Fuzzy Changes-in-Changes

Clément de Chaisemartin    Xavier D’Haultfœuille
Warwick University          CREST

September 2014
Outline

Motivation

The standard CIC

The IV-CIC model

Inference

Applications

Conclusion
The DID method

- **DID**: commonly used method to measure effect of a non-randomized treatment.

- Basic framework: 2 groups, 2 periods, the treatment group becomes fully treated in the 2nd period.

- **DID estimator** compares the evolution of the mean of the outcome in the 2 groups.

- Identifying assumption: common trends.
Some issues with common trends, and the CIC model

- The common trend condition is not credible if the effect of time heterogeneous: job training example.

- Several alternative have been proposed so far:
  - DID matching (Heckman, Ichimura & Todd, 1997; Blundell et al., 2004; Abadie, 2005);
  - Allowing for some variations in the way time affects the control and treatment groups (Donald & Lang, 2007; Manski & Pepper, 2012);
  - including heterogeneous effects of time, identified through an instrument (Bonhomme & Sauder, 2011).

- Another issue with the common trend assumption is that it is not invariant to scaling: Meyer, Viscusi & Durbin (1995).

- Athey & Imbens (2006) develop CIC model to tackle these two shortcomings.

- Identifying assumption: common changes.
The IV-DID method

- Many natural experiments cannot be analyzed within the standard DID framework, because one cannot construct sharp control and treatment groups.

- Only fuzzy control and treatment groups available: treatment rate increases more in the treatment than in the control group: e.g. Duflo (2001).


- Estimand arising from this regression: Wald-DID.
Some issues with IV-DID, and the IV-CIC model

- Causal interpretation of IV-DID estimand requires common trends assumptions, both for the treatment and for the outcome.

- Also, requires some kind of common effects assumption across groups (de Chaisemartin, 2013):
  - Assume treatment effect $> 0$, constant within groups, heterogeneous across groups: twice as large in the control, and treatment rate increases twice as much in the treatment group.
  - Then Wald-DID=0, while everybody has a positive treatment effect.

- We study an IV-CIC model which allows for heterogeneous effects of time, is invariant to the scaling of the outcome, and does not impose common effects assumptions in the treatment and in the control groups.
Main results

▶ Under our IV-CIC model, average and quantile treatment effects are point identified when the treatment rate is stable in the control group.

▶ Otherwise, we provide bounds on treatment effects. The smaller the change in the treatment rate in the control group, the tighter the bounds.

▶ We prove that these bounds are sharp under testable conditions.

▶ We develop inference for both the point and partially identified cases.
Main results

- Finally, we revisit results in Field (2007) and Duflo (2001):
  - Field (2007) is in the point identified case. Our IV-CIC results on average effects are in line with her IV-DID results. Heterogeneous effects along the distribution of the outcome.
  - Duflo (2001) is in the partially identified case. Our IV-CIC model yields wide and uninformative bounds.
  - We also consider an application “in between”, where the model is partially identified but we can still obtain informative results.
- Take away: in fuzzy settings, researchers must find a control group with a relatively stable treatment rate to derive informative estimates of treatment effects without homogeneity and linearity assumptions.
Outline

Motivation

The standard CIC

The IV-CIC model

Inference

Applications

Conclusion
The framework

- We are interested in the effect of a binary treatment $D$ on a continuous outcome $Y$. $Y(d)$ denotes the potential outcome with treatment $d$, while $Y = Y(D)$ is the observed outcome.

- We have repeated cross section data with two periods, $T = 0$ and 1, and two groups, the control and treatment group ($G = 0$ and 1, resp.). The control group is not treated while the treatment group becomes treated at period 1.

- Main assumption: had there been no treatment in period 1, two individuals in the control and treatment groups with identical $Y$ at period 0 would have had also a similar $Y$ at period 1.
The CIC transform.
CIC: a production function model.

Athey and Imbens consider the following model: $Y(0) = h_0(U_0, T)$, and make the following assumptions:

C.1: monotonicity
$h_0(., t)$ is strictly increasing for $t \in \{0; 1\}$.

C.2: Time invariance within groups
$U_0 \perp \perp T|G$.

C.3: Support restriction
$\text{Supp}(Y|G = 1) \subseteq \text{Supp}(Y|G = 0)$.

