathrm{E}\left[Y_i(0) | D=1\right] - \mathrm{E}\left[Y_i(0) | D=0\right]}_{\text{Selection Bias (SB)}}$$

(1) (2)

- In terms of the regression

$$\underbrace{\mathrm{E}\left[\hat{b}_1\right]}_{\text{Observed Effect}} = \underbrace{\mathrm{E}\left[\beta_{1i} | D=1\right]}_{\text{TOT}} + \underbrace{\mathrm{E}\left[\beta_{0i} | D=1\right] - \mathrm{E}\left[\beta_{0i} | D=0\right]}_{\text{Selection Bias (SB)}}$$

- SB is the difference in the no-treatment outcomes for the treated and untreated

(Missing) Counterfactuals

- Fundamental problem: Cannot see counterfactual

	Treatment (D_i) 0	1
Observe	$Y_i(0) = \beta_{0i}$	$Y_i(1) = \beta_{0i} + \beta_{1i}$
Counterfactual	$Y_i(1) = \beta_{0i} + \beta_{1i}$	$Y_i(0) = \beta_{0i}$

- No data on $Y_i(1)$ when $D_i = 0$ and $Y_i(0)$ when $D_i = 1$
- TOT measures the effect conditional on receiving treatment
  - Missing counterfactual: $\mathrm{E}\left[Y_i(0)|D=1\right]$
- Observed effect is contaminated with selection bias

Example: Financial Stress and Payday Loans

- Example: Financial Stress and Payday loans
- Outcome is a measure of financial distress: 90-days delinquent on a debt
- Treatment is taking out a payday loan
- TOT: Difference in delinquency if loan taken or not *given* loan wanted ($D = 1$)
- SB: Difference in outcome if loan not taken for those who want a loan and those who do not want a loan
  - Plausible TOT is negative but SB is positive
  - Positive SB if
  $$\mathrm{E}\left[\beta_{0i}|D=1\right] > \mathrm{E}\left[\beta_{0i}|D=0\right]$$
    - Default rates absent a loan are higher for loan takers than for non-takers
  - Observed effect could have either sign

Experimental and Quasi-Experimental Data

Randomization

- Randomization removes selection bias
- Well executed Randomized Controlled Trials are the gold standard for causal effect estimation
- A RCT ensures that

$$\{\beta_{0i}, \beta_{1i}\} \perp\!\!\!\perp D_i \text{ and } \{Y_i(0), Y_i(1)\} \perp\!\!\!\perp D_i$$

- Randomly give loans only to those seeking them
  - Creates group with $Y_i(0)$ *as if* $D = 1$

Independence and Conditioning

If Z and W are independent random variables, then
$\mathrm{E}[Z|W = w_1] = \mathrm{E}[Z|W = w_2] = \mathrm{E}[Z]$.

- Knowledge of W provides no information about Z.

Gains to Randomization

$\mathrm{E}\left[Y_i(0)|D=0\right] = \mathrm{E}\left[Y_i(0)|D=1\right]$ and $\mathrm{E}\left[\beta_{0i}|D=1\right] = \mathrm{E}\left[\beta_{0i}|D=0\right]$ since treatment independent of desire to be treated

TYPO

D=1 D=0

- Track outcomes of both groups

$$\underbrace{\mathrm{E}\left[Y_i|D=1\right] - \mathrm{E}\left[Y_i|D=0\right]}_{\text{Observed Effect with Randomization}} = \mathrm{E}\left[Y_i(1)|D=1\right] - \mathrm{E}\left[Y_i(0)|D=0\right]$$

$$= \mathrm{E}\left[Y_i(1)|D=1\right] - \mathrm{E}\left[Y_i(0)|D=1\right]$$

CF

- In the notation of a regression model

$$\begin{aligned}
\mathrm{E}\left[\hat{b}_1\right] &= \mathrm{E}\left[\beta_{1i}|D=1\right] + \left\{\mathrm{E}\left[\beta_{0i}|D=1\right] - \mathrm{E}\left[\beta_{0i}|D=0\right]\right\} \\
&= \mathrm{E}\left[\beta_{1i}|D=1\right] + \left\{\mathrm{E}\left[\beta_{0i}|D=1\right] - \mathrm{E}\left[\beta_{0i}|D=1\right]\right\} \\
&= \mathrm{E}\left[\beta_{1i}|D=1\right]
\end{aligned}$$

Issues Affecting RCT Validity

- Internal Validity: are the results valid for the sample used?
  - Is the assignment actual random?

$$X_i = \alpha + \beta D_i + \epsilon_i, \;\; H_0 : \beta = 0, H_1 : \beta \neq 0$$

  - Are participants complying?
  - Are there spill-overs of non-rival treatments to non-treated?
  - Hawthorne Effect: studying a subject changes their behavior
- External Validity: do the results generalize to a broader sample?
  - Is the RCT sample representative of the target population?
  - Are there other key personnel that are essential for success?

