14.387 Recitation 3

LATEs, Differences, and Changes

Peter Hull

Fall 2014

Part 1: Some More LATE

Basic LATE Setup

- Potential outcomes/treatment: $Y_i(d,z)$, D_{1i}, D_{0i}, binary Z_i
- Four assumptions
  1. *Independence:* $(\{Y_i(d,z); \forall d,z\}, D_{1i}, D_{0i}) \perp\!\!\!\perp Z_i$
  2. *Exclusion:* $Y_i(1,z) = Y_{1i}$, $Y_i(0,z) = Y_{0i}$, $\forall z$, $\forall i$
  3. *Monotonicity:* $D_{1i} \geq D_{0i}$, $\forall i$
  4. *First stage:* $E[D_{1i} > D_{0i}] \neq 0$
- Assumptions 1-4 in words?
- Linking observed to potentials:

$$D_i = D_{0i} + (D_{1i} - D_{0i})Z_i$$
$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i$$

The LATE Theorem

Prop: Under Assumptions 1-4,

$$\frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{E[D_i|Z_i=1]-E[D_i|Z_i=0]}=E[Y_{1i}-Y_{0i}|D_{1i}>D_{0i}]$$

Direct proof:

$$\begin{aligned}E[D_i|Z_i=1]-E[D_i|Z_i=0] =& E[D_{1i}|Z_i=1]-E[D_{0i}|Z_i=0]\\ =& E[D_{1i}-D_{0i}]\\ =& P(D_{1i}>D_{0i})\end{aligned}$$

and

$$\begin{aligned}E[Y_i|Z_i=1]-E[Y_i|Z_i=0] =& E[Y_{0i}+(Y_{1i}-Y_{0i})D_{1i}|Z_i=1]\\ & -E[Y_{0i}+(Y_{1i}-Y_{0i})D_{0i}|Z_i=0]\\ =& E[(Y_{1i}-Y_{0i})D_{1i}-(Y_{1i}-Y_{0i})D_{0i}]\\ =& E[(Y_{1i}-Y_{0i})(D_{1i}-D_{0i})]\\ =& E[Y_{1i}-Y_{0i}|D_{1i}>D_{0i}]P(D_{1i}>D_{0i})\end{aligned}$$

LATE with Covariates

- In class we saw Abadie's (2003) weighting approach to identifying LATE when Assumptions 1-4 hold conditional on covariates X_i

  1. Estimate (perhaps non-parametrically) $E[Z_i|X_i]$ (first step)
  2. Form kappa:

  $$\kappa(D_i, Z_i, X_i) = 1 - \frac{D_i(1-Z_i)}{1-E[Z_i|X_i]} - \frac{(1-D_i)Z_i}{E[Z_i|X_i]}$$

  3. Estimate by WLS

  $$(\alpha, \beta) = \arg\min_{a,b} E[\kappa(D_i, Z_i, X_i)(Y_i - a - bD_i)^2]$$

- Requires correcting WLS standard errors for first-step estimation, parametric choice of $E[Z_i|X_i]$; predictions may be outside $(0, 1)$

- Can we handle covariates just with 2SLS?

Conditional LATE with 2SLS: "Saturate and Weight"

- Suppose we can saturate in X_i (e.g. stratified RCT with two-sided compliance)
- At each x in the support can identify

$$\frac{E[Y_i|Z_i=1,X_i=x]-E[Y_i|Z_i=0,X_i=x]}{E[D_i|Z_i=1,X_i=x]-E[D_i|Z_i=0,X_i=x]} \equiv \beta_{LATE}(x)$$

- MHE advice: "saturate and weight"

$$Y_i = \alpha_X + \beta_{SW} D_i + \varepsilon_i$$
$$D_i = \gamma_X + \pi_X Z_i + \nu_i$$

i.e. we interact the instrument with every cell of X_i (over-id)

- What does 2SLS of a fully-saturated model identify?

"Saturate and Weight" (cont.)

- Recall (e.g. PS#1) that β_{SW} is identified by OLS of

$$Y_i = \alpha_X + \beta_{SW}\hat{D}_i + \varepsilon_i$$

where $\hat{D}_i$ is the first-stage predicted value of D_i.

- Recall (also PS#1) that, since this is OLS saturated in X_i,

$$\beta_{SW} = \frac{E[\beta_{SW}(X_i)\sigma^2_{\hat{D}}(X_i)]}{E[\sigma^2_{\hat{D}}(X_i)]}$$

- Since the first stage is also saturated,

$$\begin{aligned} \beta_{SW}(x) &= \beta_{LATE}(x) \\ \sigma^2_{\hat{D}}(x) &= Var(\hat{D}_i|X_i = x) \\ &= \pi_x^2 Var(Z_i|X_i = x) \\ &= P(D_{1i} > D_{0i}|X_i = x)^2 \sigma^2_Z(x) \end{aligned}$$

Pros and Cons of "Saturate and Weight"

$$\beta_{SW} = \frac{E[\beta_{LATE}(X_i)P(D_{1i} > D_{0i}|X_i)^2\sigma_Z^2(X_i)]}{E[P(D_{1i} > D_{0i}|X_i)^2\sigma_Z^2(X_i)]}$$

- Identify a convex combination of covariate-specific LATEs
- Get correct (2SLS) standard errors for free, but
  - Why these weights? (*square* of complier share?)
  - Weighting over full histogram of X_i (not complier histogram)
  - Specification could be heavily over-id'd; might lead to bias
- Is there a better way?

