The Significance of Data-Sharing Policy

Zohid Askarov, Anthony Doucouliagos, Hristos Doucouliagos, T.D. Stanley

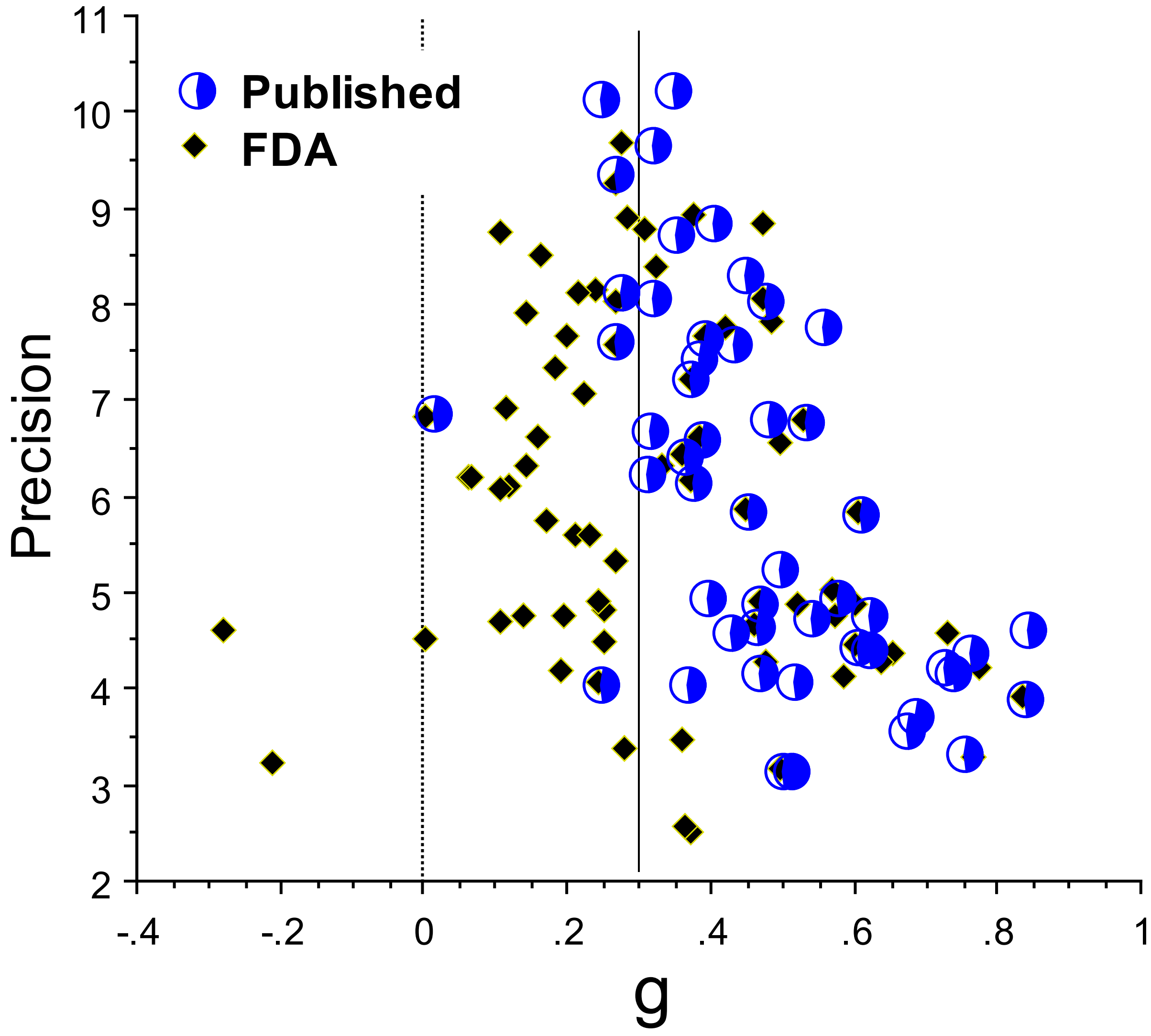
Journal of the European Economic Association

Motivation

Serious concerns regarding credibility of economics research:

- Selection for statistical significance
- Exaggeration of economic effects
- Low statistical power
- Falsely positive reported results

Antidepressant Trials



Credibility is an issue in *all* science, not just economics.

74 effect sizes from randomized clinical trials of antidepressants; the gold standard. All registered with the US FDA, because that is the law (black diamonds). However, only selected results get published in the medical journals (the blue half moons).

Many of the *published* trials report larger effects than the same experiment reported to the FDA.

Source: Turner et al. New England J. Med (2008).

Collective action problems

"Researchers are human" (Martin Paldam, 2018)

"People are people" (Gordon Tullock, 2022)

Individual researchers unlikely to take actions that are costly for themselves but beneficial to society, e.g.:

- reporting unbiased estimates
- making data available

Solutions to collective action problems

- Change norms
- Change formal rules

This Paper

The impact of mandating data-sharing in 24 leading economics journals on:

1. statistical significance: *t*-statistics
   - Observed in reported studies

2. excess statistical significance (ESS)
   - *necessary* condition for publication selection bias
   - Not observed, but can detect as a statistical trait

How might data-sharing affect reporting?

Change researcher behavior

- more careful & transparent data collection & analysis
  - e.g., fewer *accidental* errors

- reduce questionable research practices
  - e.g., *deliberate* specification searching & publication selection bias.

On the other hand ....

- Little impact if policy is not enforced

- Strong incentives to publish & exaggerate findings

- Many research design choices are independent of data-sharing.

The data

Meta-research data

- Research on research

- Inferences from many research areas

Meta-analysis

- The goal of meta-analysis is to integrate all comparable empirical estimates of a given economic phenomenon and explain their variation.
- Collect a whole body of research
    - What does the whole literature say about a specific issue?
- Take individual studies and create standardized parameters. These can be then compared.

Meta-analysis

- Distribution of *t*-statistics across different research areas cannot be expected to be the same at different journals regardless of editorial policies involved.

- Impact of journal policies may be more clearly identified if the research topic investigated and associated subject area are held constant.

