The Distribution of Treatment Effects in Experimental Settings: Applications of Nonlinear Difference-in-Difference Approaches

Susan Athey - Stanford

Guido W. Imbens - UC Berkeley

Introduction

- Based on: "Identification and Inference in Nonlinear Difference-in-Differences Models" with Guido Imbens, forthcoming in *EMA*
- Ideas and Themes for Practice
  - Heterogeneous effects (remain in presence of randomization)
  - Focus on distributions of outcomes
    - Counterfactual dist'ns of what would have happened in absense of treatment, in presence of treatment
    - Use all of the information in the data, e.g. 4 dist'ns
  - Economic v. functional form assumptions
    - Identification v. estimation
  - Verifying assumptions
    - What are crucial assns and what are convenient?
    - Are "standard" checks, e.g. group similarity on obs., most impt.?
- Questions our approach will address:
  - What is distribution of effects of TOT and TOC?
    - How do they differ? Optimal policy adoption?
  - Which individuals are affected?
  - Why quantile regression does not make sense in many apps

Basic Setup: 2×2 **Case**

- Four subpopulations
- "Rows" and "Columns": Groups and time periods

	Group A	Group B
Time 2		Treated
Time 1		

Examples: Overview

- State Policy Change
  - 2 states, 2 time pds, law changes in one state in 2nd pd
  - Groups are different
  - Worried about time trend
  - Ex: Minimum wage (Card-Krueger), Disability benefits (Meyer-Viscusi-Durbin)
- Experiment with Heterogeneous Subjects
  - 2 groups, 2 treatments, 2nd treatment has differential impact
  - Groups are different (and heterogeneity w/i group)
  - Interested in group difference relative to difference in baseline treatment
  - Examples:
    - Lab experiment: men and women, 1st treatment measures ability, 2nd treatment tournament
    - Auction identical items on eBay and Amazon, vary auction conditions
- Dynamics and Learning in the Lab
  - Control and treatment group, dynamic game or learning
  - Use initial play to learn about intrinsic differences in two games
  - See how differences evolve over time

- Hospital Technology Adoption Field Experiment
  - Two hospitals, two time periods, one adopts surgical tech in 2nd pd
  - Hospitals are different
  - Patients randomly assigned in each pd
  - Pre-surgical technology changes over time
- Effects of Training/Experience on Ability Tests
  - Subjects take a test to measure ability
  - Randomly assigned to different tests
  - Randomly assigned to training (or to practice test)
  - Does training/experience have different distn of benefits for different tests?

Formal Model

- 2×2 case
- N observations on (Y, T, G)
- Y is outcome
- $T \in \{1, 2\}$ is time period,
- $G \in \{A, B\}$ is group.
- Group $G = B$ in period $T = 2$ is the only group/period exposed to intervention.
  - Y^I is outcome if treated; observed if $G = B$, $T = 2$
  - Y^N is outcome if untreated; observed unless $G = B$, $T = 2$
- Let Y_{gt} be r.v. with dist'n of $Y|G = g, T = t$.

Standard DID Model

- Model for outcome in absence of intervention:

$$Y^N = \alpha + \beta \cdot T + \eta \cdot G + \varepsilon,$$

  with

$$\varepsilon \perp (T, G).$$

  – Weaker assn: mean-indep

- Average outcome for subpop $(B, 2)$ in absence of intervention:

$$\begin{aligned} \mathbb{E}[Y^N_{B2}] &= \alpha + \beta + \eta \\ &= \mathbb{E}[Y_{B1}] + [\mathbb{E}[Y_{A2}] - \mathbb{E}[Y_{A1}]] \end{aligned}$$

- Average treatment effect on treated group:

$$\begin{aligned} \tau^{DID} &= \mathbb{E}[Y^I_{B2}] - \mathbb{E}[Y^N_{B2}] \\ &= \mathbb{E}[Y_{B2}] - \mathbb{E}[Y_{B1}] - [\mathbb{E}[Y_{A2}] - \mathbb{E}[Y_{A1}]] \end{aligned}$$

  – Note: no model required for Y^I_{B2}

Fitting Example Into Standard DID Model: State-level Policy Adoption

- States choose whether to adopt health program
  - State A has more sick people than State B
  - Program more effective on sicker patients
    - State B adopts in response to voter demand & cost-benefit analysis
  - Hospital outcomes observed
  - Medical treatments change each year
    - Benefit sickest patients most
    - Expect average time trend to differ across counties in absence of program
- Standard model doesn't apply
- Questions our approach will address:
  - What was benefit to State B?
  - What would the benefit be to State A?
  - Which patients helped most?

