Economics 696F: Lecture Note 13 (corrected 3/30/07)

Differences-in-Differences

Example: Card (1990) Study of the Mariel Boatlift

Q: Do low-skilled immigrants displace low-skilled US citizens in the labor market?

Intervention: Mariel Boatlift, a large-scale migration of Cubans to Miami, which increased the local labor force by about 7% in 1980.

Unemployment rate in Miami:

Whites: 5.1 in 1979, 3.9 in 1981.

Blacks: 8.3 in 1979, 9.6 in 1981.

Y_i: indicator for unemployed.

$Y_i(0)$: unemployment without intervention

$Y_i(1)$: unemployment with intervention

Estimand:

$$\tau := E[Y_i(1)|Miami, 1981] - E[Y_i(0)|Miami, 1981].$$

Simple difference in means estimates:

$$\begin{aligned} E[Y_i|Miami, 1981] - E[Y_i|Miami, 1979] &= E[Y_i(1)|Miami, 1981] - E[Y_i(0)|Miami, 1979] \\ &\neq E[Y_i(1)|Miami, 1981] - E[Y_i(0)|Miami, 1981] \end{aligned}$$

since unemployment rate may change over time due to other factors (macro shocks, etc).

Solution: find comparison cities that did not experience the intervention, but plausibly experienced similar "other" labor market shocks.

Cities: Atlanta, Los Angeles, Houston, Tampa-St.Petersburg.

Notation:

Group: $G_i = 0, 1$ (1 for Miami, 0 for comparison cities)

Time period: $T_i = 0, 1$ (1 for 1981, 0 for 1979)

Treatment/Intervention:

$$I_i := 1(G_i = 1, T_i = 1) = G_i \cdot T_i.$$

Standard diffs-in-diffs model:

$$Y_i(0) = \alpha + \beta T_i + \eta G_i + \epsilon_i$$

$$Y_i(1) = Y_i(0) + \tau.$$

$$\Rightarrow \quad Y_i = \alpha + \beta T_i + \eta G_i + \tau I_i + \epsilon_i$$

Can estimate using OLS.

Equivalently:

$$\hat{\tau} = E[Y_i|G_i \widehat{= 1,} T_i = 1] - E[Y_i|G_i \widehat{= 0,} T_i = 1] - \Big\{ E[Y_i|G_i \widehat{= 1,} T_i = 0] - E[Y_i|G_i \widehat{= 0,} T_i = 0] \Big\}$$

Results:
Whites: $\hat{\tau} = -1.1(1.5)$
Blacks: $\hat{\tau} = -1.0(2.8)$

Can also extend to allow conditioning on X_i:

$$Y_i = \alpha + \gamma' X_i + \beta T_i + \eta G_i + \tau I_i + \epsilon_i$$

Drawbacks: strong functional form assumptions, esp. additivity, which depend on the scaling of outcomes. Hard to motivate functional form from economic theory.

Athey-Imbens Model:

for simplicity, consider the case where the outcome is continuous, with compact support.

Assumptions: (drop i subscript and regard observations as random variables)

1. $Y(0) = h(U, T)$.
2. $h(u, t)$ is strictly increasing in u.
3. $U \perp T | G$.
4. $supp[U|G = 1] \subset supp[U|G = 0]$.

Here *supp* means support of the random variable.

Notation:

$$\begin{aligned} Y_{gt}(0) &:= Y(0)|G=g,T=t \\ Y_{gt}(1) &:= Y(1)|G=g,T=t \\ Y_{gt} &:= Y|G=g,T=t \\ U_g &:= U|G=g \end{aligned}$$

Recall $I := G \cdot T$.

Let $F_{Y(0),gt}$ denote CDF of $Y(0)|G=g,T=t$, etc.

So the data identifies $F_{Y,gt}$ for all g,t.

Therefore

$$F_{Y(1),11} = F_{Y,11}$$

can be estimated from the data.

However, we cannot directly estimate $F_{Y(0),11}$.

Treatment effect of interest is

$$\begin{aligned} \tau &= E[Y(1)|G=1,T=1] - E[Y(0)|G=1,T=1] \\ &= \int y dF_{Y(1),11}(y) - \int y dF_{Y(0),11}(y) \\ &= \int y f_{Y(1),11}(y)dy - \int y f_{Y(0),11}(y)dy. \end{aligned}$$

Here $f(y) = dF(y)/dy$ is the density function associated with the CDF $F(y)$.

Key Theorem: Under assumptions 1-4,

$$F_{Y(0),11}(y) = F_{Y,10}\left(F^{-1}_{Y,00}(F_{Y,01}(y))\right).$$

Since each function on the right hand side is estimable, we have identified $F_{Y(0),11}$.

"Intuitive" Sketch of Proof:

For a given value of y, a CDF $F(y)$ gives the 'ranking' on a percentile scale (like SAT score percentiles). The inverse CDF, $F^{-1}(u)$ (for $u \in [0,1]$) gives the outcome associated with that percentile ranking.

Consider

$$F_{Y,01}(y) = Pr(Y(0) \leq y|G=0,T=1).$$

This is the percentile ranking of a person with outcome value y in the $G=0,T=1$ set.

Since $Y(0) = h(u, t)$ is stricly increasing in u, and since $U \perp T|G$, this person would have the same ranking if moved to the $G = 0, T = 0$ group. So her outcome would be

$$F^{-1}_{Y,00}(F_{Y,01}(y)).$$

In other words, we have transformed an outcome in period 1 to a corresponding outcome in period 0 for the control group. Define the inverse of this transformation:

$$k(y) = F^{-1}_{Y,01}(F_{Y,00}(y)).$$

This transforms the distribution of outcomes from period 0 to period 1.

