Part D: Instrumental Variables

D4: Judge IV Design

Kirill Borusyak

ARE 213 Applied Econometrics

UC Berkeley, Fall 2023

D4 Outline

1. Judge IV / Examiner IV / Judge FE / Leniency designs

2. A few words on control functions

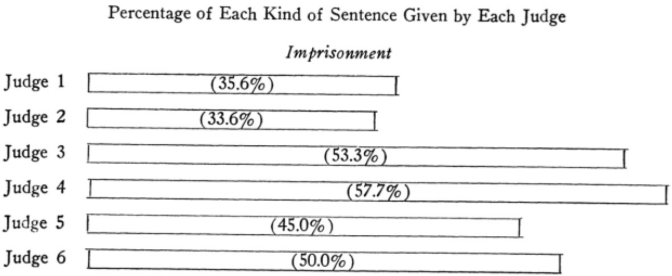
Readings: Scott Cunnigham's "Mixtape" textbook, Ch. 7.8.2

Setting

- There are many situations where:
  1. the treatment (usually binary) is decided by one of K "judges" (examiners, caseworkers, ...)
  2. judge's decision is discretionary
  3. judges are assigned to cases randomly (perhaps within strata, e.g. location-period)
- Examples of treatments:
  - Incarceration *(Kling 2006, Mueller-Smith 2015)*, bail *(Arnold, Dobbie, Yang 2018)*
  - Patent granting *(Sampat and Williams 2019)*
  - Credit ratings *(Rieber and Schechinger 2019)*
  - Psychotherapy treatment *(Blæhr and Søgaard 2021)*
  - Several types of job training programs for the unemployed *(Humlum, Munch, Rasmussen 2023)*

Idea

- Judges typically vary by leniency



(Gaudet, Harris and John 1933, reproduced from Scott Cunningham's "Mixtape")

- Can use this heterogeneity to instrument for treatment
  - Leniency is unobserved $\implies$ what should we do?
  - Under which assumptions is the answer causal?

Estimated leniency?

- Notation: judge assignment $Q_i = k \in \{1, \dots, K\}$; $Z_{ki} = \mathbf{1}[Q_i = k]$
- Consider binary treatment. Popular naïve idea:
  - Measure judge leniency as % of lenient decisions: $\hat{L}_k = \frac{\sum_i Z_{ki} D_i}{\sum_i Z_{ki}}$
  - Then instrument D_i with $\hat{L}_{Q_i}$
- Problem 1: assignment is only random within strata
  - We want to avoid variation in $\hat{L}_k$ reflecting the types of cases k's strata receives
- Problem 2: $\hat{L}_k$ is noisy if there are many judges (and not so many cases per judge)
  - $\hat{L}_{Q_i}$ is influenced by D_i, which correlates with $\varepsilon_i \implies$ bias
  - Conventional inference doesn't take into account estimation noise

Correct IV framework

- Note that $\hat{L}_k$ = **judge fixed effects**, i.e. OLS estimates from a first-stage

$$D_i = \sum_k L_k Z_{ik} + u_i$$

  - And $\hat{L}_{Q_i}$ are first-stage fitted values

  $\Rightarrow$ Using $\hat{L}_{Q_i}$ as IV $\iff$ 2SLS with $Z_1, \ldots, Z_K$ instruments
- Problem 1 $\implies$ include strata FEs as covariates
  - Identify judge FEs relative to strata mean but that's fine
- Problem 2 is the standard many weak IV problem
  - Without covariates, JIVE is intuitive: uses leave-out leniency $\hat{L}_i = \frac{\sum_j \mathbf{1}[Q_j = Q_i] D_j}{\sum_j \mathbf{1}[Q_j = Q_i]}$
  - With (many) covariates, better to use UJIVE ("unbiased JIVE"; Kolesar 2013): a version of JIVE that is consistent with many controls

Underlying assumptions

With judge assignment dummies $Z_{i1}, \ldots, Z_{iK}$ as IVs, what about:

- Independence?
- Exclusion?
- Monotonicity?

Underlying assumptions

With judge assignment dummies $Z_{i1}, \dots, Z_{iK}$ as IVs, what about:

- Independence?
  - Guaranteed by random assignment, as long as strata FEs are controlled for
- Exclusion?
  - Does the judge directly make only one decision D_i?
  - Can the judge indirectly affect others treatments, e.g. by affecting who will be making those decisions?
- Monotonicity: very strong
  - A judge who is more lenient on average should be weakly more lenient on everyone
  - Violated if judges put different weights on different characteristics of the case

Partial tests for monotonicity and exclusion

- Reject monotonicity if leniency in subgroups (based on observables) don't have the same ranks across judges as overall leniency (see Dobbie, Goldin, Yang 2018)
- Frandsen, Lefgren, Leslie (2023):
  - Under LATE assumptions, comparing any two judges gives causal effects for some complier population
  - Causal effects cannot be too large: e.g. bounded by the range of possible outcomes
  - Reject exclusion or monotonicity if the relationship between $\mathbb{E}[Y_i \mid Q_i]$ and $\mathbb{E}[D_i \mid Q_i]$ is too steep for some pair of judges
- Should we panic if monotonicity doesn't hold?
  - Not with homogeneous effects (see also de Chaisemartin 2017 "Tolerating defiers")
  - Frandsen et al.: 2SLS identifies a convex average of causal effects under "average monotonicity": for all i, $D_i(k)$ is positively correlated with L_k

Outline

2SLS as a control function estimator

- We have two strategies for dealing with endogenous D_i: adding controls and IV
  - Are they related to each other?
- IV isolates "clean" variation in D_i
  - 2SLS estimates first-stage $D_i = \hat{\pi}' Z_i + \hat{u}_i$ and regresses Y_i on $\hat{\pi}' Z_i$
- What if we control for "dirty" variation in D_i (**control function** estimator):

$$Y_i = \tau D_i + \mu \hat{u}_i + \text{error}$$

- *Exercise*: the two estimators are numerically the same (even with covariates)
- Curious. But is the control function approach to IV useful?
  - Equivalence breaks with nonlinearities: e.g. first-stage is probit or has square terms
  - Helpful when combining IVs with parametric restrictions: e.g. Heckman selection model that imposes joint normality of ε_i and latent first-stage error