Difference In Differences Methods in Econometrics

Guido W. Imbens - UC Berkeley

Johns Hopkins University

October 21st, 2005

General Motivation

Two or more groups, two or more periods. In some periods some groups are exposed to the treatment.

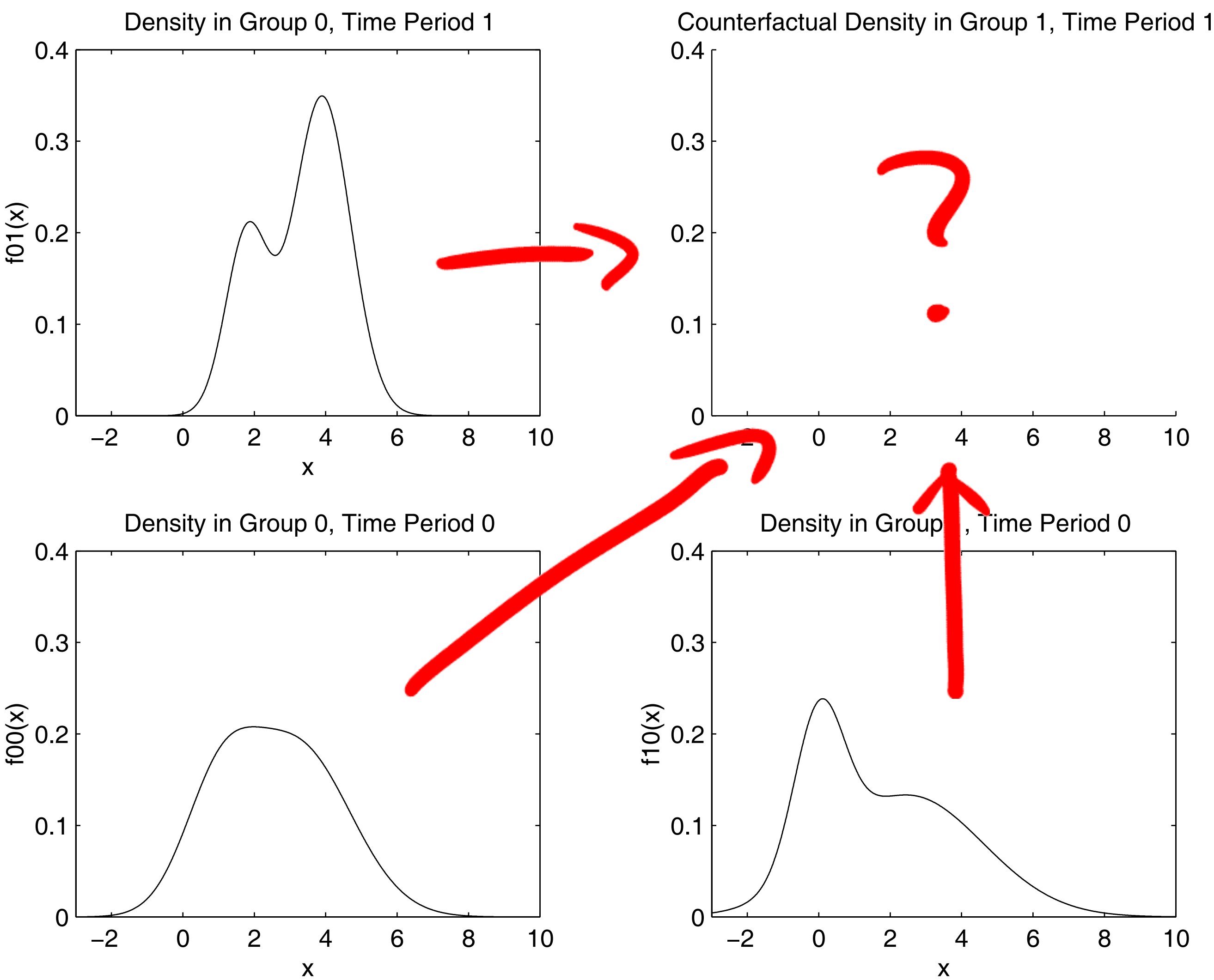
Here mainly two-group/two-period case:

$G_i \in \{0, 1\}$ (group indicator), $T_i \in \{0, 1\}$ (time indicator)

Treatment indicator:

$$W_i = \begin{cases} 1 & \text{if } G_i = 1, T_i = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Data are from repeated cross-sections, not panels. (still works with panels, but other things are possible, e.g., assuming unconfoundedness)



Density in Group 0, Time Period 1
Counterfactual Density in Group 1, Time Period 1
Density in Group 0, Time Period 0
Density in Group 1, Time Period 0
f01(x)
f00(x)
f10(x)
x
x
x
?

