Data Sharing and Citations

Causal Evidence

Garret Christensen[1] Allan Dafoe[2] Edward Miguel[3]

[1]Berkeley Institute for Data Science, UC Berkeley

[2]Department of Political Science, Yale University

[3]Department of Economics, UC Berkeley

November 7, 2017 MSDSE

PRELIMINARY–Please do not cite.

Data Sharing Incentives

- Shared data is a public good. (See Newton 1675)
- Public goods are often undersupplied.
- Is there private incentive?
  - Promotion & tenure
  - Citations

Existing Evidence

- Piwowar, Day, Fridsma (2007): 69% more citations for cancer microarray clinical trials papers (N=85).
- Piwowar, Vision (2013): 9% more citations for gene expression microarray papers with public data (N=10,555).

The Case of Political Science

Exploit plausibly exogenous variation in data availability caused by the abrupt change in editorial policy at a top political science journal, *The American Journal of Political Science* (*AJPS*). Rick Wilson became the editor on January 1, 2010:

> *"If a manuscript is accepted for publication it will not be published unless the first footnote explicitly notes where the data used in the study can be obtained for purposes of replication and should note any sources that funded the research."*

No policy at *APSR*

2SLS in 60 Seconds or Less

$$Y_i = \alpha + \beta \cdot X_i + u_i$$

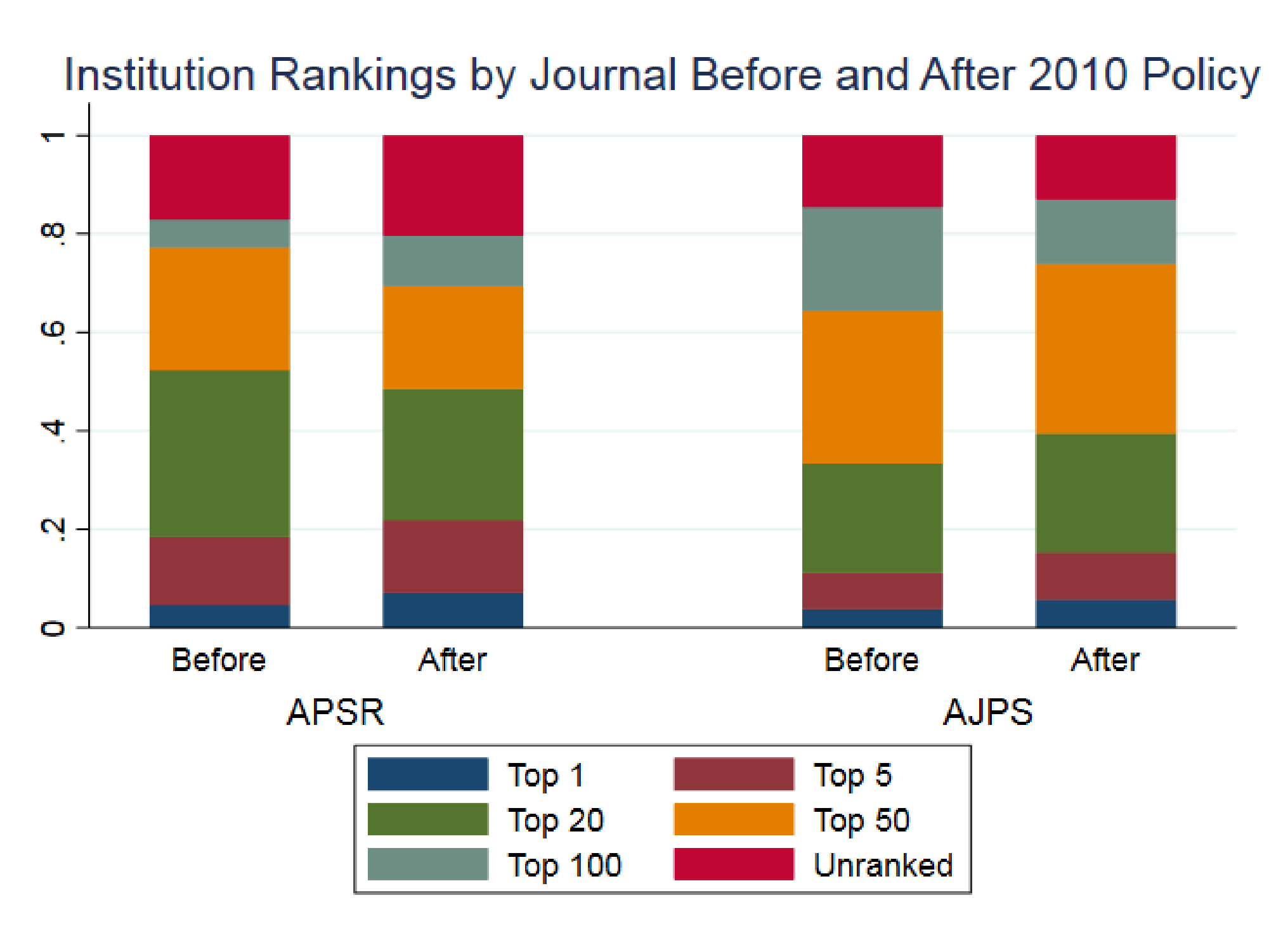
Problem: X of interest is endogenous. (Omitted variable bias–some unobservable is correlated with Y and X of interest.)