LATE: Local Average Treatment Effects

- Previous result requires perfect compliance
  - Treated if offered, not-treated if not offered
- When treatment is not random, or compliance is not perfect, simple estimators are not consistent
- Possible to use an *instrument* to recover a meaningful measure of treatment effect
- Measure is *local* in the sense that it measures the effect of a particular subgroup of the treated
- Notation
  - D_i is treatment status
  - Z_i is treatment *assignment* (offer to treat)
- Compliance
  - Perfect if $D_i = Z_i$
  - Imperfect if $D_i \neq Z_i$ for some i
- Z_i may be random even if D_i is not
  - Treatment assignment is made by lottery due to limited capacity (Z_i)
  - Treatment status *conditional on offer* depends on expected benefits (D_i)

System of Equations

- Leads to two-equation system

$$\text{Structural Equation } Y_i = \beta_{0i} + \beta_{1i} D_i$$
$$\text{Treatment Equation } D_i = \pi_{0i} + \pi_{1i} Z_i$$

- Causal chain $Z_i \rightarrow D_i \rightarrow Y_i$
- Treatment equation measures potential treatment status

$$D_i(z) = \pi_{0i} + \pi_{1i} z$$

  - $D_i(0) = \pi_{0i}$ is status when not assigned
  - $D_i(1) = \pi_{0i} + \pi_{1i}$ is status when not assigned
  - Both $D_i(0)$ and $D_i(1)$ may be 0 or 1
- Treatment responsiveness π_{1i} is heterogeneous like treatment effect β_{1i}

Independence

Assumption (Independence)

The potential outcomes and potential treatment assignments are independent of Z_i

$$\{\beta_{0i}, \beta_{1i}, \pi_{0i}, \pi_{1i}\} \perp\!\!\!\perp Z_i$$

- Often described as *as if* randomly assigned
- Note that the instrument is independent of the potential treatment status
- Z_i does not affect the probability that either occur $(\pi_{\bullet i})$
- Z_i does not affect the outcomes if treatment is taken or not $(\beta_{\bullet i})$
- Is this a reasonable assumption?
  - Often plausible when Z_i is assigned using randomization (lottery)
  - Sometimes plausible for Z_i taken from observational data

Exclusion

Assumption (Exclusion Restriction)

The instrument does not appear in the structural equation so that only treatment assignment affect the outcome.

- Violations of the exclusion restriction mean that Z_i affects Y_i through more than just D_i

$\cancel{Z_i \to Y_i}$

- Classic example is when Z_i directly affects both Y_i and D_i
- In many cases, Z_i affects D_i *and* another variable X_i which in turn affects Y_i

$$\underline{Z_i \to D_i \to Y_i}, \quad Z_i \to X_i \to Y_i$$

- Suppose selection for a randomly assigned government funding program increases probability of program participation $(Z_i \to D_i)$
- If selection also increases the probability that a firm receives series B funding, than effect confounded with fund raising $(Z_i \to X_i)$
- Exclusion restriction ensures that Z does not affect the potential outcome

$$Y(0)_i = \beta_{0i} \text{ and } Y(1)_i = \beta_{0i} + \beta_{i1} \text{ for } Z \in \{0, 1\}$$

Instrumental Variable Estimation

- The 2SLS estimator obtained by
  1. Regress $D_i = p_0 + p_1 Z_i + \eta_i$ and retain $\hat{D}_i = \hat{p}_0 + \hat{p}_1 Z_i$
  2. Regress $Y_i = b_0 + b_1 \hat{D}_i + \epsilon_i$
- In large samples

weight

$$\hat{b}_1^{2SLS} \xrightarrow{p} \frac{\mathrm{E}[\beta_{1i}\pi_{1i}]}{\mathrm{E}[\pi_{1i}]} = \mathrm{E}\left[\beta_{1i}\frac{\pi_{1i}}{\mathrm{E}[\pi_{1i}]}\right] = LATE$$

- $LATE$ is a weighted average of treatment effects
- Weights are determined by responsiveness to treatment assignment
  - Holds if either of D_i or Z_i are not binary
- If effects are not heterogeneous ($\beta_{1i} = \beta_1$ or $\pi_{1i} = \pi_1$) then $LATE = ATE$

Types of Participants

- Useful to describe structure implied by D_i and Z_i
- Four types of program participants
  - Compliers: $D_i = Z_i$ ($\pi_{0i} = 0, \pi_{1i} = 1$)
  - Always-takers: $D_i = 1$ for any Z_i ($\pi_{0i} = 1, \pi_{1i} = 0$)
  - Never-takers: $D_i = 0$ for any Z_i ($\pi_{0i} = \pi_{1i} = 0$)
  - Defiers: $D_i = 1 - Z_i$ ($\pi_{0i} = 1, \pi_{1i} = -1$)

$D_i = \underset{0}{\pi_{0i}} + \underset{1}{\pi_{1i}} Z_i$

- Compliers are the ideal candidates and ultimately what we can measure
- Defiers invalidate measurement using the instrument
- *LATE* is determined only by compliers and defiers

No Defiers

Assumption (No Defiers)

There are no defiers, so that $\pi_{1i} \geq 0$.

With this additional assumption

$$LATE = \mathrm{E}\left[\beta_{1i} | \pi_{1i} = 1\right]$$

so that $LATE$ only measures the treatment response of the compliers.

Intention-to-Treat (ITT)

- Common to report the effect on the intention-to-treat

$$
\begin{aligned}
ITT &= \mathrm{E}\left[Y_i | Z = 1\right] - \mathrm{E}\left[Y_i | Z = 0\right] \\
&= \mathrm{E}\left[\beta_{1i}\pi_{1i}\right] \\
&= LATE \times \mathrm{E}\left[\pi_{1i}\right]
\end{aligned}
$$

$\Rightarrow LATE = \dfrac{ITT}{E[\pi_{1i}]} \leftarrow$ Compliers

- Difference in outcomes conditional on only the instrument which measures the intention to treat
- With perfect compliance $ITT = LATE$

0 1

$Prob(\pi_{1i}) = 1$

Example

The Oregon Health Insurance Experiment

- 2008 Medicaid expansion in US state or Oregon
- Used a lottery to choose participants from a waiting list
- Constructed a control group from non-winners
- Not everyone selected participated in the program (imperfect compliance)
- Non-selected prohibited from participation

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J.P., Allen, H., Baicker, K. and Oregon Health Study Group, 2012. The Oregon Health Insurance Experiment: evidence from the first year. The Quarterly journal of economics, 127(3), pp.1057-1106.