▶ If $Y(0)$ is a test score, or wages, $U_0$ can be interpreted as ability.
▶ C.1: wage is a strictly increasing function of ability.
▶ C.2: within each group, the distribution of ability does not change over time. Key assumption of the model. Could be violated, e.g. if observations switch from one group to the other over time.
▶ C.3: technical restriction, could be dropped but then bounds.
Intuition on why the CIC model rationalize the CIC transform

- CIC transform may be seen as a double matching.
- 1st matching: each treatment group unit in period 0 is matched to a control group unit in period 0 with the same $y$.
- Their $y$ is $h_0(u_0, 0)$. Under C1, they must have the same $u_0$.
- 2nd matching: the control group unit in period 0 is matched to his rank counterpart in period 1. Under C1 and C2, they must have the same $u_0$.
- Therefore, the $y$ of the second control unit is equal to $h_0(u_0, 1)$: this is the $y$ that the treatment group unit would have had in period 1 if he had remained untreated.
Outline

Motivation

The standard CIC

The IV-CIC model

Inference

Applications

Conclusion
An IV model

- We assume that in period 1, the treatment group receives some specific incentives for treatment $Z = T \times G$.

- Two potential treatments $D(0)$ and $D(1)$ depending on whether one receives incentives for treatment.
  Duflo (2001): $D(1) =$ will I go to school if a new school is built close to my home?

- Exclusion restriction: the instrument has no effect per se on the outcome.
We consider the model
\[ Y(d) = h_d(U_d, T), \quad d \in \{0; 1\} \]
and we impose the following conditions.

**A.1** (monotonicity) \( h_d(., t) \) is strictly increasing.

**A.2** (no defiers) \( D(z) = 1\{V \geq v_z(T)\} \), with \( v_0(t) > v_1(t) \) for \( t \in \{0, 1\} \) (equivalent to imposing no defiers, see Vytlacil, 2002).

**A.3** (Time invariance within groups) For \( d \in \{0, 1\}, (U_d, V) \perp \perp T|G \).

\[ U_d \] can be interpreted as ability under treatment \( d \), \( V \) is one’s taste for treatment.

**A.2:** The instrument cannot decrease propensity for treatment.

**A.3:** Key extra ingredient relative to CIC. Implies \( U_d \perp \perp T|G, V \).
With an extended Roy model of selection, 
\( D(z) = 1 \{ Y(1) - Y(0) \geq C(z) \} \), A.3 is compatible with
\[
Y(d) = U_d + \eta_d T + \gamma U_d T. \tag{1}
\]
On the other hand, incompatible with \( Y(d) = U_d + \eta_d T + \gamma_d U_d T \).

Time can have heterogeneous effects on the outcome, but the treatment cannot affect heterogeneity in trends.

IV-DID is far more difficult to rationalize under a Roy selection. Typically fails under (1).
The IV-CIC model

- Additional, more minor, requirements. Let us define

\[ D_{gt} \sim D \mid G = g, T = t, \quad Y_{dgt} \sim Y \mid D = d, G = g, T = t. \]

A.4 (data restrictions)
1. \( \text{Supp}(Y_{dgt}) = \text{Supp}(Y) = [\underline{y}, \bar{y}], \ (\underline{y}, \bar{y}) \in \mathbb{R}^2 \).
2. \( F_{Y_{dgt}} \) is strictly increasing and continuous on \( [\underline{y}, \bar{y}] \).

A.5 (rank condition) \( P(D_{11} = 1) > P(D_{10} = 1) \).
As in the IV-LATE framework, it is useful to define several populations. First, remark that under A.2, A.3 and A.5, \( v_0(0) > v_1(1) \). Moreover, by A.2, \( v_0(1) \geq v_1(1) \). Then define:

- **AT** = \( \{ V \geq v_0(0) \} \): always takers in period 0.
- **NT** = \( \{ V < v_1(1) \} \): never takers in period 1.
- **TC** = \( \{ V \in [\min(v_0(0), v_0(1)), \max(v_0(0), v_0(1))] \} \): units that switch because of time (time compliers).
- **IC** = \( \{ V \in [v_1(1), v_0(1)) \} \): units that switch because of the instrument (instrument compliers).
- **C** = \( \{ V \in [v_1(1), v_0(0)) \} \): units that become treated either through the effect of \( Z \) or time. If the treatment rate increases in the control group, \( C = IC \cup TC \). Otherwise, \( C = IC \setminus TC \).
### Four populations