- Meta-analysis offers a rich panel structure that allows to control for research subject area fixed effects. By controlling for these critical research dimensions, we eliminate at least some of the sources of variation across journals and time.

The meta-research data

Number of estimates	Number of research areas (meta-analyses)	Number of papers
166,924	359	14,947
20,121 (top 24 eco journals up to 4 years post)	345	1,913

Top five journals

	Number of estimates (1)	Number of research areas (meta-analyses) (2)	Year data-sharing mandated (3)
American Economic Review	2383	117	2005
Journal of Political Economy	872	45	2006
Quarterly Journal of Economics	585	58	2016
Review of Economic Studies	209	12	2010
Econometrica	171	21	2004

Non-top five general interest journals

	Number of estimates (1)	Number of research areas (meta-analyses) (2)	Year data-sharing mandated (3)
Review of Economics and Statistics	2642	86	2010
European Economic Review	1542	89	2012
Economic Journal	989	80	2012
Journal of the European Economic Association	135	19	2011

Tier A field journals

	Number of estimates (1)	Number of research areas (meta-analyses) (2)	Year data-sharing mandated (3)
Journal of Development Economics	2796	90	2014
Journal of Public Economics	1326	57	Nm
Journal of Finance	1193	34	Nm
Journal of Financial Economics	1001	35	Nm
Journal of Monetary Economics	944	32	Nm
Journal of Money, Credit, and Banking	834	29	1998
Public Choice	821	31	Nm
Journal of Human Resources	607	23	1990
Journal of Labor Economics	570	28	2009
Health Economics	533	19	Nm
Journal of Economic Growth	425	21	2013
Journal of Business and Economic Statistics	300	11	2011
Journal of Health Economics	192	19	Nm
Journal of Econometrics	187	13	Nm
Journal of Industrial Economics	145	17	Nm