- Estimating Intent-to-Treat (ITT)

Z_i

$$Y_{ihj} = \beta_0 + \beta_1 LOTTERY_h + \mathbf{X}_{ih}\boldsymbol{\beta}_2 + \mathbf{V}_{ih}\boldsymbol{\beta}_3 + \epsilon_{ihj}$$

  - i: individual, h: household, j: domain of variable
  - $LOTTERY_h$ indicates household was selected by the lottery ($Z_i = 1$)
  - $\mathbf{X}_{ih}$ are required controls and $\mathbf{V}_{ih}$ are optional controls

- Estimating $LATE$

D_i

$$Y_{ihj} = \pi_0 + \pi_1 INSURANCE_{ih} + \mathbf{X}_{ih}\boldsymbol{\pi}_2 + \mathbf{V}_{ih}\boldsymbol{\pi}_3 + \nu_{ihj}$$

  - $INSURANCE_{ih}$ is a measure of insurance coverage (D_i)
  - Use 2SLS

D_i Z_i

$$INSURANCE_{ih} = \delta_0 + \delta_1 LOTTERY_h + \mathbf{X}_{ih}\boldsymbol{\delta}_2 + \mathbf{V}_{ih}\boldsymbol{\delta}_3 + \xi_{ihj}$$

  - Insurance is “Ever on Medicaid”

Results

The Oregon Health Insurance Experiment

TABLE VIII

FINANCIAL STRAIN (SURVEY DATA)

	Control mean (1)	ITT (2)	LATE (3)	p-values (4)
Any out of pocket medical expenses, last six months	0.555	−0.058	−0.200	[<0.0001]
	(0.497)	(0.0077)	(0.026)	{<0.0001}
Owe money for medical expenses currently	0.597	−0.052	−0.180	[<0.0001]
	(0.491)	(0.0076)	(0.026)	{<0.0001}
Borrowed money or skipped other bills to pay medical bills, last six months	0.364	−0.045	−0.154	[<0.0001]
	(0.481)	(0.0073)	(0.025)	{<0.0001}
Refused treatment because of medical debt, last six months	0.081	−0.011	−0.036	[0.01]
	(0.273)	(0.0041)	(0.014)	{0.01}
Standardized treatment effect		−0.089	−0.305	[<0.0001]
		(0.010)	(0.035)	

Observational Data

Difference-in-Difference

Difference-in-Difference

- Repeated cross-sections across multiple periods
- Examine evidence between treated group and control group
- Control chosen to be similar except for treatment
- Basic model

$$Y_{it} = \alpha_i + \beta_t + \epsilon_{it}$$

- Assume two periods, treatment only in second
- Treated individuals or firms have

$$Y_{it} = \alpha_i + \beta_t + \delta D_{it} + \epsilon_{it}$$

$$Y_{jt} = \alpha_j + \beta_t + \delta D_{jt} + \varepsilon_{jt}$$

Difference-in-Difference

- Two groups, $\mathcal{A}$ (treated) and $\mathcal{B}$ (untreated)
- Construct averages

$\alpha_A + B_1$

$\alpha_A + B_2 + \delta$

$$\mathrm{E}\left[\bar{Y}_{\mathcal{A}t}\right] = \alpha_{\mathcal{A}} + \beta_t + \delta I_{[t=2]}, \quad t = 1, 2$$
$$\mathrm{E}\left[\bar{Y}_{\mathcal{B}t}\right] = \alpha_{\mathcal{B}} + \beta_t$$

$\alpha_B + B_1$

- Difference across groups removes own effects $\alpha_B + B_2$

1   2

Treaty

$$\mathrm{E}\left[\bar{Y}_{\mathcal{A}2} - \bar{Y}_{\mathcal{A}1}\right] = \beta_2 - \beta_1 + \delta$$
$$\mathrm{E}\left[\bar{Y}_{\mathcal{B}2} - \bar{Y}_{\mathcal{B}1}\right] = \beta_2 - \beta_1$$

- Difference across time removes time effects

$$\mathrm{E}\left[\bar{Y}_{\mathcal{A}2} - \bar{Y}_{\mathcal{B}2}\right] = \alpha_{\mathcal{A}} - \alpha_{\mathcal{B}} + \delta$$
$$\mathrm{E}\left[\bar{Y}_{\mathcal{A}1} - \bar{Y}_{\mathcal{B}1}\right] = \alpha_{\mathcal{A}} - \alpha_{\mathcal{B}}$$

No B

- Solution is to difference twice

$$\mathrm{E}\left[\bar{Y}_{\mathcal{A}2} - \bar{Y}_{\mathcal{A}1} - \bar{Y}_{\mathcal{B}2} - \bar{Y}_{\mathcal{B}1}\right] = \delta$$

Difference-in-Difference

- Key assumption

Assumption (Counterfactual (Parallel Trends))