Examples I (David Card, 1990)

Muriel boatlift caused increase in low-educated labor supply in Miami.

Card compares labor market outcomes in Miami before and after the boatlift with changes in labor market of comparable cities.

Groups: individuals in Miami vs individuals in other cities

Treatment: Inflow of labor

Examples II (Eissa and Liebman, 1996)

Eissa and Liebam evaluate a tax reform. Only some groups are affected by tax change. They compare changes in outcomes for group affected by tax reform with changes in outcomes for groups not affected by reform.

Groups: individual affected by reform (defined partly by spousal income)

Treatment: change in tax rate

Examples III (Jin and Leslie, 2001)

Jin and Leslie evaluate effect of information disclosure law on restaurant profitability and sales: restaurants in LA county need to post hygiene score cards in the window.

Groups: restaurants in LA County versus restaurants in neighbouring counties.

Treatment: information disclosure requirement

Example IV: (Meyer, Viscusi, Durbin 1995)

MVD evaluate effect of increase in disability payments. This was affected through change in maximum disability payments. This applies only to workers with high earnings (who would otherwise hit the maximum), not to low earners.

Groups: high and low earners

Treatment: change in disab. payment

	Mean Duration in Levels		Mean Dur in Logs	
	$T=0$	$T=1$	$T=0$	$T=1$
$G=0$	6.27	7.03	1.27	1.33
$G=1$	11.18	12.89	1.38	1.58

Standard DID

Unconfoundedness would suggest comparing the $(G=1, T=1)$ and $(G=1, T=0)$ data, or the $(G=1, T=1)$ and $(G=0, T=1)$ data.

DID suggests comparing the $(G=1, T=1)$ and $(G=0, T=1)$, but adjusting for differences we see in the initial, pre-program, period between the $(G=1, T=0)$ and $(G=0, T=0)$.
Standard DID:

$$Y_i = \beta_0 + \beta_1 \cdot G_i + \beta_2 \cdot T_i + \tau \cdot G_i \cdot T_i + \varepsilon_i,$$

with ε independent of G and T, leading to

$$\tau = \mathbb{E}[Y|G=1, T=1] - \mathbb{E}[Y|G=0, T=1]$$

$$-\left(\mathbb{E}[Y|G=1, T=0] - \mathbb{E}[Y|G=0, T=0]\right).$$

Problems with Standard DID

1. Functional form dependent:

MVD data in levels: 0.951 (s.e. 1.26)
MVD data in logs: 0.191 (s.e. 0.068)

Different models, different assumptions.

2. What if heterogeneity in effect of treatment?

3. What is effect for group that was not treated?

Alternative Approach, Athey-Imbens, 2004

Model outcome under control treatment:

$$Y(0) = h(U, T)$$

Outcome under control treatment depends on time period T, and on unobserved individual component U.

Assumptions:
1. $U \perp T | G$ (distribution of U does not vary over time within a group
2. $h(u, t)$ is monotone in u.
3. The support of $U|G = 1$ is a subset of the support of $U|G = 0$.

Comparison with Standard DID Model

Standard DID also assumes:

1. (additivity) $U - \mathbb{E}[h(U,T)|G] \perp G$
2. (single index) $h(u,t) = \phi(u + \delta \cdot t)$
3. (identity function) $\phi(a) = a$

All assumptions are difficult to justify here. They make the results functional form dependent.

Identification of the Distribution of $Y^N|G = 1, T = 1$

$F_{Y(w),gt}(y)$ is the cumulative distribution function of $Y(w)$ given $T = t, G = g$.

$F_{Y,gt}(y)$ is the cumulative distribution function of $Y = WY(1) + (1 - W)Y(0)$ given $T = t, G = g$.

Given the three assumptions:

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

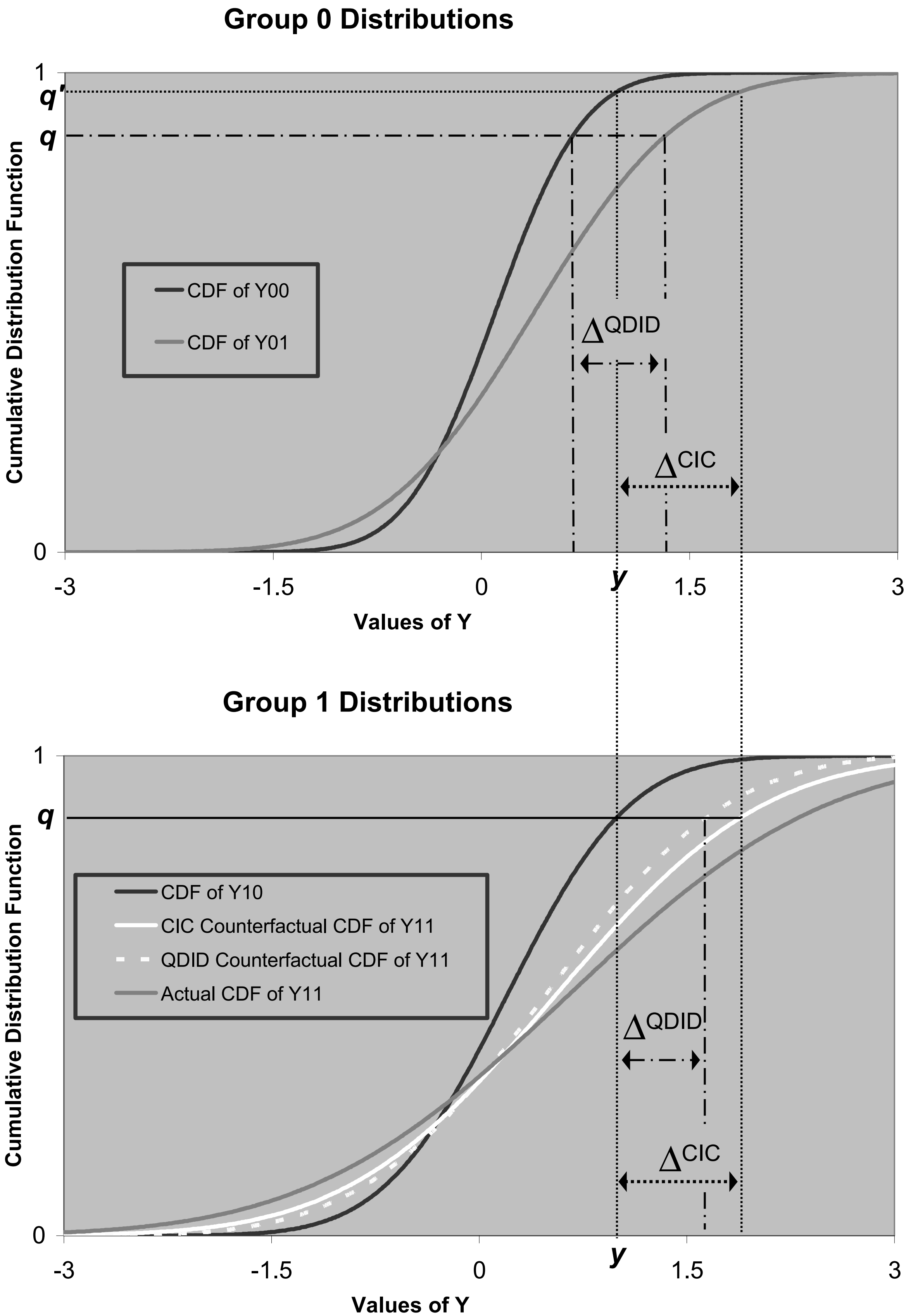


Figure 1: Illustration of Transformations

Interpretation

Take a person in the first period treatment group, with outcome y. What would happen to that person in the second period without the exposure to the treatment.

1. Look at someone with the same value of y in the control group first period. This person must have had the same u.

2. Someone with that value of u in the second period control group would have an outcome y' at the same quantile ($y' = F_{01}^{01}(F_{00}(y))$.)

3. So, outcome distribution $Y_{11}^{N} \sim F_{01}^{-1}(F_{00}(Y_{10}))$

Interpretation of Transformation

So, under this model

$$Y_{11}^N \sim k(Y_{10}),$$

where

$$k(y) = F_{01}^{-1}(F_{00}(y)).$$

Standard DID model:

$$k^{\mathrm{DID}}(y) = y + \mathbb{E}[Y_{01}] - \mathbb{E}[Y_{00}].$$

Advantages Relative to Standard DID Model

- Standard DID depends on functional form: it can hold in levels or logs, but not in both at the same time.

  This model is invariant to monotone transformations. If assumptions 1-3 hold for y, than they hold for $z = g(y)$ if $g(\cdot)$ is strictly monotone. The function $h(u,t)$ changes to $h'(u,t) = g(h(u,t))$, but this still satisfies all assumptions.

- The model is just identified (nothing testable).