Some patterns

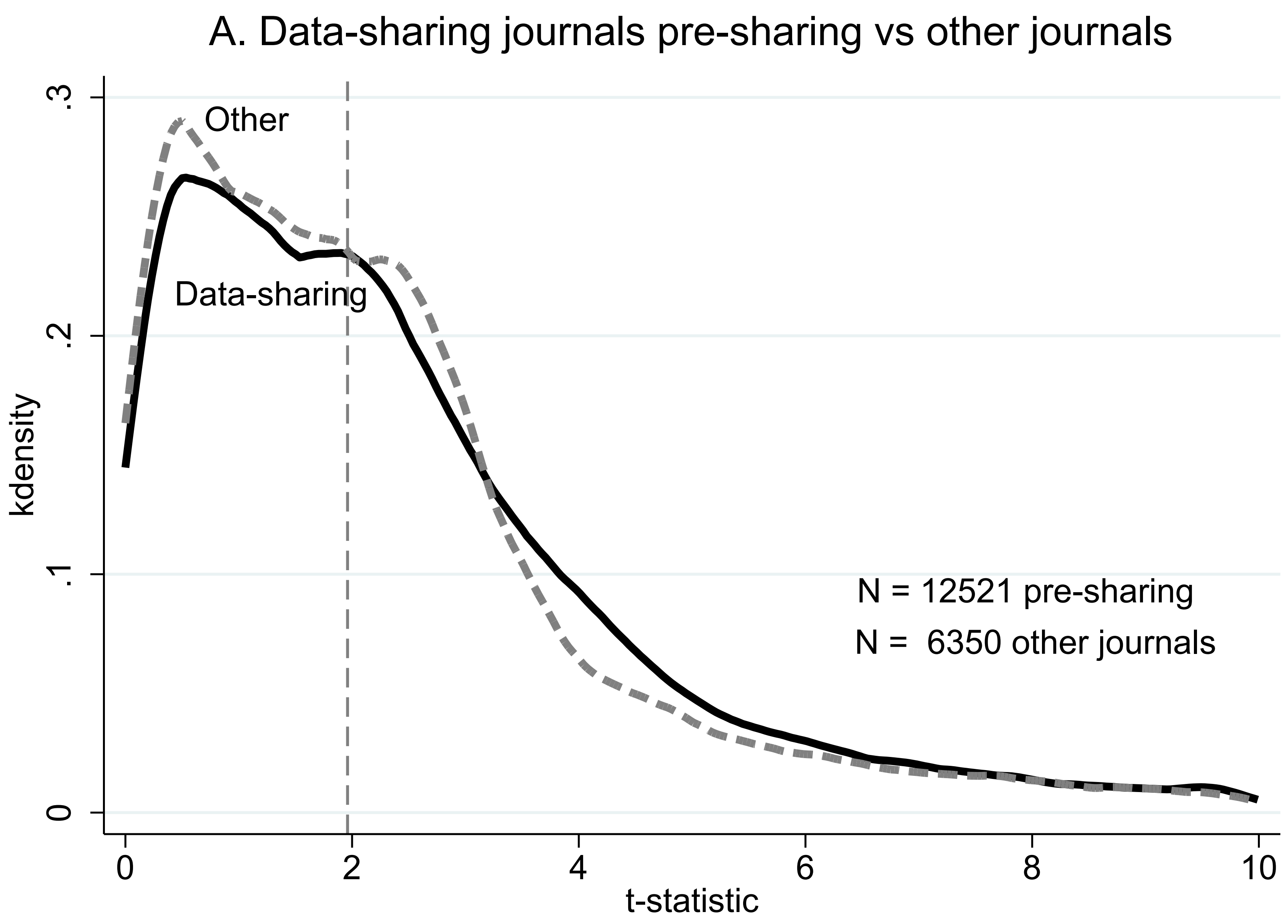
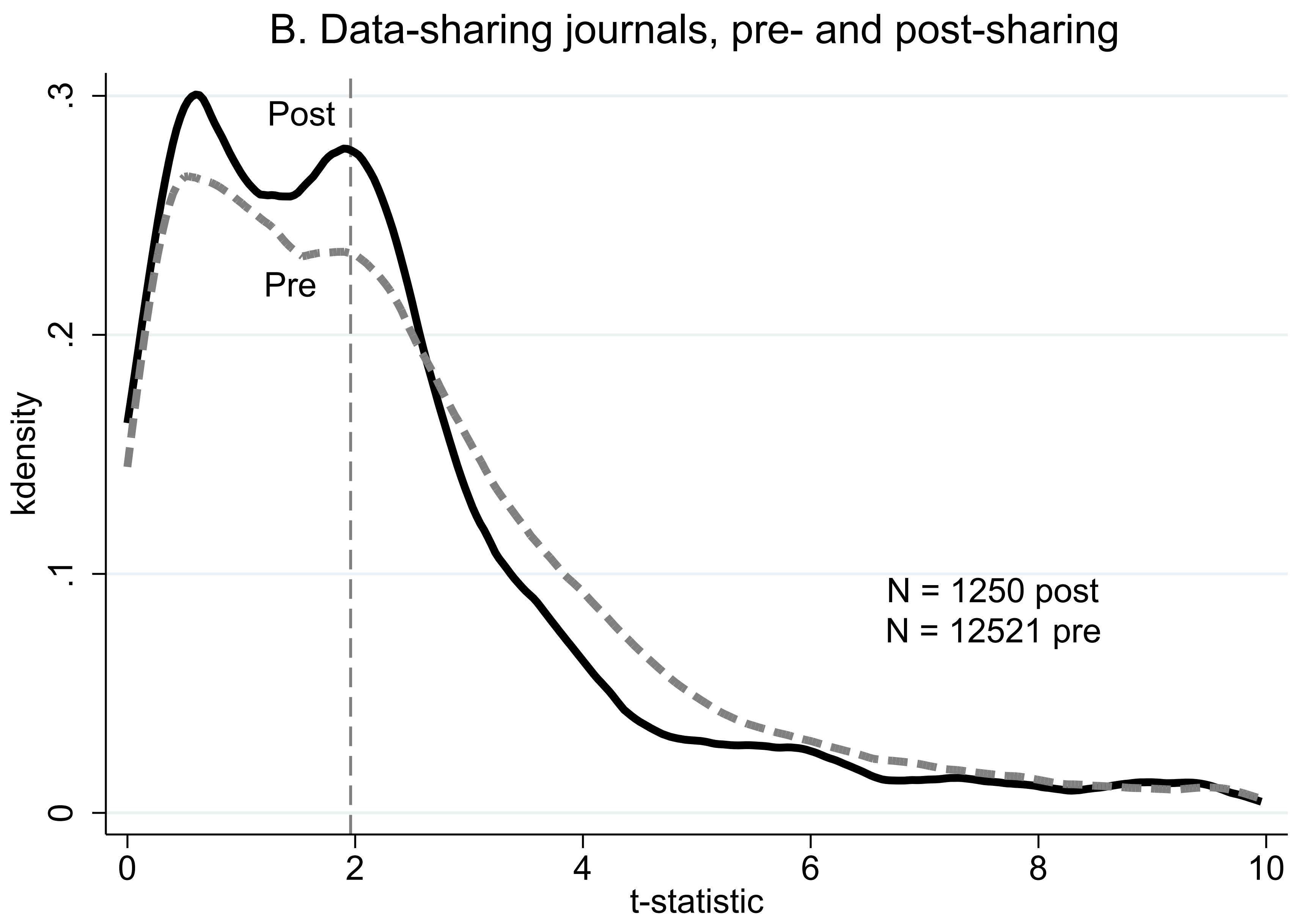
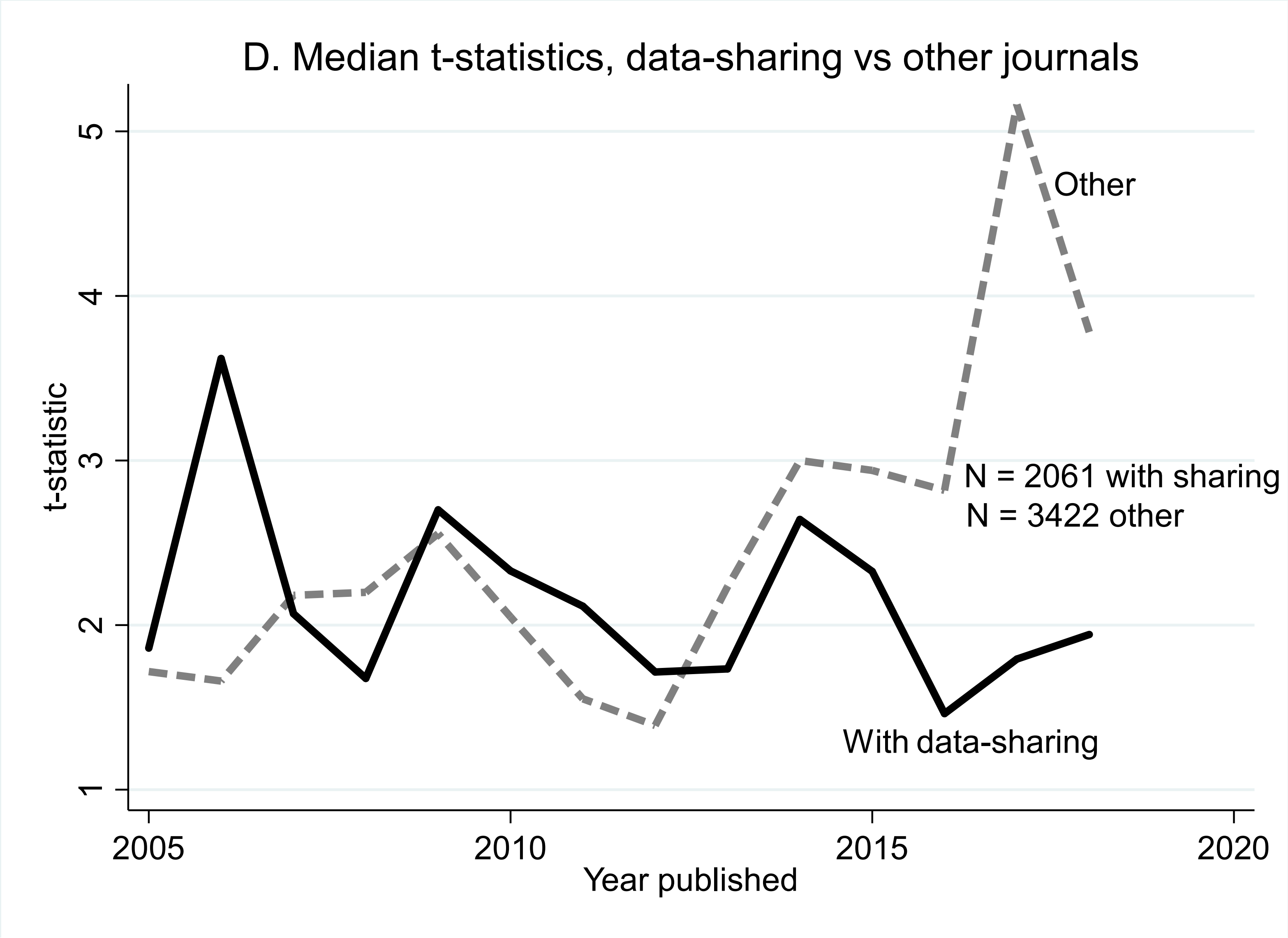