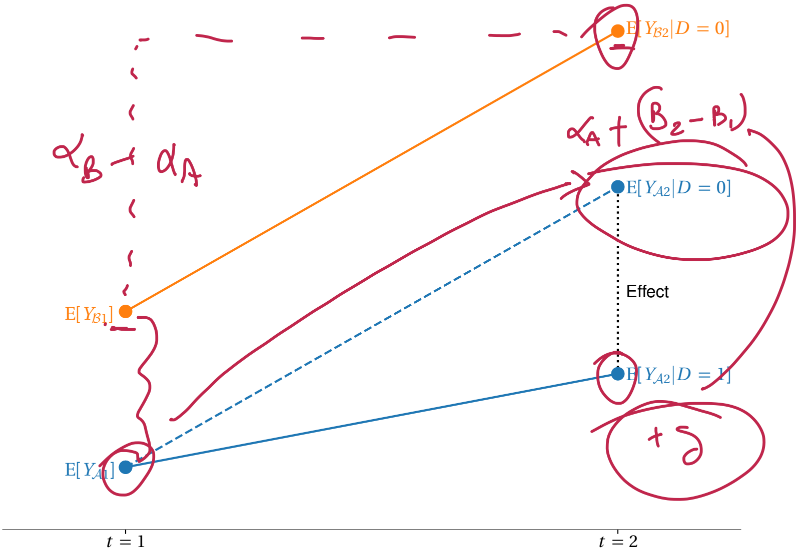
$$\overset{\beta_2 - \beta_1}{\mathrm{E}\left[Y_{i2} - Y_{i1} | D = 1\right]} = \overset{\beta_2 - \beta_1}{\mathrm{E}\left[Y_{i2} - Y_{i1} | D = 0\right]}$$

The growth rate does not depend on whether the unit is treated ($D = 1$) or untreated.

- Uses lagged average of $\mathcal{A}$ to remove group-specific effects
- Generates counter-factual using group $\mathcal{B}$

Difference-in-Difference

- Counter-factual can be interpreted as parallel line



Difference-in-Difference

- Estimated as a regression

$$Y_{it} = \alpha_i + \beta_t + \delta I_{[i \in \mathcal{A}, t=2]} + \epsilon_{it}$$

  - Extends trivially to include other controls

$$Y_{it} = \alpha_i + \beta_t + \delta I_{[i \in \mathcal{A}, t=2]} + \mathbf{X}_{it}\boldsymbol{\gamma} + \epsilon_{it}$$

- Uses dummy variables

$$Y_{it} = \alpha_{\mathcal{A}} I_{[i \in A]} + \alpha_{\mathcal{B}} I_{[i \in \mathcal{B}]} + \beta_1 I_{[t=1]} + \beta_2 I_{[t=2]} + \delta I_{[i \in \mathcal{A}, t=2]} + \epsilon_{it}$$

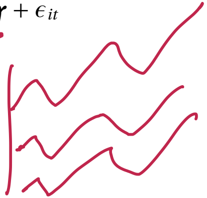
  - Extends to multiple groups and time periods

$$Y_{it} = \sum_{g=1}^{G} \alpha_g I_{[i \in \mathcal{G}_g]} + \sum_{j=1}^{T} \beta_j I_{[t=j]} + \delta I_{[t \geq \tau_i]} + \mathbf{X}_{it}\boldsymbol{\gamma} + \epsilon_{it}$$

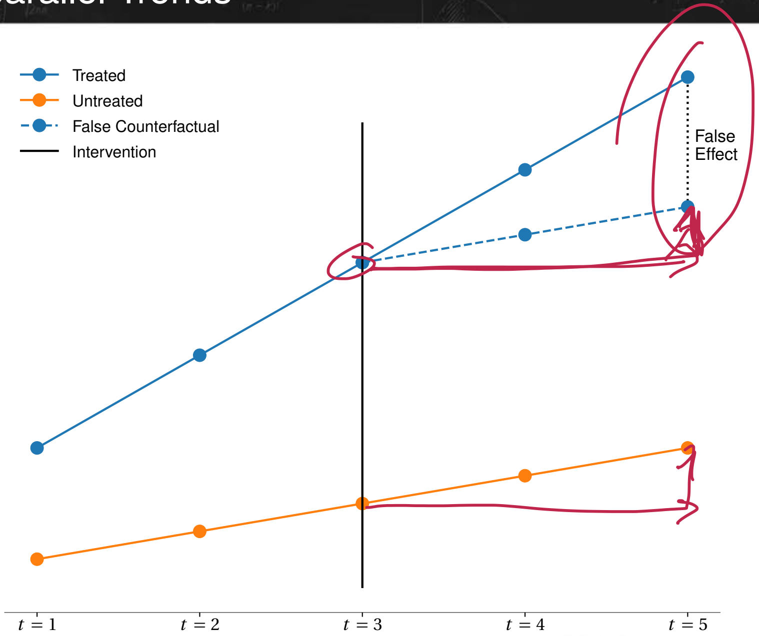
where τ_i is the treatment time for individual i

- Main issue: Violation of parallel trends assumption

$$\beta_{\mathcal{A}t} \neq \beta_{\mathcal{B}t}$$



Non-parallel Trends



Example

The Distortive Effects of Too Big To Fail: Evidence from the Danish Market for Retail Deposits

- Examine how uninsured deposits are affected by implicit *Too Big Too Fail* (TBTF) guarantees
- Variation introduced by changes in the limits of the amounts insured
- Main model