Summary of Problems with Standard DID Model

We will address:

- Linearity/additivity
  - Rules out interesting economics
    - Cannot have effect of unobservable change over time, mean-variance shift over time
  - Assumption not invariant to transformation of dep. variable
- Get effect of the treatment on the treated, not control
  - Leads to constant treatment effect assn's or assn of "exogenous" policy adoption
- Problems with binary/discrete outcomes
  - Linear model can predict outcomes out of bounds

We will ignore:

- Treatment affects group comp. (Heckman 1996; Marrufo 2001)
- Issues with standard errors (Donald and Lang, 2001; Bertrand, Duflo, and Mullainathan, 2001)

Our Baseline Model: "Changes-in-Changes"

A0 (Model) $\quad Y^N = h(U, T).$

- h is nonlinear, unknown function
  - "Production function" does not vary with group
  - All diff's across groups due to dist'n of U

A1 (Time Invariance) $\quad U \perp T \mid G.$

- Group composition does not change over time
  - Ex: state or county, short time periods

A2 (Monotonicity) $\quad h(u, t)$ is strictly montone in u.

- Not a restriction in a single period
  - Restrictive in conjunction with (A1)
  - Enables inference of change in prod fn over time

- Std model has two add'l restrictions:

$$U = \eta \cdot G + \varepsilon$$

$$h(U, T) = \alpha + \beta \cdot T + U$$

  - Note: No assn's (yet) on effect of treatment. Focus first on treatment on treated.

Interpretations

- Fits application where:
  - Group same over time–e.g. state or county, short time periods
  - Similar services, etc. except for policy
  - Some group differences in "technology" can be incorporated in U
- Compare across groups and over time (not additivity)
  - Look across groups at time t :
    - What is comparable: Production fn, thus level of outcome
    - What is different: Distribution of individuals
  - Look over time within group g :
    - What is comparable: Quantile of outcome (i.e. u)
    - What is different: "Production function"
- Assumptions not symmetric for group and time

Fitting Example into CIC Model: State Policy Change

- Economic assns
  - States differ in distn of unobs factors affecting health U : policy assignment NOT "random"
  - (A1) Groups stable over time: distn of U does not vary w/i group
  - Mapping from U to outcomes, $h^N(U,t)$ and $h^I(U,t)$ are (A2) monotone and depend on (i) time (t) and (ii) treatment status, but (A0) not directly on group
- Interpreting (A2) and (A3):
  - Pd. 1: health outcome $h^N(U,1) \equiv U$
    - Incorporates some differences in infrastructure, etc.
  - Pd. 2: $h^N(U,2) \neq U$ due to changes in health technology that apply to both treatment and control group and do not change ranking of outcomes
  - w/o treatment, mapping from unobs to outcomes same in both states due to similar technology changes

Identification of (Continuous) CIC Model

Theorem 1 *Assume:*
(i) CIC model: A0-A2,
(ii) supp$[Y_{B1}] \subseteq$*supp*$[Y_{A1}]$.
Then the distribution of Y_{B2}^N *is identified and given by*

$$F_{Y^N,B2}(y) = F_{Y,B1}(F_{Y,A1}^{-1}(F_{Y,A2}(y))).$$

See paper for nonparametric estimation, CAN, efficiency.

"Proof"

- Pick a first period treated unit with outcome y.
- Find someone with the same outcome y in the first period control group. By the model these units must have the same unobs. u.
- Find the rank of this unit in the $(A, 1)$ distribution, $F_{Y,A1}(y)$.
- By monotonicity, a control person with the same value of u in period 2 would have outcome

$$F_{Y,A2}^{-1}(F_{Y,A1}(y)) = h(h^{-1}(y;1),2).$$

- Apply this transformation to outcome in first period treatment unit, so that

$$\begin{aligned}\Pr(Y_{B2}^N \leq y) &= F_{Y,B1}\left(h(h^{-1}(y;2),1)\right) = \Pr(F_{Y,A2}^{-1}(F_{Y,A1}(Y_{B1})) \leq y) \\ &= F_{Y,B1}(F_{Y,A1}^{-1}(F_{Y,A2}(y))).\end{aligned}$$