<table>
<thead>
<tr>
<th></th>
<th>Period 0</th>
<th>Period 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control Group</strong></td>
<td>30% treated: Always takers</td>
<td>35% treated: Always takers and Time Compliers</td>
</tr>
<tr>
<td></td>
<td>70% untreated: Never Takers, Instrument Compliers and Time Compliers</td>
<td>65% untreated: Never Takers and Instrument Compliers</td>
</tr>
<tr>
<td><strong>Treatment Group</strong></td>
<td>25% treated: Always takers</td>
<td>60% treated: Always takers, Instrument Compliers and Time Compliers</td>
</tr>
<tr>
<td></td>
<td>75% untreated: Never Takers, Instrument Compliers and Time Compliers</td>
<td>40% Untreated: Never Takers</td>
</tr>
</tbody>
</table>
Parameters of interest

- Here we focus on the population of compliers instead of instrument compliers. We cannot identify anything about instrument compliers in general.

- We consider two types of parameters of interest:
  - The average treatment effect on compliers $\Delta = E(Y(1) - Y(0) | C)$.
  - Quantile treatment effects $\tau_q = F_{Y_{11}(1)}^{-1}(\tau | C) - F_{Y_{11}(0)}^{-1}(\tau | C)$ for any $\tau \in (0, 1)$.

- Let us introduce the P-P and Q-Q transforms, $H_d = F_{Y_{d10}} \circ F_{Y_{d00}}^{-1}$ and $Q_d = F_{Y_{d01}}^{-1} \circ F_{Y_{d00}}$. 
When the treatment rate is stable in the control group: point identification.

- Suppose first $P(D_{00} = 1) = P(D_{01} = 1)$. Applies for instance to extensions of a public policy. The “control group” is already partly treated at period 0 (cf. Field, 2007).

**Theorem 1**

If A.1-A.5 hold and, for $d \in \{0, 1\}$, $P(D_{00} = d) = P(D_{01} = d) > 0$,

$$F_{Y_{11}(d)|c} = \frac{P(D_{10} = d)H_d \circ F_{Y_{d01}} - P(D_{11} = d)F_{Y_{d11}}}{P(D_{10} = d) - P(D_{11} = d)}.$$  \hspace{1cm} (2)

Hence, $\Delta$ and $\tau_q$ are identified when $P(D_{00} = 0) = P(D_{01} = 0) \in (0, 1)$. Moreover,

$$\Delta = \frac{E(Y|G = 1, T = 1) - E(Q_D(Y)|G = 1, T = 0)}{E(D|G = 1, T = 1) - E(D|G = 1, T = 0)}.$$  \hspace{1cm} (3)

- Cases where $P(D_{00} = 0) = P(D_{01} = 0) \in \{0, 1\}$ are also point identified under similar conditions.
Intuition (1/2)

\[
\Delta = \frac{E(Y|G = 1, T = 1) - E(Q_D(Y)|G = 1, T = 0)}{E(D|G = 1, T = 1) - E(D|G = 1, T = 0)}.
\]

- RHS \(\approx\) standard Wald ratio in treatment group, instrument = \(T\).
- Difference: \(Q_D(Y)\) instead of \(Y\).
- Time is not a standard IV, it has two effects on the outcome:
  - Indirect through treatment: treatment rate increases over time.
  - But also a direct effect: means of \(Y(0)\) and \(Y(1)\) change over time.
- We want to net out the direct effect to recover the indirect one.
- We can recover the direct effect from the control group. Treatment rate stable there, indirect effect absent. The evolution of \(Y\) in this group gives us the effect of time on the outcome.
- \(Q_D(Y)\) measures the effect of time in the control and nets it out in the treatment.
## Intuition (2/2)

<table>
<thead>
<tr>
<th></th>
<th>Period 0</th>
<th>Period 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control Group</strong></td>
<td>30% treated: Always Takers</td>
<td>30% treated: Always Takers</td>
</tr>
<tr>
<td></td>
<td>70% untreated: Never Takers and</td>
<td>70% untreated: Never Takers and</td>
</tr>
<tr>
<td></td>
<td>Compliers</td>
<td>Compliers</td>
</tr>
<tr>
<td><strong>Treatment Group</strong></td>
<td>20% treated: Always Takers</td>
<td>65% treated: Always Takers and</td>
</tr>
<tr>
<td></td>
<td>80% untreated: Never Takers and</td>
<td>Compliers</td>
</tr>
<tr>
<td></td>
<td>Compliers</td>
<td>35% Untreated: Never Takers</td>
</tr>
</tbody>
</table>

- \( Q_D(.) \) does double matching within subgroups with same treatment.
- In period 1, we observe distribution of \( Y(1) \) in treatment group for always takers and compliers.
- We would like to withdraw distribution of \( Y(1) \) for always takers to get compliers alone.
- Not observed, but can be recovered through double matching.
When the treatment rate changes in the control group: partial identification.