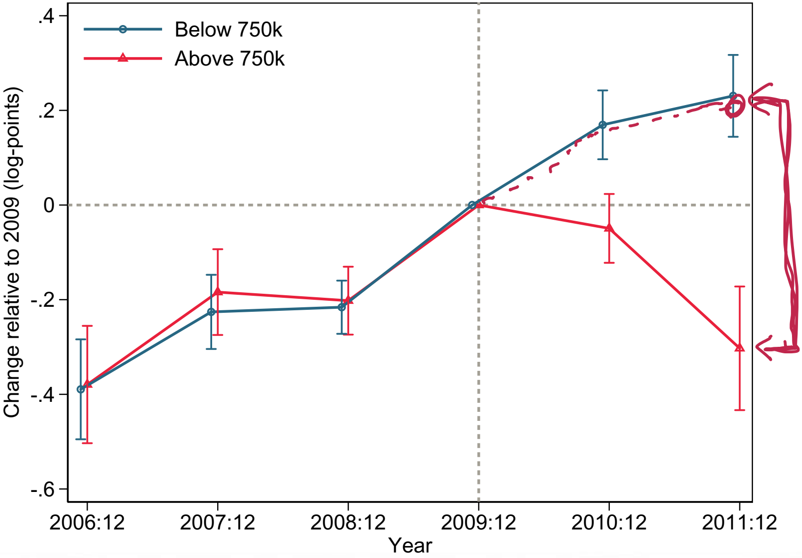
$$\ln(Deposits)_{btk} = \alpha + \beta \times Above_k + \beta_2 \times After_t + \beta_3 \times Above_k \times After_t + \mathbf{X}_i\boldsymbol{\gamma} + \epsilon_{btk}$$

  - Bank b, Year t
  - k indicates the range of deposits in DKK 50,000 bins (e.g., 0–50K, 50K–100K, ...)
  - Interest in the difference between accounts that remain insured with those that become uninsured
- Key variable is $Above_k \times After_t$ which would have a 0 coefficient if no effect

Iyer, R., Lærkholm Jensen, T., Johannesen, N., & Sheridan, A. (2019). The Distortive Effects of Too Big To Fail: Evidence from the Danish Market for Retail Deposits. The Review of Financial Studies, 32(12), 4653-4695.

Results

The Distortive Effects of Too Big To Fail: Evidence from the Danish Market for Retail Deposits



Results

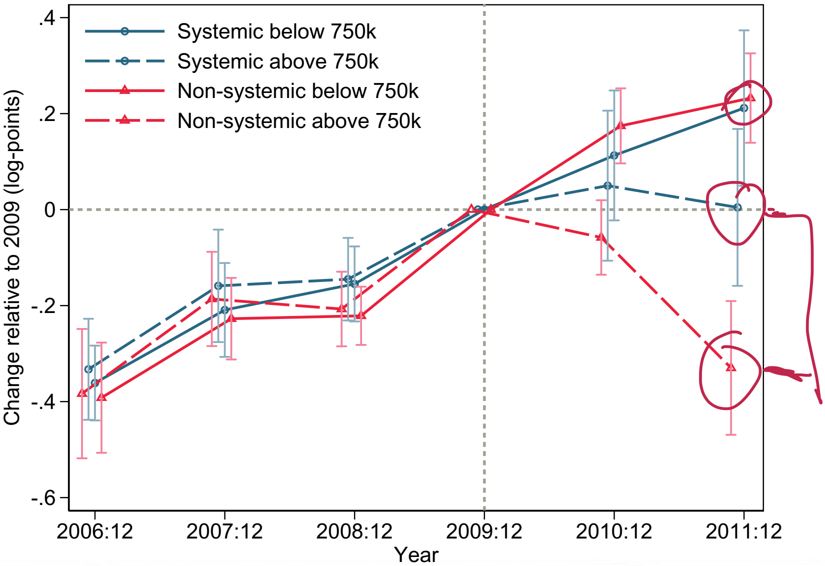
The Distortive Effects of Too Big To Fail: Evidence from the Danish Market for Retail Deposits

Table 2
Baseline results

	(1)	(2)	(3)	(4)
	Deposits (in logs)			
After reform × Above limit	−0.363***	−0.373***	−0.378***	−0.366***
	(0.0462)	(0.0430)	(0.0501)	(0.0500)
After reform	0.292***	0.304***	0.305***	
	(0.0403)	(0.0394)	(0.0456)	
Above limit	−0.578***	−0.593***		
	(0.0214)	(0.0187)		
Equity / debt (in 2007)	1.941			
	(1.452)			
Loans / assets (in 2007)	1.302			
	(0.834)			
Log of assets (in 2007)	0.802***			
	(0.0398)			
Constant	1.154			
	(0.773)			
Observations	3,507	3,507	3,507	3,507
R-squared	.869	.951	.970	.990
Bank FEs	No	Yes	Yes	Yes
Bank-range FEs	No	No	Yes	Yes
Bank-time FEs	No	No	No	Yes

Results

The Distortive Effects of Too Big To Fail: Evidence from the Danish Market for Retail Deposits



Regression Discontinuity

Regression Discontinuity

- Regression discontinuity exploits discrete jumps in treatment as a function of an observable X_i
- Individuals with $X_i \leq \overline{X}$ are untreated
- Individuals with $X_i \cancel{\leq} \overline{X}$ are treated

$X_i > \overline{X}$

- Theoretical framework examines the difference locally with $\overline{X}$

$$\mathrm{E}\left[Y_i | 0 < X_i - \overline{X} < h\right] - \mathrm{E}\left[Y_i | 0 < \overline{X} - X_i < h\right]$$

- h is a bandwidth parameter that shrinks to 0 as the sample size increase
- Intuition is that individuals are homogeneous local to $\overline{X}$

RDD Practice

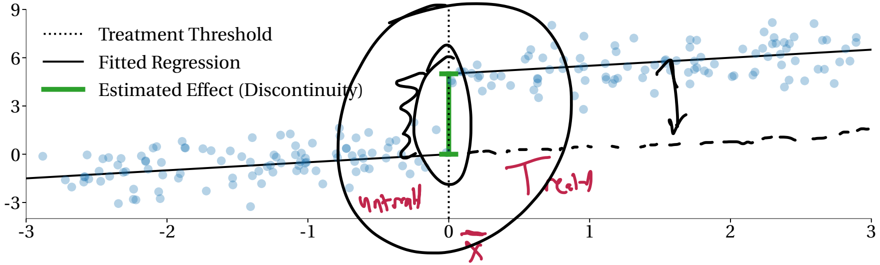
- In practice often included as part of a model

$$Y_i = \beta_0 + \beta_1 X_i + \delta I_{[X_i \geq \overline{X}]} + \epsilon_i$$