Figure compares data-sharing journals pre-data-sharing to control journals (without mandated data-sharing).



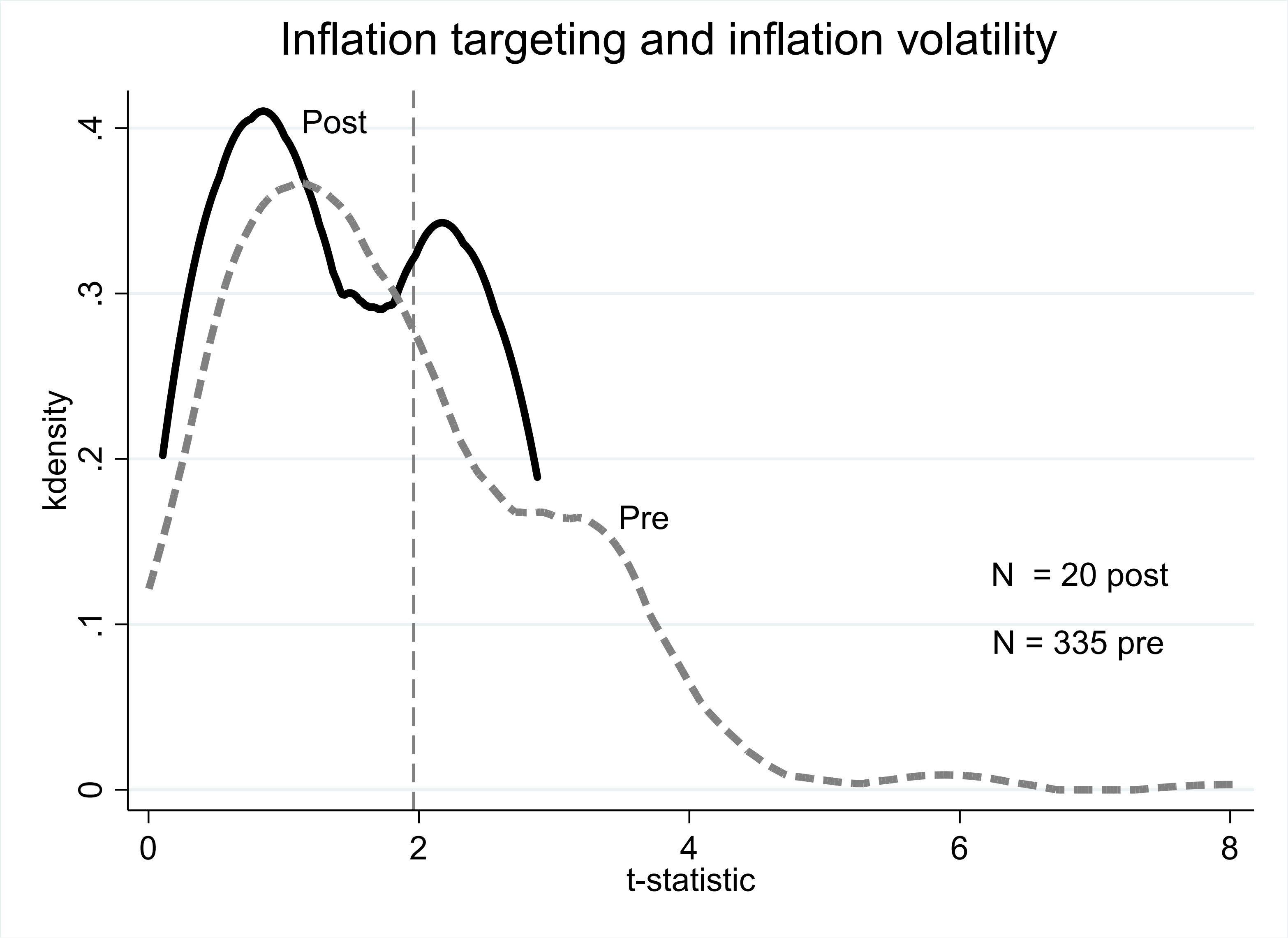
Double hump

Figure compares distributions for data-sharing journals, pre- and post-data-sharing.