- If treatment rate changes in the control group as well, we cannot recover the direct effect of time from this group.

- But we can bound it, and derive bounds for the LATE and QTE we are interested in.

- Intuition: second matching partly collapses.

<table>
<thead>
<tr>
<th></th>
<th>Period 0</th>
<th>Period 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control Group</strong></td>
<td>30% treated: Always takers</td>
<td>35% treated: Always takers and Time Compliers</td>
</tr>
<tr>
<td></td>
<td>70% untreated: Never Takers, Instrument Compliers and Time Compliers</td>
<td>65% untreated: Never Takers and Instrument Compliers</td>
</tr>
<tr>
<td><strong>Treatment Group</strong></td>
<td>25% treated: Always takers</td>
<td>60% treated: Always takers, Instrument Compliers and Time Compliers</td>
</tr>
<tr>
<td></td>
<td>75% untreated: Never Takers, Instrument Compliers and Time Compliers</td>
<td>40% Untreated: Never Takers</td>
</tr>
</tbody>
</table>
The main tool for building bounds is the following result:

**Lemma 1**

If A.1-A.5 hold and \( P(D_{00} = d) > 0 \), then:

\[
F_{Y_{11}(d)|C} = \frac{P(D_{10} = d)H_d \circ (\lambda_d F_{Y_{d01}} + (1 - \lambda_d) F_{Y_{01}(d)|TC}) - P(D_{11} = d) F_{Y_{d11}}}{P(D_{10} = d) - P(D_{11} = d)}. \tag{4}
\]

Where \( \lambda_d = \frac{P(D_{01} = d)}{P(D_{00} = d)} \).

- The only unknown quantity in (4) is \( F_{Y_{01}(d)|TC} \).

- Bounds can be obtained simply replacing it by 0 or 1.

- Actually it is possible to do better, using the fact that \( \lambda_d F_{Y_{d01}} + (1 - \lambda_d) F_{Y_{01}(d)|TC} \) and \( F_{Y_{11}(d)|C} \) should be cdf as well.
Constructing the bounds (2/2)

When \( P(D_{00} = d) > 0 \) \( \triangleright \text{other case} \), let \( T_d = F_{\gamma_{01}(d)}|T_C \), and define

\[
G_d(T_d) = \lambda_d F_{\gamma_{d01}} + (1 - \lambda_d) T_d
\]
\[
C_d(T_d) = \frac{P(D_{10} = d) H_d \circ G_d(T_d) - P(D_{11} = d) F_{\gamma_{d11}}}{P(D_{10} = d) - P(D_{11} = d)}.
\]

Then, let \( M(x) = \max(0, \min(1, x)) \), \( \mu_d = P(D_{11} = d)/P(D_{10} = d) \) and

\[
\begin{align*}
\underline{T}_d & = M \left( \frac{\lambda_d F_{\gamma_{d01}} - H_d^{-1}(\mu_d F_{\gamma_{d11}})}{\lambda_d - 1} \right), \\
\overline{T}_d & = M \left( \frac{\lambda_d F_{\gamma_{d01}} - H_d^{-1}(\mu_d F_{\gamma_{d11}} + (1 - \mu_d))}{\lambda_d - 1} \right).
\end{align*}
\]

We then define our bounds as:

\[
\begin{align*}
\underline{B}_d(y) & = \sup_{y' \leq y} C_d \left( T_d \right) (y'), \\
\overline{B}_d(y) & = \inf_{y' \geq y} C_d \left( T_d \right) (y').
\end{align*}
\]
Bounds are sharp under a condition testable from the data.

A.6 (Increasing bounds) For \((d, g, t) \in \{0, 1\}^3\), \(F_{Y_{dgt}}\) is continuously differentiable, with positive derivative on \(\text{Supp}(Y)\). Moreover, \(T_d\), \(G_d(T_d)\) and \(C_d(T_d)\) (resp. \(\overline{T}_d\), \(G_d(\overline{T}_d)\) and \(C_d(\overline{T}_d)\)) are increasing.