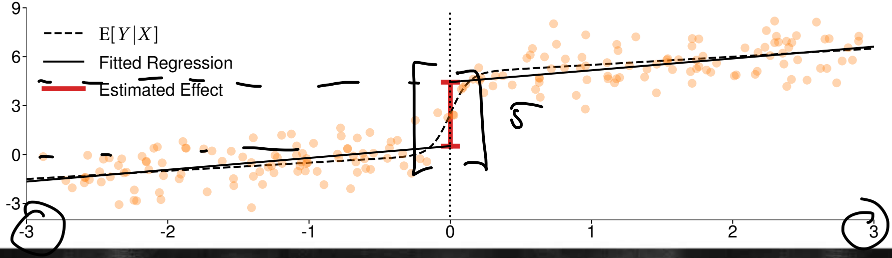
- Can use more sophisticated models
- Causality from model requires treatment indicator $I_{[X_i \geq \overline{X}]}$ to be uncorrelated with omitted variables
- Also requires functional form to be correct that that $\mathrm{E}\left[\epsilon_i | X\right] = 0$
  - Rules out neglected nonlinearities

Regression Discontinuity

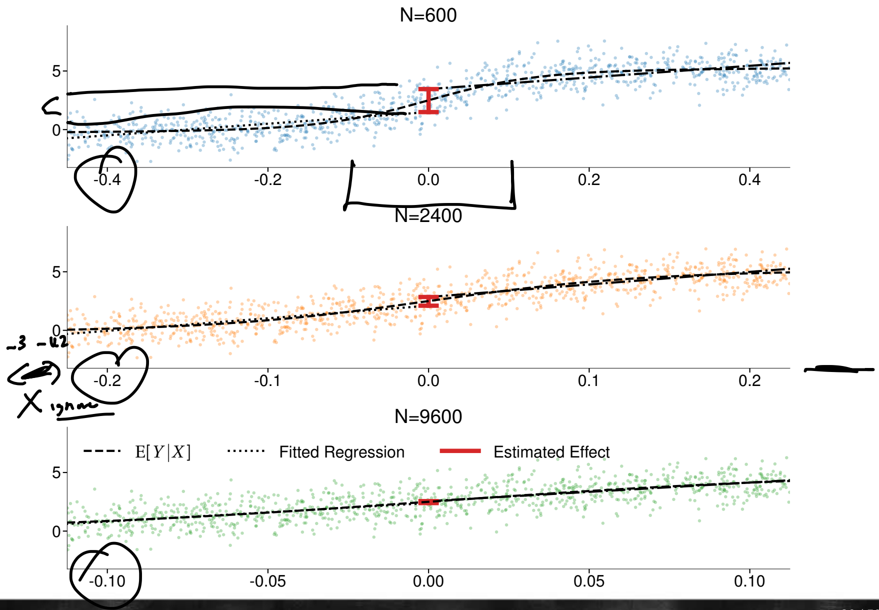
Linear Model with a Discontinuity



Neglected Nonlinearity



Local Estimation



Example

Does Corporate Social Responsibility Lead to Superior Financial Performance?

- Investigate the effect of CSR proposals that just pass or just fail board votes
- Key assumption: the difference between just passing and failing is *as if* random
  - Close failures and close passes are identical aside from the vote
- Regress abnormal returns on a dummy for passing
  - Returns computed using Carhart 4 factor model (FF3 + Momentum)
- Use RDD estimate is based on a small window near a tied vote
- Also considers a full model which is piece-wise polynomial

$$y_{it} = \beta \times D_{it} + P_l\left(v_{it}, \boldsymbol{\gamma}_l\right) + P_r\left(v_{it}, \boldsymbol{\gamma}_r\right)$$

  - $P_\bullet\left(v_{it}, \boldsymbol{\gamma}_\bullet\right)$ is a polynomial in the vote shape for left and right of passing

Flammer, C. (2015). Does corporate social responsibility lead to superior financial performance? A regression discontinuity approach. Management Science, 61(11), 2549-2568.

Results

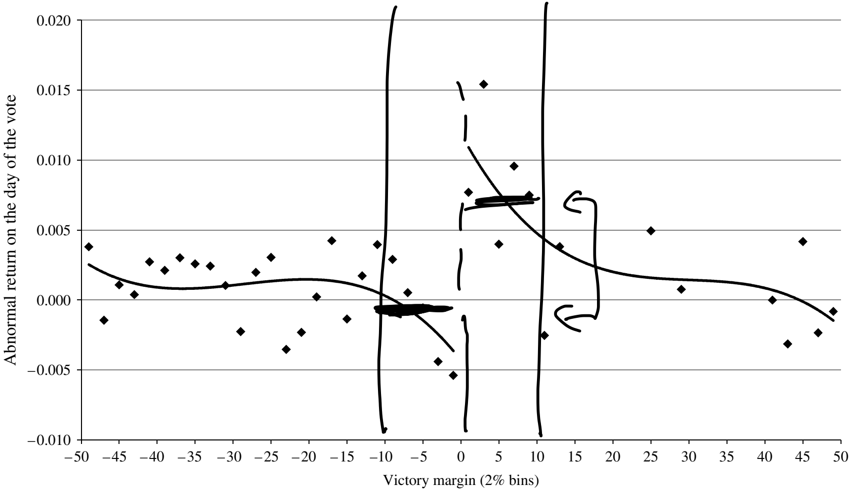
Does Corporate Social Responsibility Lead to Superior Financial Performance? A RD Approach

