Data Sharing and Citations

Causal Evidence from Political Science and Economics

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

May 8, 2018 BIDS

PRELIMINARY–Please do not cite.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

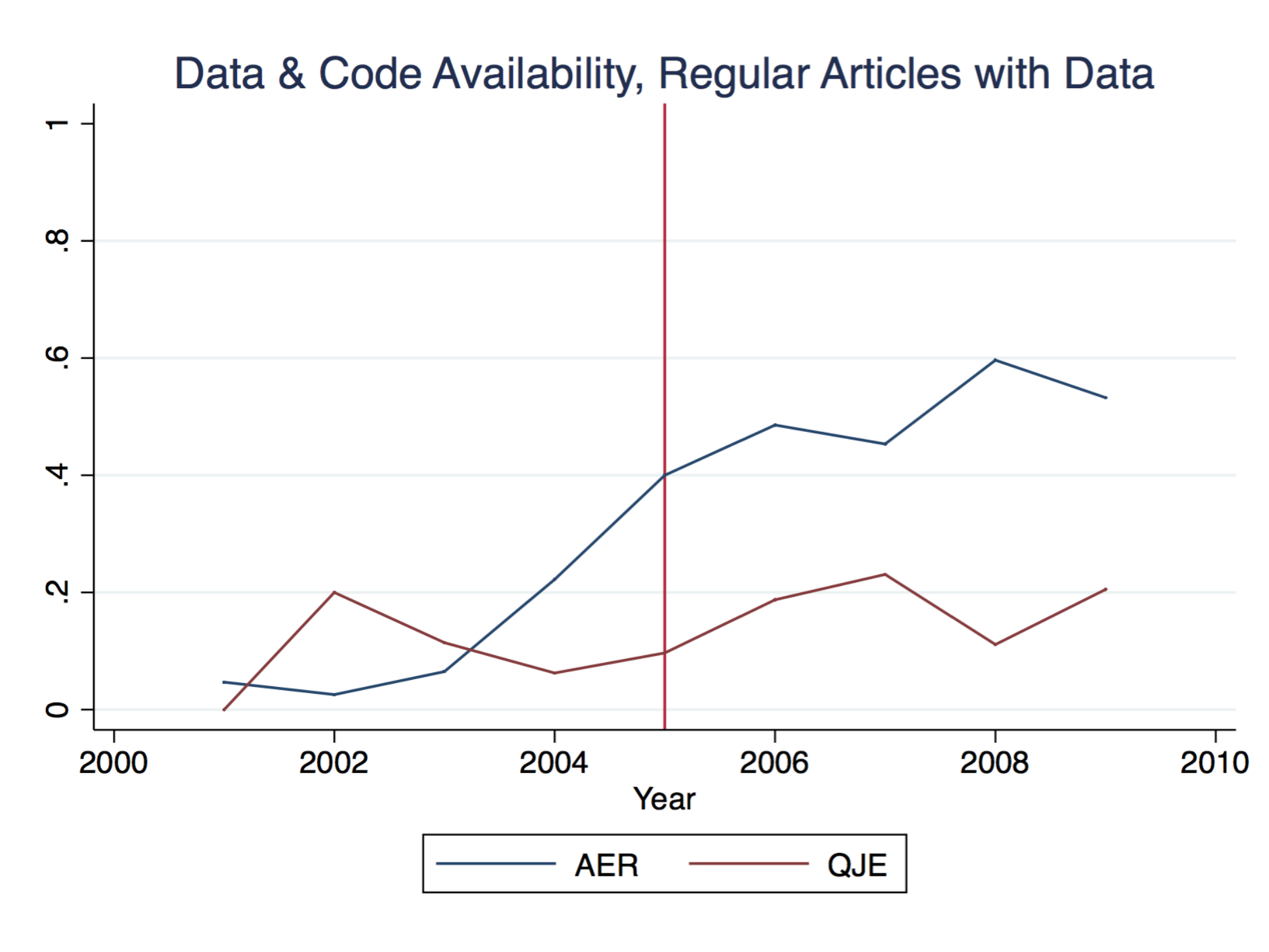
Evidence

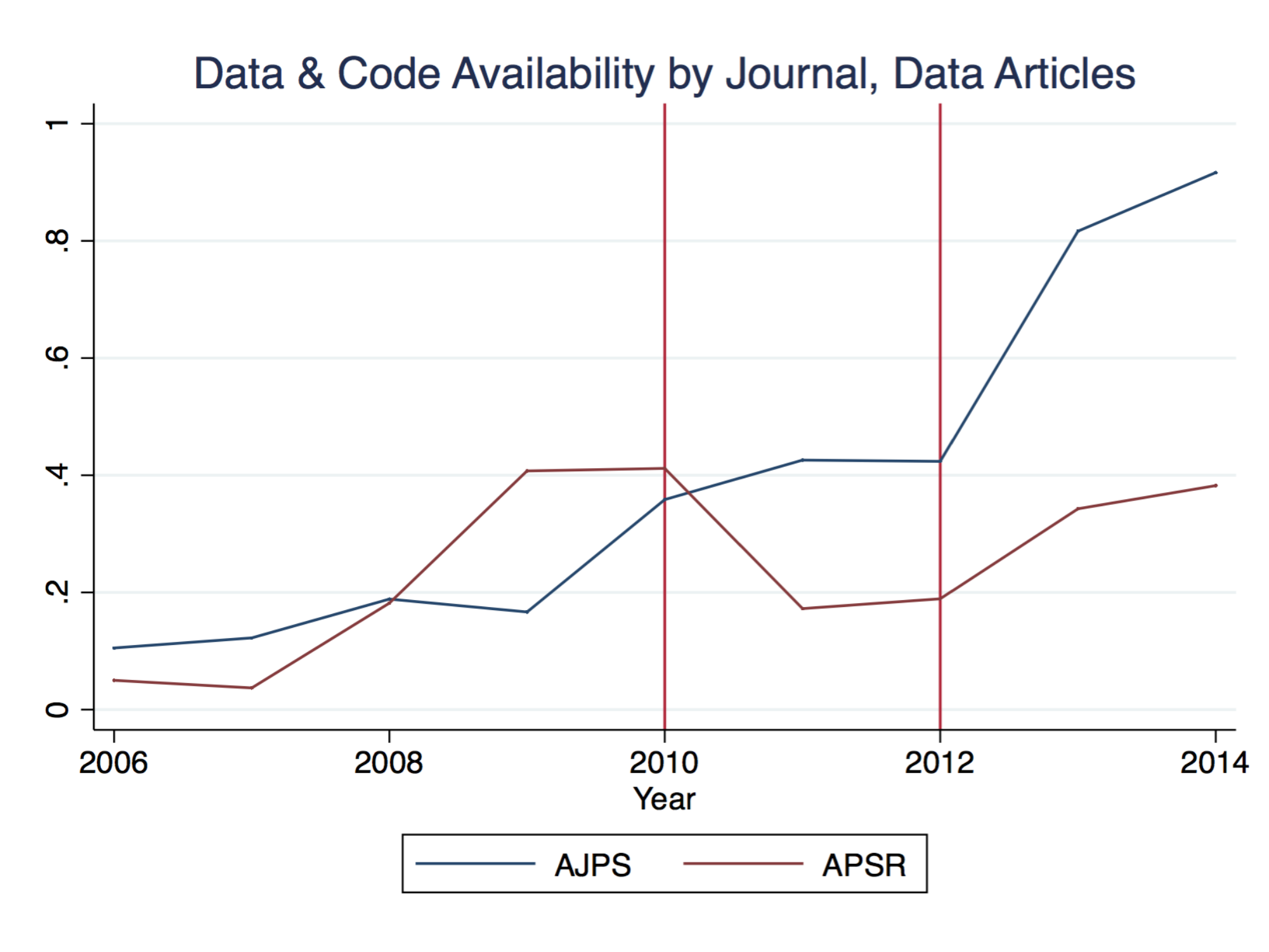
Existing Evidence: Associations

- 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).

Our Research: Causal Inference

- Sharp changes in journal policies as instrumental variables


Data & Code Availability, Regular Articles with Data
0
.2
.4
.6
.8
1
2000
2002
2004
2006
2008
2010
Year
AER
QJE


Data & Code Availability by Journal, Data Articles
0
.2
.4
.6
.8
1
2006
2008
2010
2012
2014
Year
AJPS
APSR

Preliminary Conclusions

- Top econ & political science papers with public data & code are cited more (35-50 cites, 30-45%).
- Journal policy does not appear to have changed submissions.
  - IV identification strategy OK.
- Tentative: IV estimates, which can reduce attenuation bias, even larger.