Theorem 2
If A.1-A.5 hold, we have

\[
\underline{B}_d(y) \leq F_{Y_{11}(d)}|C(y) \leq \overline{B}_d(y).
\]

Moreover, if A.6 holds, \(\underline{B}_d(y)\) and \(\overline{B}_d(y)\) are sharp.

▶ We thus prove not only that \(\underline{B}_d\) and \(\overline{B}_d\) are valid bounds for \(F_{Y_{11}(d)}|C(y)\), but also that they actually exhaust the information of the model and the data under the testable restriction A.6.

▶ Using these bounds, we derive sharp bounds on \(\Delta\) and \(\tau_q\).
Outline

Motivation

The standard CIC

The IV-CIC model

Inference

Applications

Conclusion
Inference in the point identified case

- We follow closely our identification method.

- Let \( \hat{F}_{dgt} \) (resp. \( \hat{F}_{dgt}^{-1} \)) denote the empirical cdf of \( Y \) on the subsample \( \{i : D_i = d, G_i = g, T_i = t\} \) and \( \hat{Q}_d = \hat{F}_{d01}^{-1} \circ \hat{F}_{d00} \). We also let \( \mathcal{I}_{gt} = \{i : G_i = g, T_i = t\} \) and \( n_{gt} \) denote the size of \( \mathcal{I}_{gt} \) for all \( (d, g, t) \in \{0, 1\}^3 \).

- Our estimator of the LATE is defined by:

\[
\hat{\Delta} = \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} Y_i - \frac{1}{n_{10}} \sum_{i \in \mathcal{I}_{10}} \hat{Q}_{D_i}(Y_i) - \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} D_i - \frac{1}{n_{10}} \sum_{i \in \mathcal{I}_{10}} D_i.
\]
Inference in the point identified case

- To estimate the QTE, let us define $\hat{P}(D_{gt} = d)$ as the proportion of $D = d$ in the sample $I_{gt}$, $\hat{H}_1 = \hat{F}_{110} \circ \hat{F}_{100}^{-1}$ and
  $$\hat{F}_{Y_{11}(d)|C} = \frac{\hat{P}(D_{11} = d)\hat{F}_{d11} - \hat{P}(D_{01} = 1)\hat{H}_1 \circ \hat{F}_{d01}}{\hat{P}(D_{11} = 1) - \hat{P}(D_{10} = 1)}.$$ 

- Then the estimator of the QTE is the simple plug-in estimator:
  $$\hat{\tau}_q = \hat{F}_{Y_{11}(1)|C}(q) - \hat{F}_{Y_{11}(0)|C}(q).$$
Inference in the point identified case

We impose the following conditions.

**A.7** \((Y_i, D_i, G_i, T_i)_{i=1,\ldots,n}\) are i.i.d.

**A.8** \(\text{Supp}(Y)\) is a bounded interval \([y, \bar{y}]\). Moreover, for all \((d, g, t) \in \{0, 1\}^3\), \(F_{Y_{dgt}}\) and \(F_{Y_{11}(d)|\mathcal{C}}\) are \(C^1\) with strictly positive derivatives on \([y, \bar{y}]\).

**Theorem 3**

*Suppose that Assumptions A.1-A.5, A.7-A.8 hold and \(P(D_{00} = 0) = P(D_{01} = 1) \in (0, 1)\). Then \(\hat{\Delta}\) and \(\hat{\tau}_q\) are root-n consistent and asymptotically normal. The bootstrap is consistent for both \(\hat{\Delta}\) and \(\hat{\tau}_q\).*

- Argument distinct from Athey & Imbens (2006): we rely on the functional delta method, by proving that the estimators are Hadamard differentiable functionals of standard empirical processes.

- A similar construction and reasoning applies to the partially identified case.
Outline

Motivation

The standard CIC

The IV-CIC model

Inference

Applications

Conclusion
1996-2003: Peruvian government issued very large number of property titles to urban squatters.

Field uses a survey conducted in 2000 to measure the effect of being granted a property title on labor supply. Two sources of variation:

- Squatters / non squatters.
- Neighborhood reached by the program or not.

<table>
<thead>
<tr>
<th></th>
<th>Reached after 2000</th>
<th>Reached before 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-squatters</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Squatters</td>
<td>0%</td>
<td>71%</td>
</tr>
</tbody>
</table>

**Table 1**: Share of households with a property right
Field (2007): property titles and labor supply (2/2)

- Field uses IV-DID. LATE: 27.7. Property title increases labor supply of compliant households by 27.7 hours per week.

- Similar (though significantly different) LATE with the IV-CIC model: 23.3.