Table 5 Abnormal Returns Around the Majority Threshold

	Vote share					
	±10% (3)	±5% (4)	±2.5% (5)	±1.5% (6)	Full model (7)	Full model with controls (8)
Pass	0.0107**	0.0136**	0.0109***	0.0117	0.0118**	0.0107**
	(0.0046)	(0.0058)	(0.0038)	(0.0074)	(0.0053)	(0.0052)
R-squared	0.056	0.099	0.255	0.204	0.007	0.044
Observations	122	61	23	9	2,729	1,780

Results

Does Corporate Social Responsibility Lead to Superior Financial Performance? A RD Approach



Panel Regression with Fixed Effects

Panel Data and Fixed Effects

- Panel data is double indexed

$$Y_{it}$$

  - i is the entity (or unit): Traders, Firms, Borrowers, ...
  - t is the time period
- Panels track entities over time
- N entities, T time periods
  - N is assumed to be large, T is usually small
  - Asymptotics assume $N \to \infty$

Panel Data and Fixed Effects

- Panel data can be used to estimate pooled OLS models

$$Y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \epsilon_{it}$$

Estimate

  - Ignores the panel structure

- Panel structure allows us to model unobserved heterogeneity

$$Y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \mathbf{W}_i\boldsymbol{\gamma} + \epsilon_{it}$$

Truth

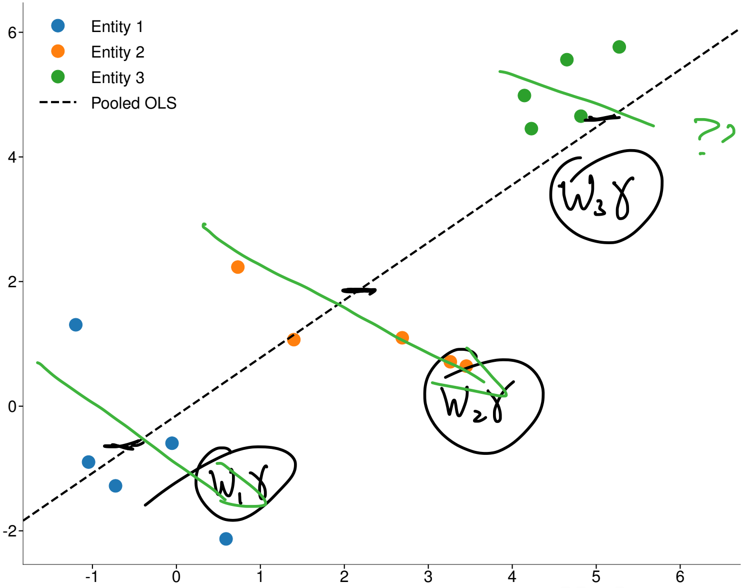
- $\mathbf{W}_i$ is a vector of entity-specific characteristics
- Key: Assumed to be time invariant
- Estimating pooled OLS results in biased coefficients if $\mathbf{W}_i$ is correlated with $\mathbf{X}_{it}$
- In large samples,

$\Lambda \neq 0$

$$\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta} + \boldsymbol{\Lambda\gamma}$$
$$\underline{\mathbf{W}_i = \mathbf{X}_{it}\boldsymbol{\Lambda} + \boldsymbol{\eta}_{it}}$$

  - This is omitted variable bias

Pooled OLS



Entity 1
Entity 2
Entity 3
Pooled OLS
6
4
2
0
-2
-1
0
1
2
3
4
5
6
$W_1\gamma$
$W_2\gamma$
$W_3\gamma$
??

Panel Data and Fixed Effects

- Could collect data on $\mathbf{W}_i$ if available, and include in the model
- In many plausible scenarios it is not observable
  - Individual ability or intrinsic motivation
  - Firm management culture
- Fixed Effect estimator allow $\boldsymbol{\beta}$ to be estimated when $\mathbf{W}_i$ is not known
- Note that $\mathbf{W}_i\boldsymbol{\gamma}$ is a constant for entity i

ω_i

$$Y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \omega_i + \epsilon_{it}$$

$\bar{\omega}_i = \omega_i$

- Demean entity-by-entity

$$Y_{it} - \bar{Y}_i = \left(\mathbf{X}_{it} - \bar{\mathbf{X}}_i\right)\boldsymbol{\beta} + (\omega_i - \bar{\omega}_i) + (\epsilon_{it} - \bar{\epsilon}_i)$$

missing

- ω_i is time-invariant so $\cancel{\omega_i - \bar{\omega}_i = 0}$

$$Y_{it} - \bar{Y}_i = \left(\mathbf{X}_{it} - \bar{\mathbf{X}}_i\right)\boldsymbol{\beta} + (\epsilon_{it} - \bar{\epsilon}_i)$$

- Note that FE models cannot estimate time-invariant effects

Panel Data and Fixed Effects

- Model is equivalent to including entity dummies

$$Y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + D_i\lambda_i + \epsilon_{it}$$