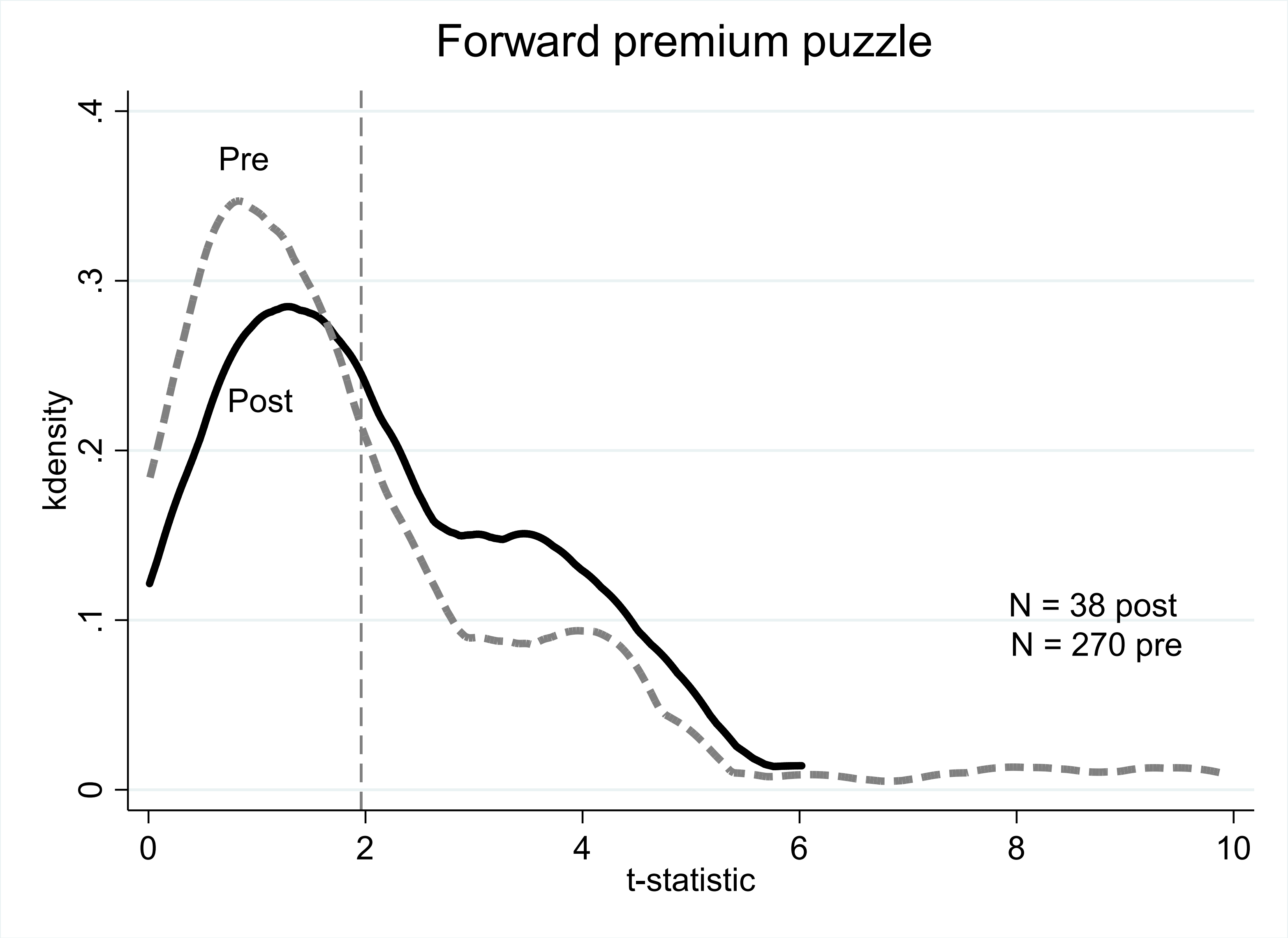
Divergence.
General equilibrium effects?
Spillover?

Figure compares the annual median *t*-statistic for journals with data-sharing, post-data-sharing, to control journals, for the period 2005-2018.