Now apply this transformation to the $G = 1$ group:

$$\begin{aligned} F_{Y(0),11}(y) &= Pr(Y(0) \leq y|G = 1, T = 1) \\ &= Pr(k(Y(0)) \leq y|G = 1, T = 0) \\ &= Pr(Y(0) \leq k^{-1}(y)|G = 1, T = 0) \\ &= F_{Y(0),10}(k^{-1}(y)) \\ &= F_{Y,10}\left(F^{-1}_{Y,00}(F_{Y,01}(y))\right). \end{aligned}$$

Estimation

Since

$$E[Y(1)|G = 1, T = 1] = E[Y|G = 1, T = 1]$$

The first component of τ can be estimated by simply taking the sample average of Y in the $G = 1, T = 1$ set:

$$E[Y(1)|\widehat{G = 1}, T = 1] := \frac{\sum_{i=1}^{n} Y_i \cdot 1(G_i = 1, T_i = 1)}{\sum_{i=1}^{n} 1(G_i = 1, T_i = 1)}.$$

The second term in τ is $E[Y(0)|G = 1, T = 1]$.

Recall that the transformation $k(y) = F^{-1}_{Y,01}(F_{Y,00}(y))$ transforms outcomes from period 0 to period 1. Thus we can apply it to $Y = Y(0)$ outcomes in the $G = 1, T = 0$ set to get counterfactual $Y(0)$ outcomes for $G = 1, T = 1$. This suggests:

$$E[Y(0)|\widehat{G = 1}, T = 1] := \frac{\sum_{i=1}^{n} \hat{F}^{-1}_{Y,01}(\hat{F}_{Y,00}(Y_i)) \cdot 1(G_i = 1, T_i = 0)}{\sum_{i=1}^{n} 1(G_i = 1, T_i = 0)}$$

where $\hat{F}_{Y,01}$ is the estimated CDF of Y given $G = 0, T = 1$ and $\hat{F}_{Y,00}$ is the estimated CDF of Y given $G = 0, T = 0$.

Our estimator is then

$$\hat{\tau} = E[Y(1)|\widehat{G=1}, T=1] - E[Y(0)|\widehat{G=1}, T=1]$$

Side Note: Estimating CDFs

Suppose $Y_1, \ldots, Y_n$ are IID with CDF $F(y)$. Then since

$$F(y) = Pr(Y \leq y) = E[1(Y \leq y)],$$

a natural estimator is the sample analog:

$$\hat{F}(y) := \frac{1}{n} \sum_{i=1}^{n} 1(Y_i \leq y).$$

This is called the empirical CDF, because it is the CDF of the discrete distribution that places probability $1/n$ on each observed value of Y_i.

Note that to get the whole function, you need to estimate this for each value y in the support of Y. The empirical CDF is a step function, with jumps of height $1/n$ at each observed value of Y_i.

There are two ways to view the empirical CDF estimator. One is to fix a single value of y, and consider $\hat{F}(y)$ as a (scalar) estimator for $F(y)$. The other way is to view $\hat{F}(\cdot)$ as an estimated *function.*

Consider the first view, where we fix a single y. Since Y_i is an IID random variable, so is $1(Y_i \leq y)$. It is a Bernoulli variable, equal to 1 with probability $Pr(Y_i \leq y) = F(y)$. It has variance $F(y)(1 - F(y))$.

Therefore, by the simple law of large numbers,

$$\hat{F}(y) \xrightarrow{p} F(y),$$

and by the central limit theorem

$$\sqrt{n}(\hat{F}(y) - F(y)) \xrightarrow{d} N(0, F(y)(1 - F(y))).$$

The second view requires more care. (Note: you do not need to know this, but I am including it to give a flavor of some of the underlying theory needed to work rigorously with empirical CDFs.)

First, we need some notion of distance between the estimate $\hat{F}(\cdot)$ and the function $F(\cdot)$. A useful notion is the supremum distance:

$$d(\hat{F}, F) := \sup_y |\hat{F}(y) - F(y)|.$$

This is the maximum pointwise distance between the two functions.

According to the Glivenko-Cantelli theorem:

$$d(\hat{F}, F) \xrightarrow{p} 0.$$

(It can also be shown to converge a.s.)

So this provides a functional version of the consistency of the empirical CDF.

What about convergence in distribution? First, consider the case where we are estimating $F(y)$ at two points, y_1 and y_2. The vector

$$V_i := \begin{pmatrix} 1(Y_i \leq y_1) \\ 1(Y_i \leq y_2) \end{pmatrix}$$

has mean $(F(y_1), F(y_2))'$ and covariance matrix

$$Cov(V) = \begin{pmatrix} F(y_1)(1 - F(y_1) & F(\min(y_1, y_2)) - F(y_1)F(y_2) \\ F(\min(y_1, y_2)) - F(y_1)F(y_2) & F(y_2)(1 - F(y_2) \end{pmatrix}$$

So by the multivariate CLT,

$$\sqrt{n}(\bar{V}_n - E[V_i]) \xrightarrow{d} N(0, Cov(V)).$$

Donsker's Theorem extends this to function-valued estimates. It says that

$$\sqrt{n}(\hat{F} - F) \xrightarrow{d} \mathbb{G},$$

where $\mathbb{G}$ is a gaussian process with covariances

$$Cov(\mathbb{G}(y_1), \mathbb{G}(y_2)) = F(\min(y_1, y_2)) - F(y_1)F(y_2), \quad \forall y_1, y_2$$

For more on this, a good source is van der Vaart, *Asymptotic Statistics*, Cambridge University Press, 1998.