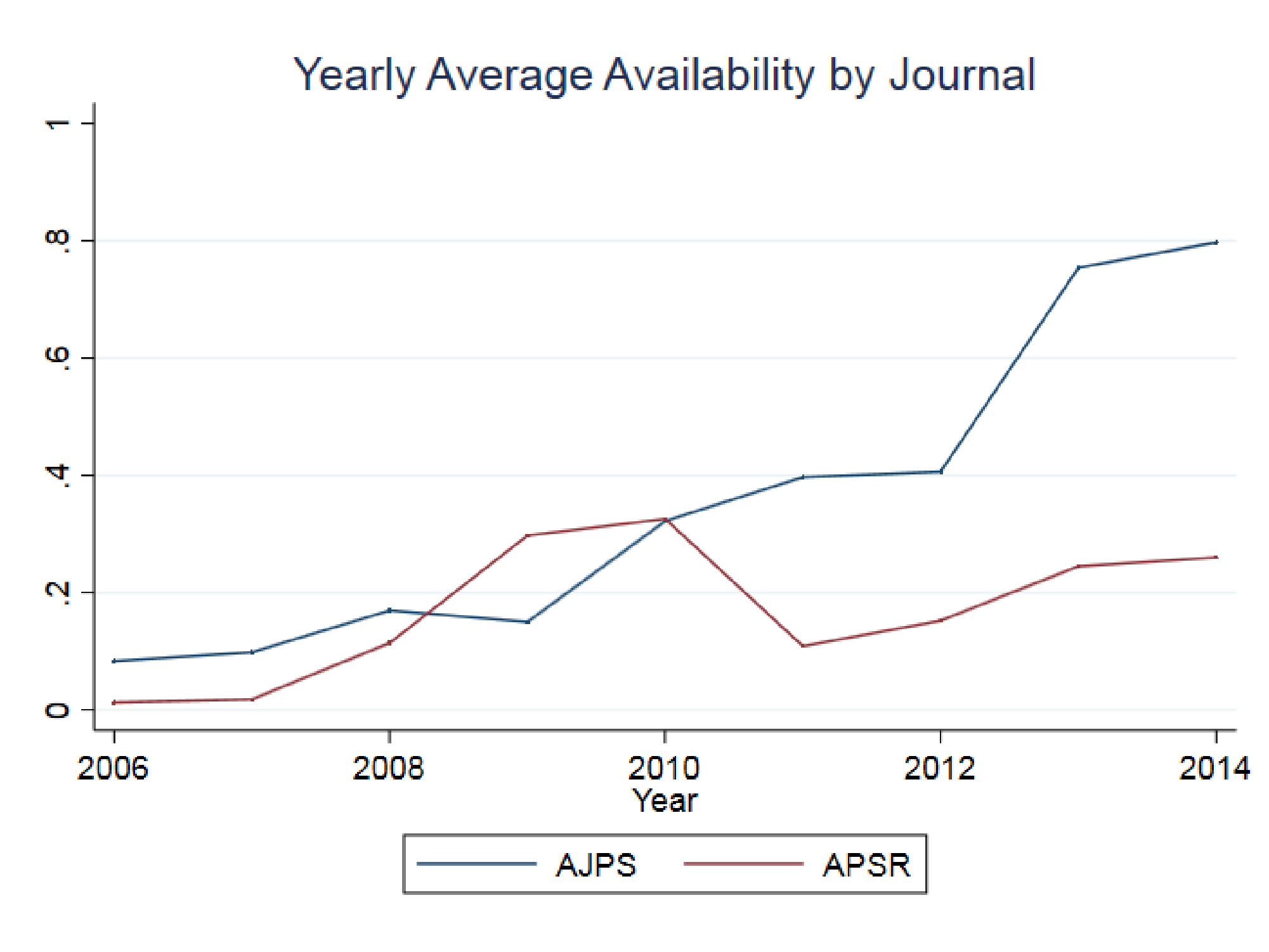
Solution: Find an instrument–something that predicts X but doesn't have the endogeneity problem itself.

$$corr(Z, u) = 0$$

$$corr(Z, X) \neq 0$$

Biased, but consistent. (Get large N!)


Institution Rankings by Journal Before and After 2010 Policy
0
.2
.4
.6
.8
1
Before
After
Before
After
APSR
AJPS
Top 1
Top 5
Top 20
Top 50
Top 100
Unranked


Yearly Average Availability by Journal
0
.2
.4
.6
.8
1
2006
2008
2010
2012
2014
Year
AJPS
APSR

Preliminary Conclusions

- Top political science papers with public data are cited more (11-13 cites, 30-45%).
- Journal policy does not appear to have changed submissions.
  - IV identification strategy OK.
- Underpowered for question of causality, so we're adding econ and other disciplines.