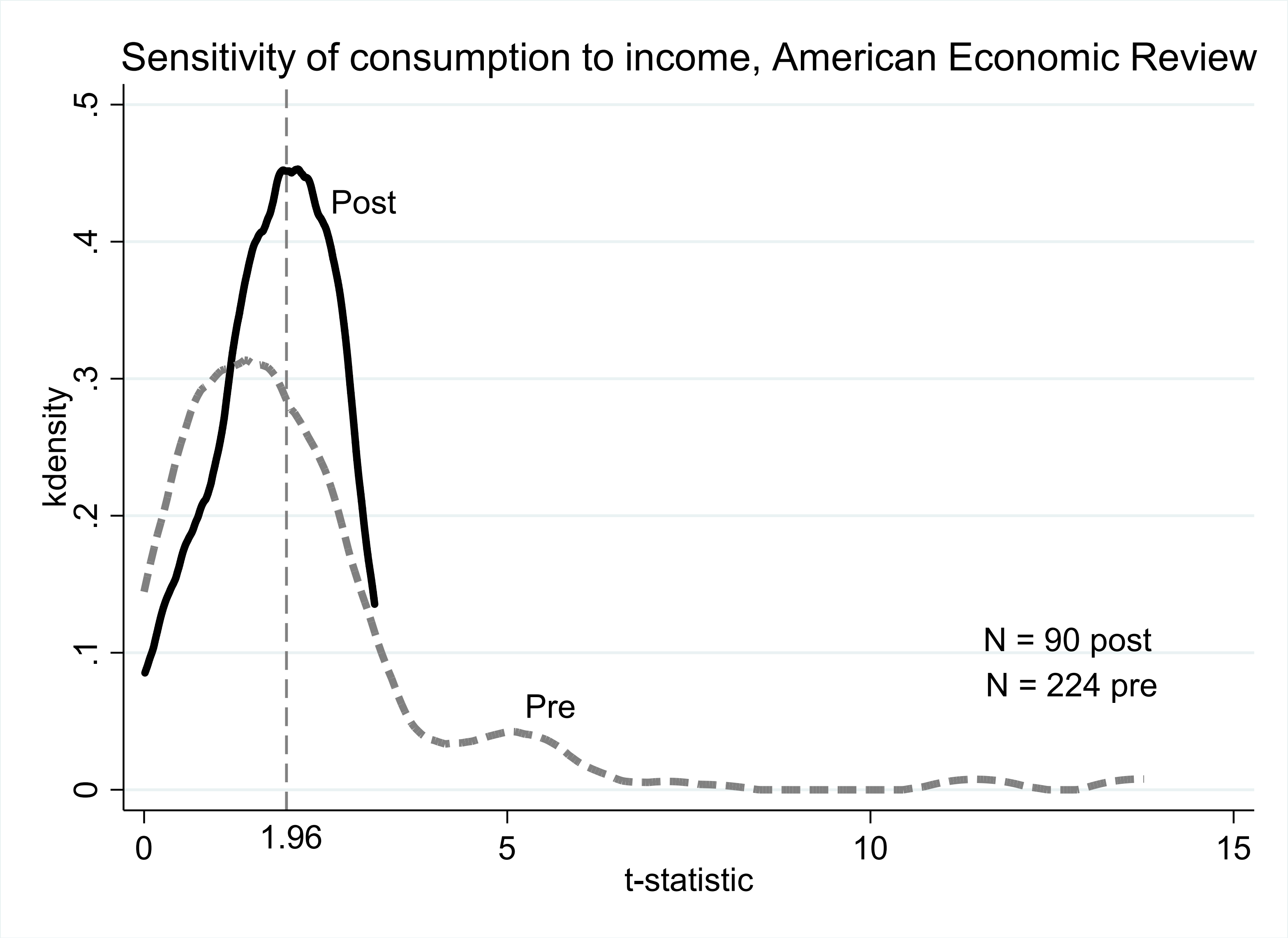


Shorter tails

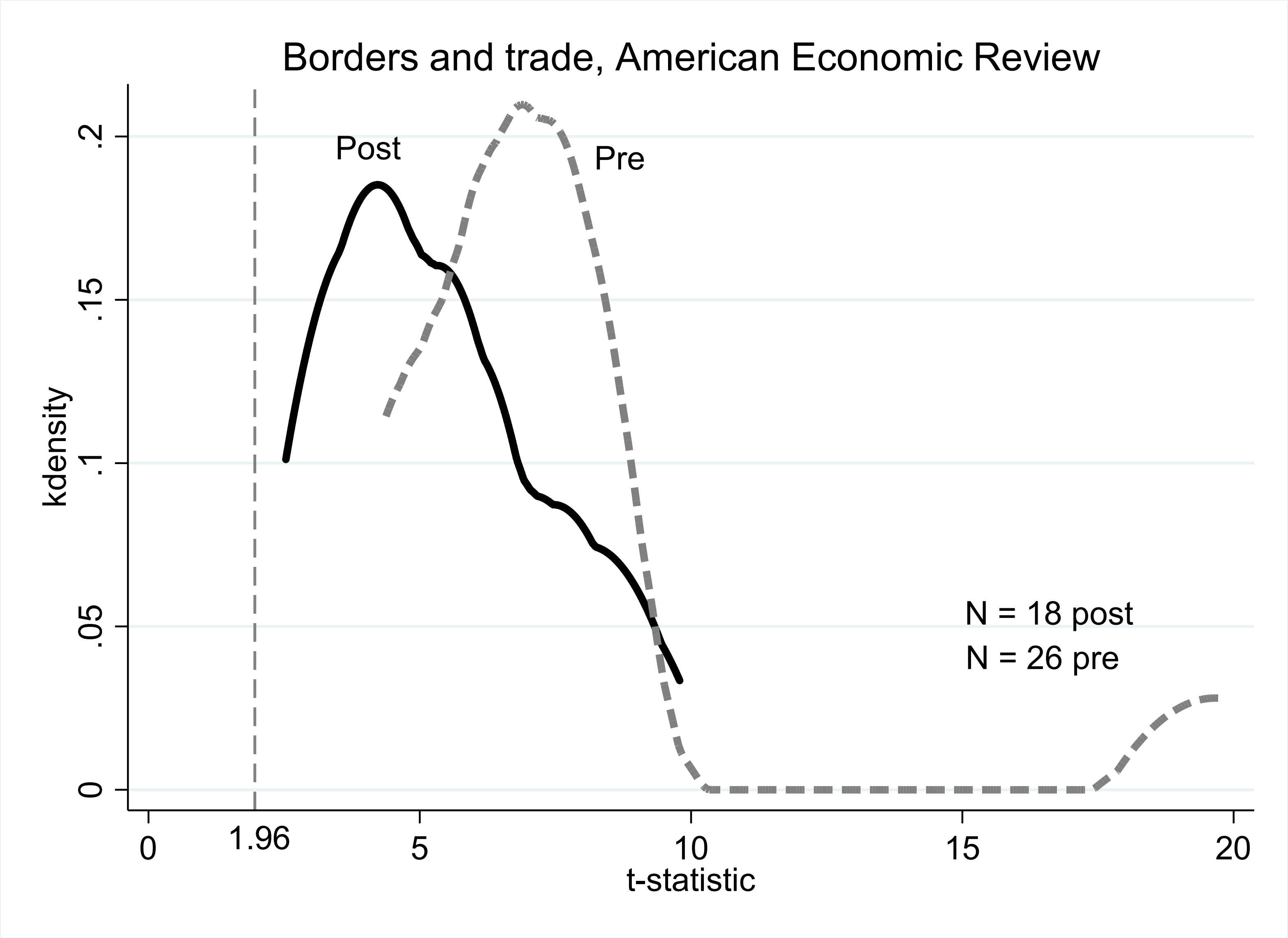
Distribution of test statistics for inflation targeting research area, pre- and post-data-sharing. Dashed vertical line denotes the 1.96 threshold.



Distribution of test statistics for forward premium research area, pre- and post-data-sharing. Dashed vertical line denotes the 1.96 threshold.