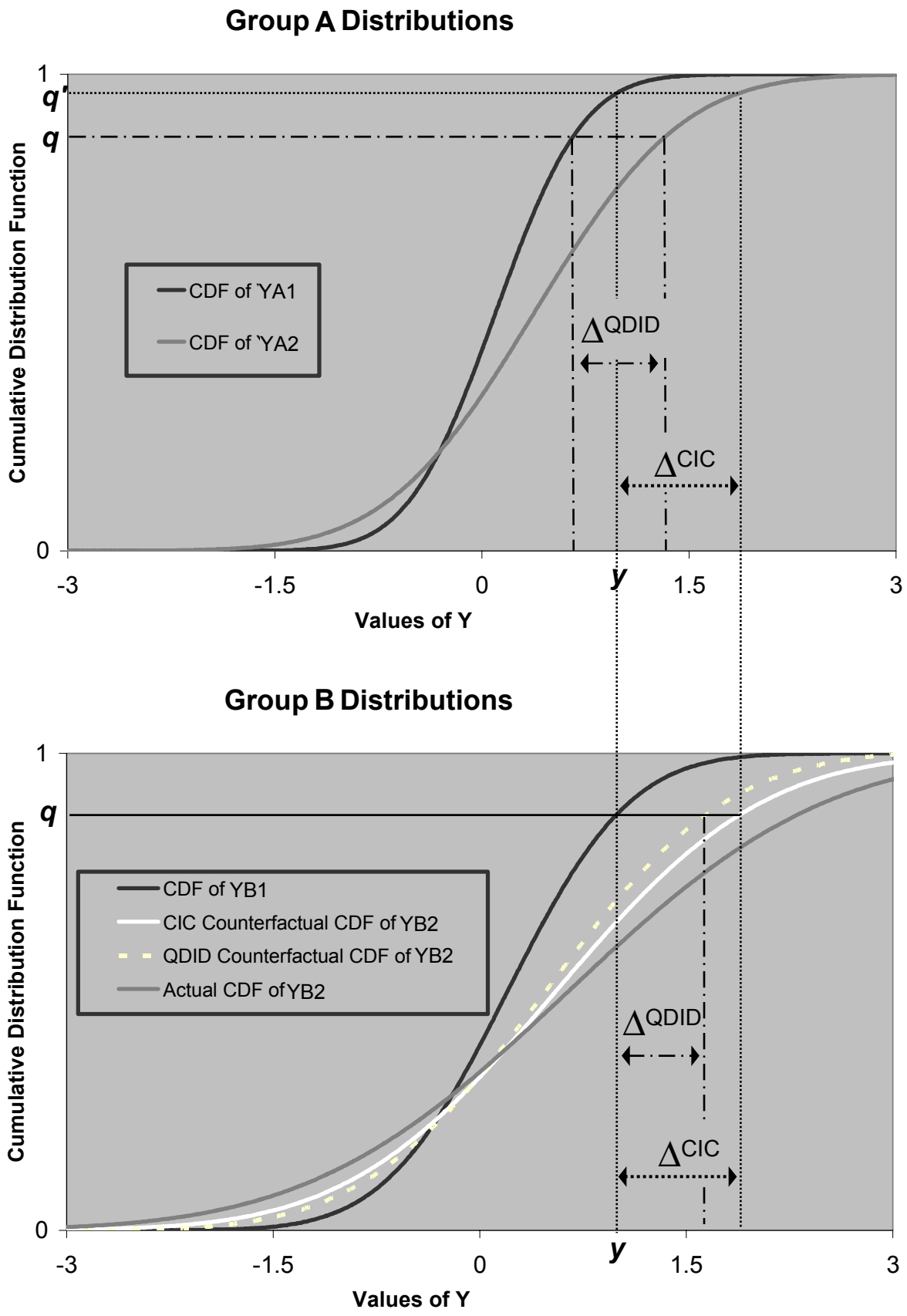

Figure 1: Illustration of Transformations

Interpretation in Terms of Transformation

- Model defines a transformation,

$$k^{CIC}(y) = F^{-1}_{Y,A2}(F_{Y,A1}(y)),$$

  such that

$$Y^N_{B2} \overset{d}{\sim} k^{CIC}(Y_{B1})$$

- Standard DID model has simple linear transformation:

$$k^{DID}(y) = y + \mathbb{E}[Y_{A2}] - \mathbb{E}[Y_{A1}].$$

The Effect of the Treatment on the Control Group

- Analogous model assumption:

$$Y^I = h^I(U, T)$$

- Goal: Compute distribution of Y^I_{A2}.
  - Problem seems diff't: only 1/4 subpop's treated
  - But under assn's, exactly analogous.
- Apply transformation to the time 1, group 1 outcomes:

$$F^{-1}_{Y^I,B2}(F_{Y,B1}(y)) = h^I(h^{-1}(y;1),2)$$

  so that

$$Y^I_{A2} \overset{d}{\sim} F^{-1}_{Y^I,B2}(F_{Y,B1}(Y_{A1}))$$

- End result
  - Same procedure
  - Reverse roles of treatment and control group

Fitting Example into CIC Model: Experiment with Heterogeneous Groups

- See Gneezy, Niederle and Rustichini (2003)
- Men and women differ in ability to perform baseline task, define $h^N(U,1) = U$
- (A1) Groups stable over time: distn of U does not vary w/i group
  - Same individuals or different cross-sections
- Instead of "treated" and "untreated," there is a male and female production function for "treatment" task
  - Mapping from U to outcomes, $h^{male}(U,2)$ and $h^{female}(U,2)$ are (A2) monotone
  - Treatment task doesn't cause low-ability individuals to pass high-ability individuals
- Question: what would distn of female performance be, given underlying ability, if they had male production function for treatment task, and vice-vera
  - Distn of $h^{male}(U,2)|female$ and $h^{female}(U,2)|male$

Fitting Example into CIC Model: Field Experiments Auctions on eBay and Amazon

- Auction identical objects in different formats
- Different sets of bidders at two sites, stable over time, so auction price is draw from different distn
- Baseline differences, incorporating different bidders and site differences: price=$h^N(U, 1) = U$
  - All baseline differences accounted for by distn of U varying across sites
- Compare relative differences from varying auction design parameters
  - Reserve price, buy-it-now, experience rating of seller, etc.

The CIC-r Model: Reverse the Roles of Group and Time Period

- Recall that we make different assumptions for groups and periods
- Reverse:
  - Two groups have identical distributions of unobservables within a time period (e.g. random assignment)
  - Production function stays the same over time in absence of the treatment
- Result: TOT is identified, apply same formula with roles reversed
- TOC not identified w/o extra assumptions
  - Issue: given that we think groups have different production functions, not clear what data tells us about production function of control group in presence of treatment.

Example (2): Hospital Technology Adoption Field Experiment

- Setup
  - Groups: hospitals Time: before and after CATH
  - Group B in period 2 is "treated"
  - Patients randomly assigned to hospitals in each period
- Economic assns (imply distn of TOT identified, *not* TOC)
  - (A1′) Within a period, distn of U same for both hospitals
  - Distn of individuals changes over time (new heart drugs available to patients of both hospitals)
  - Hospitals have different production functions
  - Mapping from U to outcomes in absence of treatment, $h^N(U, g)$ is (A2) monotone and depends on hospital (g) but (A0′) not directly on time
- Ideas
  - Control hospital tells us how distn of patients changed over time, use to calculate counterfactual outcomes if hospital production function had not changed
  - Issue: given that we think hospitals have different production functions, not clear what data tells us about production function of control hospital in presence of treatment.

The QDID Model: Quantile Regression at Each Quantile

- Assumptions
  - Distribution of unobservables same for each subpopulation
    - Random assignment to variations on treatments
    - Otherwise, why compare quantiles?
  - Production function monotone in unobservables
    - Ranking of outcomes same for all treatment variations
  - Group effect and time effect are additive

$$Y = h^G(U, g) + h^T(U, t) + h^I(U)$$

    - Imposes testable restrictions on the data
    - Implies average TOT and TOC are the same
- Gives different answer than CIC or CIC-r models
- May be applicable in experimental contexts
- Our value-added
  - Unified uderlying structural model motivating regression at all quantiles
  - Test validity
  - Generate entire distribution of counterfactual outcomes

Conclusions

- Approach
  - Essence of DID: Control group provides information about what would have happened to treatment group in the absence of the treatment
  - Data are four distributions
  - Use three distributions to predict counterfactual for fourth
    - Economic assumptions tell you how
  - Nonparametric structural model, primitives are production function and distribution of unobservables
  - Focus on distributions of effects
- Benefits of Approach
  - New model relaxes assumptions of standard DID, can be more or less efficient
  - Scale invariance
  - No problems with out-of-bounds predictions for discrete variables
  - Treatment can be endogenous to both level of outcome in each group and anticipated incremental benefit of policy
  - Identify effect of treatment on treated and control groups, can compute structural parameters
  - Fewer, but still many, settings where CIC is not applicable
- See paper for discrete outcomes, multiple groups, treatments