- Quantile effects in the number of hours worked are similar at each quantile, implying a large heterogeneity in relative terms (+40% at the 1st quartile, +10% at the 3rd quartile).

Figure 1: Estimated QTE on the number of hours worked.
Effectiveness of a smoking cessation treatment (1/2).


- We study its effect on smoking intensity, through carbon monoxide concentration. Database from 17 French smoking cessation clinics, in which doctors, nurses, and psychologists help smokers quit.

- Very strong predictor of clinics prescription rate: share of their staff holding a degree in tabacology. Proxy of proneness to adopt technological innovations

- We use this share to construct a control and treatment group.

Table 2: Share of patients prescribed varenicline

<table>
<thead>
<tr>
<th></th>
<th>Before February 2007</th>
<th>After February 2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control clinics</td>
<td>0%</td>
<td>4.9%</td>
</tr>
<tr>
<td>Treatment clinics</td>
<td>0%</td>
<td>25.4%</td>
</tr>
</tbody>
</table>

- $F_{Y(1)|C}$ is point identified while $F_{Y(0)|C}$ is partially identified.
Effectiveness of a smoking cessation treatment (2/2).

Figure 2: Estimated bounds for QTE, including baseline CO as a control.

- 50% of smokers quit irrespective of whether they receive Varenicline or not.
- But we obtain significant negative QTE for large quantiles.
- If a rank invariance condition holds, this means that the effect is concentrated on heavy smokers.


- Cohorts who had already completed primary school not exposed + variation in treatment intensity across regions: government allocated more schools to districts with low initial enrolment.

<table>
<thead>
<tr>
<th></th>
<th>Older cohort</th>
<th>Younger cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low treatment regions</td>
<td>89.8%</td>
<td>94.3%</td>
</tr>
<tr>
<td>High treatment regions</td>
<td>81.2%</td>
<td>90.0%</td>
</tr>
</tbody>
</table>

Table 3: Share of individuals completing primary school

- In this case, $\lambda_0 \approx 1.79$ is far away from 1 $\Rightarrow$ wide bounds on $F_{Y(0)|C}$.

- Translates into uninformative results, e.g. $QTE(0.5) \in [-6.29; 6.81]$. 
Outline

Motivation

The standard CIC

The IV-CIC model

Inference

Applications

Conclusion
Summary

- We develop an extension of the CIC model, the IV-CIC model, which tackles important shortcomings of IV-DID: allows for heterogeneous effects of time, invariant to scaling of the outcome, does not impose common effects assumptions across groups.

- Under IV-CIC assumptions, LATE and QTE on compliers are:
  - Point identified when % treated stable in the control group.
  - Partially identified when % treated changes in the control: the smaller this increase, the tighter the bounds.

- We use those results to revisit results in Field (2007) and Duflo (2001).

- In applications in which treatment rate stable in the control group, better use IV-CIC: relies on weaker assumptions. In very fuzzy applications, one can use IV-DID but important to discuss the common trends and effects assumptions.
Here we identify the effect of an endogenous treatment using time variation, even though time has an effect on the outcome.

As a continuation, two research projects on non-standard instruments:

- Identification of treatment effects for a continuous treatment, using also time variation (with Hoderlein and Sasaki);

- Identification of the effect of a continuous, endogenous regressors when instruments are also included, so that $Y = g(\varphi(X, Z) , \varepsilon)$ (also with Hoderlein and Sasaki).

Key ideas: (i) we have to impose some structure on how the instrument affects $Y$; (ii) the instrument should have an heterogenous effect on $X$, so that we can separate the direct effect of $Z$ from its indirect effect, through $X$. 


Partial identification

When $P(D_{00} = d) = 0$, the bounds on $F_{Y_{11}(d)|C}$ are much more simple. We simply bound $F_{Y_{11}(1)|(2d-1)\geq(2d-1)v_0(0)}$ by 0 and 1, which yields

$$B_d(y) = M_0 \left( \frac{P(D_{10} = d) - P(D_{11} = d)F_{Y_{11}|D=d}}{P(D_{10} = d) - P(D_{11} = d)} \right),$$

$$\overline{B}_d(y) = m_1 \left( \frac{-P(D_{11} = d)F_{Y_{11}|D=d}}{P(D_{10} = d) - P(D_{11} = d)} \right).$$

When $d = 0$, we obtain in particular the trivial bounds $B_0(y) = 0$ and $\overline{B}_1(y) = 1$.