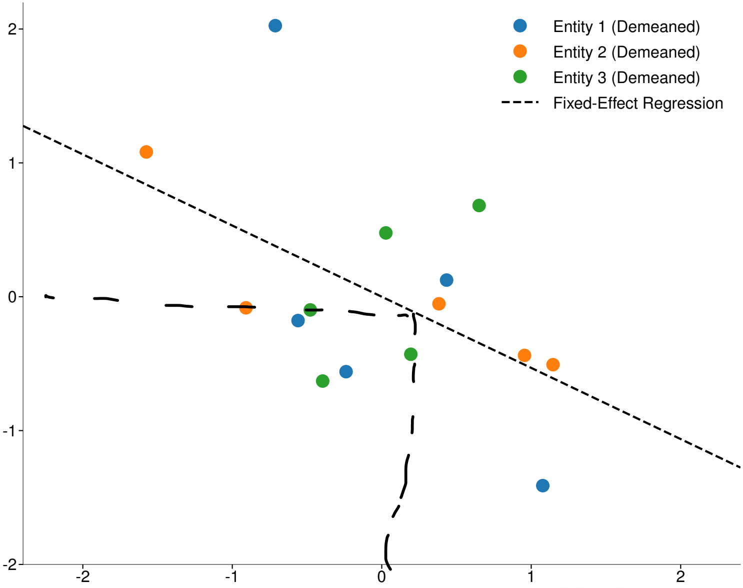
- When $T = 2$ identical to first-difference estimator

$$Y_{i2} - Y_{i1} = (\mathbf{X}_{i2} - \mathbf{X}_{i1})\,\boldsymbol{\beta} + (\epsilon_{i2} - \epsilon_{i1})$$

$\Delta Y_i = \Delta X_i \beta + u_i$

- Inefficient to use first difference estimator for $T \geq 3$

Fixed Effects Regression



Panel Data and Fixed Effects

- Estimates of $\hat{\omega}_i$ are not consistent when T is finite
- Estimated using OLS on entity-wise demeaned data

$$\tilde{Y}_{it} = \tilde{\mathbf{X}}_{it}\boldsymbol{\beta} + \tilde{\epsilon}_{it}$$

  - Known as the Least Squares Dummy Variable (LSDV) estimator
  - Intercept is not meaningful in FE models when reported

Inference in Fixed Effects Models

- Robust inference requires a clustered variance covariance estimator
  - White Covariance

$$\Sigma_{XX}^{-1}\mathbf{S}\Sigma_{XX}^{-1}$$

$$\mathbf{S} = \mathrm{E}\left[\tilde{\epsilon}_{it}^{2}\tilde{\mathbf{X}}_{it}'\tilde{\mathbf{X}}_{it}\right], \ \hat{\mathbf{S}} = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\tilde{\epsilon}}_{it}^{2}\tilde{\mathbf{x}}_{it}'\tilde{\mathbf{x}}_{it}$$

  - Clustered (Rogers) covariance

$$\mathbf{S}_C = \mathrm{E}\left[\xi'\xi\right], \ \hat{\mathbf{S}} = N^{-1}\sum_{i=1}^{N}\hat{\xi}'\hat{\xi}$$

$$\xi = \sum_{t=1}^{T}\tilde{\epsilon}_{it}\tilde{\mathbf{X}}_{it}$$

    - Replace $\mathbf{S}$ when an estimator that allows dependence within entity
    - $\xi'\xi$ contains all squares and cross-products
    - Imposes no restrictions on the dependence *within* an entity

Time Effects

$\hat{\gamma}_t \rightarrow \gamma_t$

- Panels models often include time effects

$$Y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \omega_i + \gamma_t + \epsilon_{it}$$

- γ_t is a shock that affects all observations in period t
  - Commonly used for model common aggregate movement
- Estimated regression use data demeaned using entity and time period
- When N is large and T is small, time effects are consistently estimated
  - Does not need special treatment for inference
  - Identical to including dummies for each time period
- In general, fixed effects can be used to remove constant effects in any dimension with repeated observations
  - Industry
  - City, County, State, or Regional

Example

Real effects of workers' financial distress: Evidence from teacher spillovers

- Examine how test pass rate is affected by *teacher* financial distress

$$y_{cgt} = \beta \times Bankruptcy_{cgt} + \bar{\mathbf{X}}_{cgt}\boldsymbol{\gamma} + \mathbf{Z}_{cgt}\boldsymbol{\lambda} + \underbrace{\delta_{dt} + \eta_{gt} + \phi_{cg}}_{\text{Fixed Effects}} + \epsilon_{cgt}$$

  - Subscripts:
    - c: campus
    - g: grade of student
    - t: year
  - $Bankruptcy_{cgt}$ is *teacher* bankruptcy
  - $\bar{\mathbf{X}}_{cgt}$ are average teacher characteristics
  - $\mathbf{Z}_{cgt}$ are student demographic characteristics

Maturana, G., & Nickerson, J. (2019). Real effects of workers' financial distress: Evidence from teacher spillovers. Journal of Financial Economics.

Example

Real effects of workers' financial distress: Evidence from teacher spillovers

$$y_{cgt} = \beta \times Bankruptcy_{cgt} + \mathbf{X}_{cgt}\boldsymbol{\gamma} + \mathbf{Z}_{cgt}\boldsymbol{\lambda} + \underbrace{\delta_{dt} + \eta_{gt} + \phi_{cg}}_{\text{Fixed Effects}} + \epsilon_{cgt}$$

- Fixed Effects
  - δ_{dt} - District-Year: Control for local economic conditions
  - η_{gt}- Grade-Year: Control for changes to the test across grade and time
  - ϕ_{cg}- Campus-Grade: Control for heterogeneity across different campuses and grades
- Estimates are computed as deviations from all three FE
- Standard errors are clustered by campus-grade and campus-year
  - CG allows arbitrary correlation within all students in a single grade and campus in any year
  - CY allows arbitrary correlation within all students in a single campus and year across all grades