Partially-Linear IV (New!)

- Suppose we instead run the just-identified IV model

$$\begin{aligned} Y_i =& \theta_X + \beta_{PL} D_i + \varepsilon_i \\ D_i =& \delta_X + \pi Z_i + \nu_i \end{aligned}$$

First/second stage saturated in X_i but linear in Z_i/D_i

- From Abadie (2003) we know

$$\begin{aligned} (\theta_X, \beta_{PL}) &= \arg\min_{b,t_X} E[\kappa(D_i, Z_i, X_i)(Y_i - t_X - bD_i)^2] \\ &= \arg\min_{b,t_X} E[(Y_i - t_X - bD_i)^2 | D_{1i} > D_{0i}] \end{aligned}$$

since $E[Z_i|X_i]$ is fit perfectly when X_i saturates (Prop. 5.1 in paper)

Partially-Linear IV (cont.)

$$(\theta_X, \beta_{PL}) = \arg\min_{b,t_X} E[(Y_i - t_X - bD_i)^2 | D_{1i} > D_{0i}]$$

Again by PS#1/Angrist '98 logic

$$\beta_{PL} = \frac{E[\beta_{PL}(X_i)\sigma^2_{D_i,C}(X_i)|D_{1i} > D_{0i}]}{E[\sigma^2_{D_i,C}(X_i)|D_{1i} > D_{0i}]}$$

Now:

$$\begin{aligned} \beta_{PL}(x) &= \beta_{LATE}(x) \\ \sigma^2_{D,C}(x) &= Var(D_i|X_i = x, D_{1i} > D_{0i}) \\ &= Var(Z_i|X_i = x, D_{1i} > D_{0i}) \\ &= \sigma^2_Z(x) \end{aligned}$$

Pros of Partially-Linear IV

$$\beta_{PL} = \frac{E[\beta_{LATE}(X_i)\sigma_Z^2(X_i)|D_{1i} > D_{0i}]}{E[\sigma_Z^2(X_i)|D_{1i} > D_{0i}]}$$

- Also identify a convex comb. of $\beta_{LATE}(X_i)$ with a partially-linear first stage

- Get correct standard errors for free
- Weights more intuitive (like usual FE weights)
- Weighting by *complier* histogram of X_i
- Potential efficiency argument (not yet worked out)

- Bottom line: can easily handle saturating covs with IV
- When you can't saturate, do kappa

Part 2: Differences and Changes

Diff-in-diffs and Fixed Effects Regression

- Model:

$$Y_{it} = \alpha_i + \gamma_t + \rho D_{it} + X_{it}'\beta + \varepsilon_{it}$$

where $i = 1, ..., N$ are potentially treated states, $t = 1, ..., T$ is time, d_{it} is an indicator for treatment in state i at time t, and x_{it} are controls

- Can get rid of state effects by first-differencing the data

$$\Delta Y_{it} = \Delta\gamma_t + \rho \Delta D_{it} + \Delta X_{it}'\beta + \Delta\varepsilon_{it}$$

This is how Card (1992) does Diff-in-Diffs

- Alternatively can demean the data within states

$$\tilde{Y}_{it} = \tilde{\gamma}_t + \rho \tilde{D}_{it} + \tilde{X}_{it}'\beta + \tilde{\varepsilon}_{it}$$

This is what a fixed-effects regression does

Diff-in-diffs and FEs (cont.)

- If the model is correct and all variables are measured properly, both methods are consistent for ρ
  - Motivates measurement error tests (Griliches and Hausman, 1986)
- With only two periods, the methods are *numerically equivalent*
- Proof of this: write

$$
\begin{aligned}
Y_{it} &= \alpha_i + W_{it}'\mu + \varepsilon_{it} \\
\tilde{Y}_{it} &\equiv Y_{it} - \frac{Y_{i1} + Y_{i2}}{2} \\
\Delta Y_i &\equiv Y_{i2} - Y_{i1}
\end{aligned}
$$

Diffs=FEs when T=2

Note that

$$\begin{aligned}\tilde{Y}_{i1} &= Y_{i1} - \frac{Y_{i1} + Y_{i2}}{2} \\ &= \frac{1}{2}(Y_{i1} - Y_{i2}) \\ &= -\frac{1}{2}\Delta Y_i\end{aligned}$$

Similarly

$$\begin{aligned}\tilde{Y}_{i2} &= Y_{i2} - \frac{Y_{1i} + Y_{i2}}{2} \\ &= \frac{1}{2}\Delta Y_i\end{aligned}$$

Diffs=FEs when T=2 (cont.)