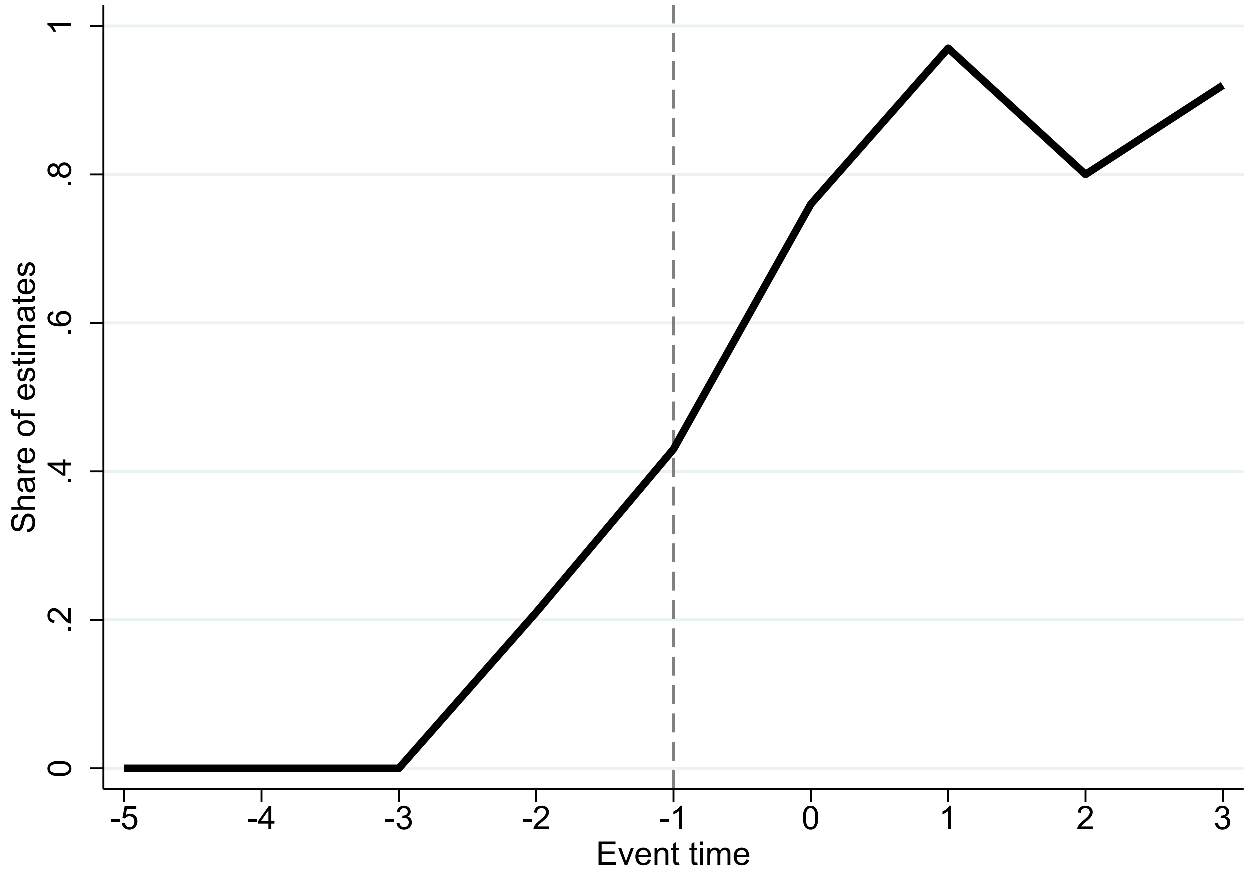


Distribution of test statistics for sensitivity of consumption to income research area, pre- and post-data-sharing, AER estimates. Dashed vertical line denotes the 1.96 threshold.



Distribution of test statistics for borders and trade research area, pre- and post-data-sharing, AER estimates. Dashed vertical line denotes the 1.96 threshold.

Share of estimates reported with data available



Policy anticipation?

Notes: Black line denotes share of estimates with data made available. We use only studies with a known submission date. Event time 0 denotes the year data-sharing mandated. Event time 1 denotes the prior year and indicates anticipation.

Empirical strategy

- **Standard static DD study:**

$$Y_{ijt} = \beta_0 D_{jt} + \boldsymbol{x}_{ijt}\gamma + \alpha_j + \alpha_t + \alpha_m + \varepsilon_{ijt},$$

- **Event study (dynamic):**

$$Y_{ijt} = \alpha_j + \alpha_t + \alpha_m + \sum_{y=-5}^{-1} \beta_y \, D_j \cdot Yr_y + \sum_{y=1}^{4} \delta_y \, D_j \cdot Yr_y + \boldsymbol{x}_{ijt}\gamma + \varepsilon_{ijt},$$

Difference-in-differences (DD) analysis

- Allowing for:
  - staggered introduction of mandated data-sharing
    - Journals adopt data-sharing at different points in time
  - heterogeneous treatment effects
  - anticipation of policy change.

Estimation

- OLS
  - all data (biased) Goodman-Bacon (2018), Sun and Abraham (2021)
  - ‘stacked’, comparable data careful choice of ‘control’ untreated journals.

- New DD imputation estimator of Borusyak *et al.* (2022, *RES*)

Excess statistical significance (ESS)

- Difference between observed statistically significant results & statistically significant results expected based on statistical power & in the absence of selective reporting.

- If there is publication selection, then ESS results *must* have been produced (Stanley et al. 2021)

Excess statistical significance (ESS) for estimate *i*, in journal *j*, at time *t*, and research area *m* is:

$$ESS_{ijmt} = SIG_{ijmt} - ESig_{ijmt},$$

where SIG_{ijmt} (0/1) denotes whether a reported estimate is statistically significant or not.

Following Stanley et al. (2021), we calculate *ESig* as:

$$ESig_{ijm} = 1 - \Phi\left(Z_{ijmt}\right),$$

where $\Phi\left(Z_{ijm}\right)$ denotes the cumulative standard normal probability &

$Z_{ijmt} = (1.96 * SE_{ijmt} - \left|\hat{\delta}_m\right|)/\sqrt{SE_{ijmt}^2 + \hat{\tau}_m^2}$, *SE* denotes the standard error for the i[th] estimate $of\ \delta_m$, the mean effect for research area *m,* and $\hat{\tau}_m^2$ is the estimated heterogeneity variance for each area of research.

Estimating the mean of the distribution

1. UWLS (Stanley & Doucouliagos, 2017)
   - conservative estimate of the mean

2. Robustness: FAT-PET-PEESE conditional estimator (Stanley & Doucouliagos, 2004)
   - correction for selection (less conservative)

Controls

- Fixed effects:
    - Journal
    - time (year submitted or year published)
    - research area fixed effects
- Type of research (observational, experimental)
- Number of co-authors (Ioannidis 2012; Brodeur et al. 2016; Fanelli et al. 2017)
- Temporal rank (Ioannidis 2005)
- Editorial decision makers (number of editors and change in editorial board)

Results

Sample means
(data-sharing journals)

Outcome	Pre-data-sharing	Post-data-sharing
t-statistics	3.55	2.95
		-17%
Excess statistical significance (ESS)	0.25	0.19
		-26%

With policy anticipation

	No fixed effects or covariates	Plus journal and time fixed effects	Plus field of research effects	Plus covariates	Plus journal trends
	(1)	(2)	(3)	(4)	(5)
t-value (1)	-0.539	-1.311	-1.055	-1.109	-1.549
	(0.265)	(0.458)	(0.452)	(0.483)	(0.695)
Pre-trend test	0.063	0.695	0.509	0.514	0.734
ESS (2)	-0.084	-0.052	-0.076	-0.090	-0.197
	(0.027)	(0.044)	(0.055)	(0.058)	(0.064)
Pre-trend test	0.620	0.157	0.454	0.490	0.379
N	20,121	20,121	19,946	19,946	19,946

Notes: Dependent variable is the absolute value *t*-statistic & ESS in Rows (1) & (2). Each cell reports the ATT from the BJS DD imputation estimator. Clustered standard errors at the journal article level reported in parentheses. Model allows for anticipation of policy change one year prior to mandatory data-sharing. The pre-trend test reports the *p*-values of the joint statistical significance of seven pre-trend coefficients.