Asymptotic Properties

$$\hat{\tau} = \frac{1}{N}\sum_{i=1}^{N_{11}} Y_{11,i} - \sum_{i=1}^{N_{10}} \hat{F}_{01}^{01}(\hat{F}_{00}(Y_{10,i}))$$

Then: (i) $\hat{\tau} - \tau = O_p\left(N^{-1/2}\right)$,

and (ii) $\sqrt{N}\,(\hat{\tau} - \tau) \xrightarrow{d} \mathcal{N}\left(0, V^p/\alpha_{00} + V^q/\alpha_{01} + V^r/\alpha_{10} + V^s/\alpha_{11}\right).$

An initial step in the argument is to linearize the estimator by showing that

$$\hat{\tau} = \tau + \frac{1}{N_{00}}\sum_{i=1}^{N_{00}} p(Y_{00,i}) + \frac{1}{N_{01}}\sum_{i=1}^{N_{01}} q(Y_{01,i})$$

$$+\frac{1}{N_{10}}\sum_{i=1}^{N_{10}} r(Y_{10,i}) + \frac{1}{N_{11}}\sum_{i=1}^{N_{11}} s(Y_{11,i}) + o_p\left(N^{-1/2}\right).$$

Effect of Treatment on Control Group

Model: $Y(1) = h^1(U, T)$,

1. $U \perp T | G$ (dist. of U does not vary over time within group)
2. $h^1(u, t)$ is monotone in u.
3. The supp of $U|G = 0$ is a subset of the supp of $U|G = 1$.

Then

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

Binary Case

In the binary case monotonicity is not an attractive assumption: $U \in \{0, 1\}$, then $\Pr(Y_{01} = 1) = \Pr(Y_{00} = 1)$, or no change over time in control group. This is directly testable.

Standard DID model is not attractive either:

$$\mathbb{E}[Y(0)|G = 1, T = 1] = \mathbb{E}[Y(0)|G = 1, T = 0]$$

$$+ (\mathbb{E}[Y(0)|G = 0, T = 1] - \mathbb{E}[Y(0)|G = 0, T = 0])$$

Suppose $\mathbb{E}[Y_{00}] = 0.8$, $\mathbb{E}[Y_{01}] = 0.2$, $\mathbb{E}[Y_{10}] = 0.5$. Then $\mathbb{E}[Y(0)|G = 1, T = 1)] = -0.1$ which is impossible.

Modification for Binary (and Discrete) Data

1. $h(u, t)$ is weakly monotone in u
2. $U|G$ is continuous
3. $U \perp G|Y, T$ (conditional independence)

The last assumption is key. It is trivially true in the case with $h(u, t)$ monotone in u because U and Y are one-to-one. With weak monotonicity it has content.

Without this assumption we can only infer bounds for τ.

Result for Binary Case

$$\mathbb{E}[Y_{11}^N] = \begin{cases} \frac{\mathbb{E}[Y_{01}]}{\mathbb{E}[Y_{00}]}\mathbb{E}[Y_{10}] & \text{if } \mathbb{E}[Y_{01}] \leq \mathbb{E}[Y_{00}] \\ 1 - \frac{1-\mathbb{E}[Y_{01}]}{1-\mathbb{E}[Y_{00}]}(1-\mathbb{E}[Y_{10}]) & \text{if } \mathbb{E}[Y_{01}] > \mathbb{E}[Y_{00}] \end{cases}$$

Without the conditional independence assumption:

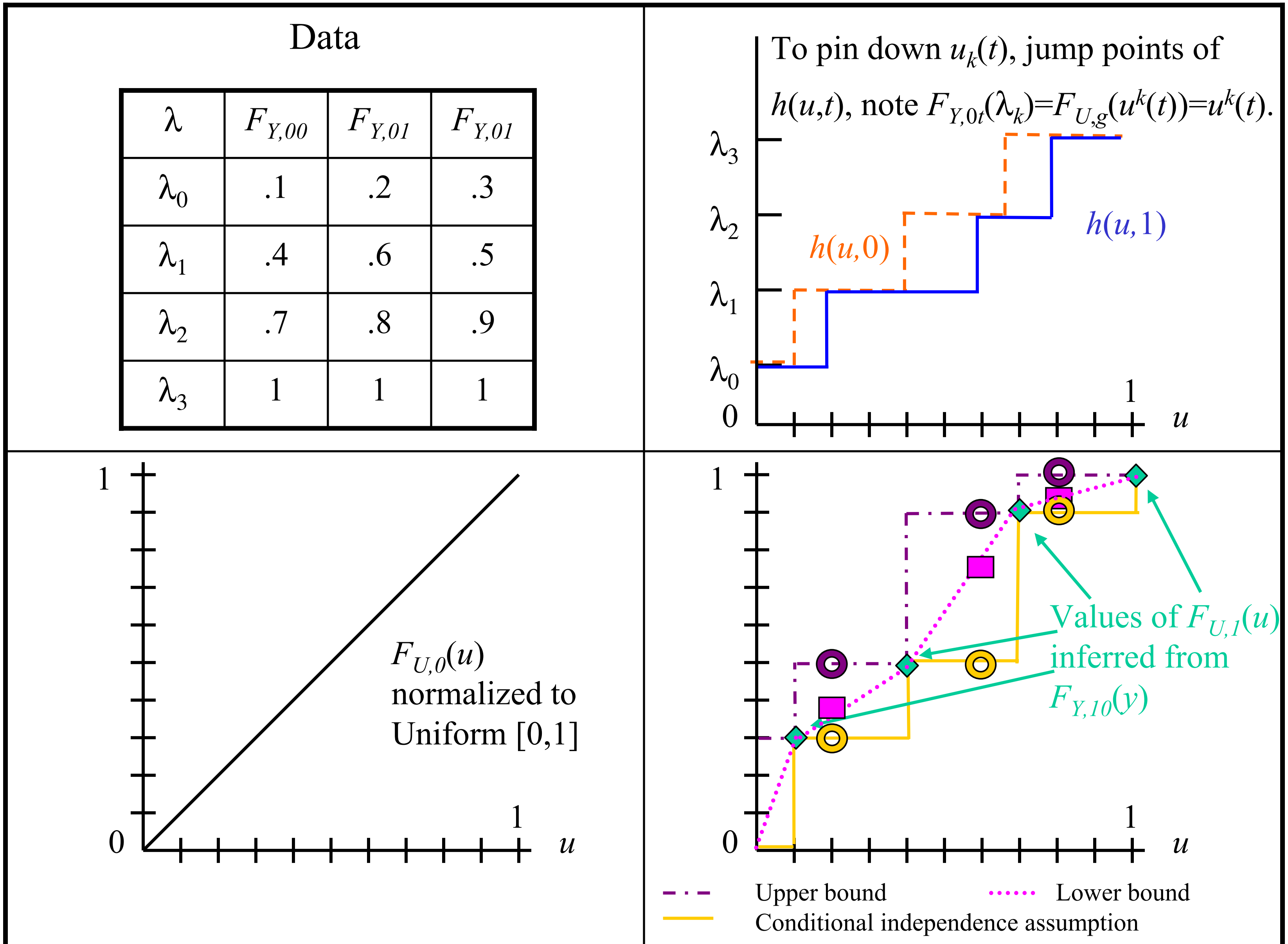
$$\tau \in \begin{cases} [\mathbb{E}[Y_{11}^I] - 1,\ \mathbb{E}[Y_{11}^I] - \mathbb{E}[Y_{10}]] & \text{if } \mathbb{E}[Y_{01}] > \mathbb{E}[Y_{00}] \\ \mathbb{E}[Y_{11}^I] - \mathbb{E}[Y_{10}] & \text{if } \mathbb{E}[Y_{01}] = \mathbb{E}[Y_{00}] \\ [\mathbb{E}[Y_{11}^I] - \mathbb{E}[Y_{10}],\ \mathbb{E}[Y_{11}^I]] & \text{if } \mathbb{E}[Y_{01}] < \mathbb{E}[Y_{00}]. \end{cases}$$

Extension to General Discrete Case

The results for the binary case can be extended to the general discrete case. This includes lower/upper bound, and conditional independence case.

With many points of support all three discrete estimators will get close to continuous case.

Figure 2: Bounds and the Conditional Independence Assumption in the Discrete Model



λ	$F_{Y,00}$	$F_{Y,01}$	$F_{Y,01}$
λ_0	.1	.2	.3
λ_1	.4	.6	.5
λ_2	.7	.8	.9
λ_3	1	1	1

Estimates for Meyer-Viscusi-Durbin Data

	mean weeks	(s.e.)	mean logs	(s.e.)	percentiles 25th	50th	75th	90th
DID-level	0.95	(1.27)	–	–	-0.77	0.23	1.23	5.23
DID-logs	1.63	(1.26)	0.19	(0.07)	-0.02	0.97	1.94	5.87
CIC ci	0.39	(1.57)	0.18	(0.07)	0.00	1.00	2.00	5.00
CIC lower	0.07	(1.60)	0.14	(0.11)	0.00	1.00	1.00	4.00
CIC upper	1.08	(1.64)	0.58	(0.14)	1.00	2.00	2.00	5.00