Thus

$$\begin{aligned}
\hat{\mu}_{FE} &= \left[\sum_{i=1}^{N} \tilde{W}_{i1}\tilde{W}_{i1}' + \tilde{W}_{i2}\tilde{W}_{i2}'\right]^{-1} \left[\sum_{i=1}^{N} \tilde{W}_{i1}\tilde{Y}_{i1} + \tilde{W}_{i2}\tilde{Y}_{i2}\right] \\
&= \left[\sum_{i=1}^{N} \frac{1}{4}\Delta W_i \Delta W_i' + \frac{1}{4}\Delta W_i \Delta W_i'\right]^{-1} \left[\sum_{i=1}^{N} \frac{1}{4}\Delta W_i \Delta Y_i + \frac{1}{4}\Delta W_i \Delta Y_i\right] \\
&= \left[\sum_{i=1}^{N} \Delta W_i \Delta W_i'\right]^{-1} \left[\sum_{i=1}^{N} \Delta W_i \Delta Y_i\right] \\
&= \hat{\mu}_{FD}
\end{aligned}$$

$\square$

Change-in-Changes

- Recall an inherent drawback of difference-in-differences: taking functional form seriously
    - Can't (typically) have parallel trends in both levels and logs
- Athey and Imbens (2006) consider a semiparametric framework to overcome this using estimated distributions of potential outcomes
    - AI'09 assumptions are strong (unlike AI'94): no D-in-D free lunch!
- Basic idea: treatment preserves rank in the outcome distribution; impute the counterfactual change in outcomes for treated using the change in control distribution

C-in-C Setup

- Start with a weak assumption: untreated outcomes satisfy

$$Y_i^0 = h(U_i, T_i)$$

  where $T_i \in \{0,1\}$ denotes time and U_i is unobserved heterogeneity

- Nests canonical D-in-D setup:

$$\begin{aligned} U_i &= \alpha + \gamma G_i + \varepsilon_i \\ h(u,t) &= u + \delta \cdot t \\ \varepsilon_i &\perp\!\!\!\perp (G_i, T_i) \end{aligned}$$

  where $G_i \in \{0,1\}$ denotes groups. With treatment $I_i \equiv G_i \cdot T_i$ and constant effects,

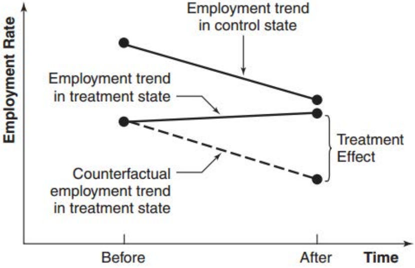
$$\begin{aligned} Y_i &= Y_i^0 + \rho I_i \\ &= \alpha + \gamma G_i + \delta T_i + \rho G_i \cdot T_i + \varepsilon_i \end{aligned}$$

C-in-C Assumptions

Assumption 1 (Strict monotonicity): $h(u,t)$ is strictly increasing in u

Assumption 2 (Time invariance): $U_i \perp\!\!\!\perp T_i | G_i$

- Imply rank-preservation: if my U_i puts me at the q-th quantile of the distribution of Y_i^0 at $T_i = 0$, it also puts me there at $T_i = 1$
  - Very strong: like Chernozhukov and Hansen (2005) for quantile IV
- Allow us to impute the change in non-treated outcomes for treated in $T_i = 1$; conceptually very similar to D-in-D:



C-in-C Identification

Proposition: Under **A**1 & **A**2 (and a few other technical conditions)

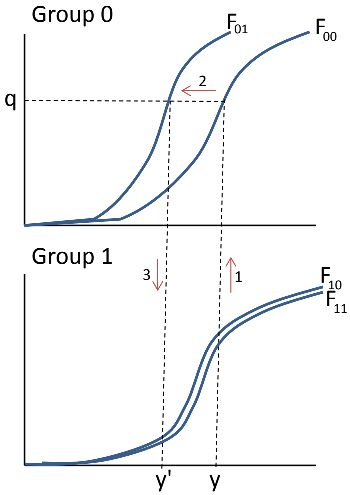
$$E[Y_{11}^1 - Y_{00}^0] = \underbrace{E[Y_{11}^1]}_{\text{Observed}} - \underbrace{E[F_{Y,01}^{-1}(F_{Y,00}(Y_{10}))]}_{\text{Counterfactual}}$$

For every value y in the pre-treatment distribution of the treated (F_{10}):

- Find quantile $q = F_{Y,00}(y)$
- Observe among controls that quantile q changed to $y' = F_{Y,01}^{-1}(q)$

Average these new y' according to the distribution of y

C-in-C Mechanics



- Inference on C-in-C estimator: not a walk in the park (see paper)

14.387 Applied Econometrics: Mostly Harmless Big Data
Fall 2014