Impact of data-sharing

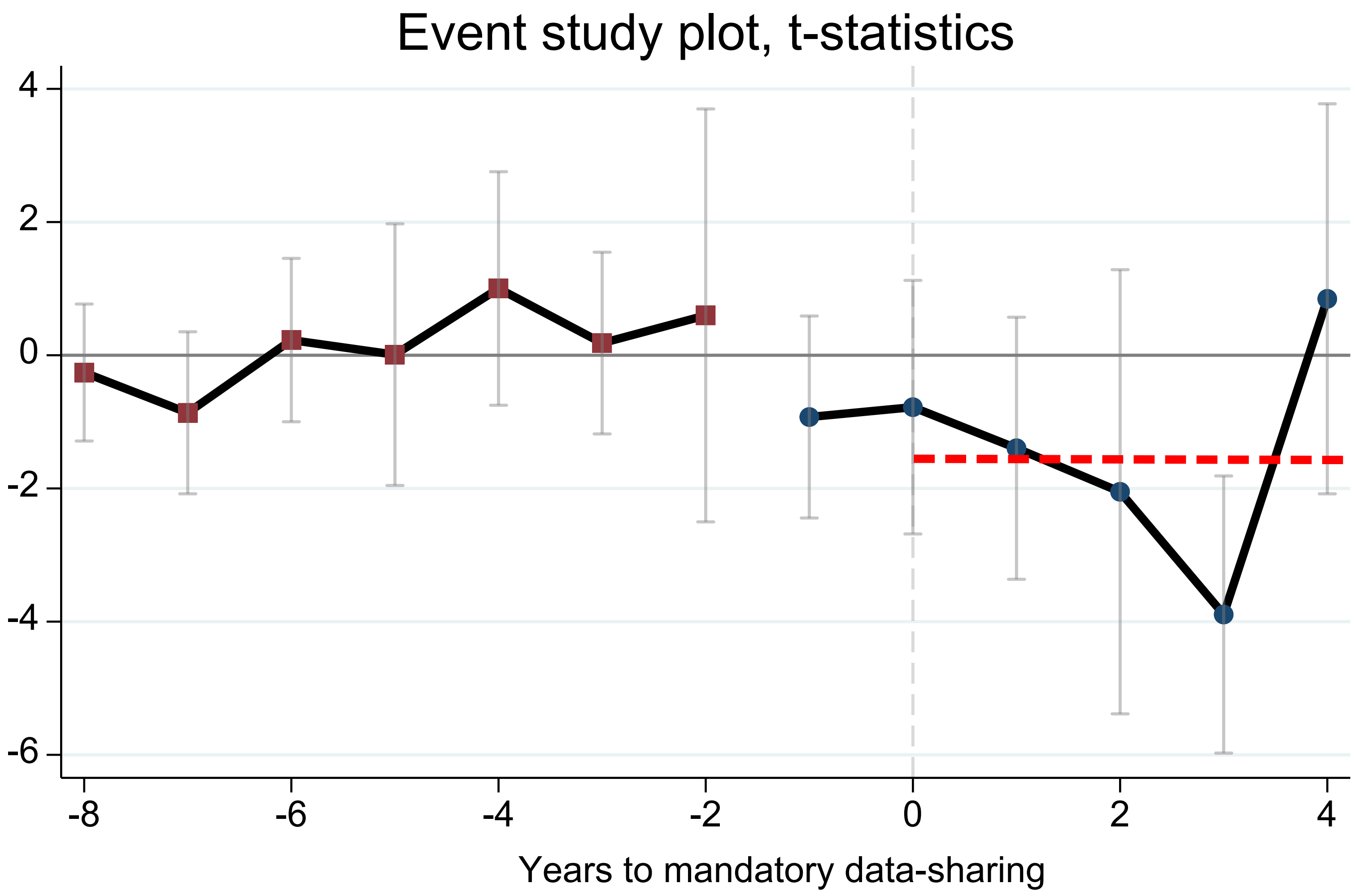
- 31% decrease in reported *t*-values
- ESS fell by 36% (but this may not be causal)

Robustness – subsamples

	Top five journals	Non-top five journals	Without JHR & JMCB	Post-1999	Pre- & post- research areas	Balanced panel	Macro research
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
t-value (1)	-0.705	-1.889	-2.270	-2.044	-4.517	-4.381	-1.193
	(1.071)	(0.808)	(0.797)	(1.014)	(1.001)	(0.867)	(0.720)
Pre-trend test	0.074	0.379	0.271	0.040	0.008	0.346	0.056
ESS (2)	0.254	-0.240	-0.141	-0.159	-0.275	-0.200	-0.009
	(0.088)	(0.099)	(0.079)	(0.099)	(0.090)	(0.072)	(0.067)
Pre-trend test	0.159	0.000	0.033	0.000	0.001	0.079	0.057
Number of observations	10,822	16,183	19,273	11,814	9,159	12,957	12,155

Notes: Dependent variable is the absolute value t-statistic & ESS in Rows (1) & (2). Each cell reports the ATT from the BJS DD imputation estimator. Clustered standard errors at the journal article level reported in parentheses. Model allows for anticipation of policy change one year prior to mandatory data-sharing. The pre-trend test reports the p-values of the joint statistical significance of seven pre-trend coefficients.

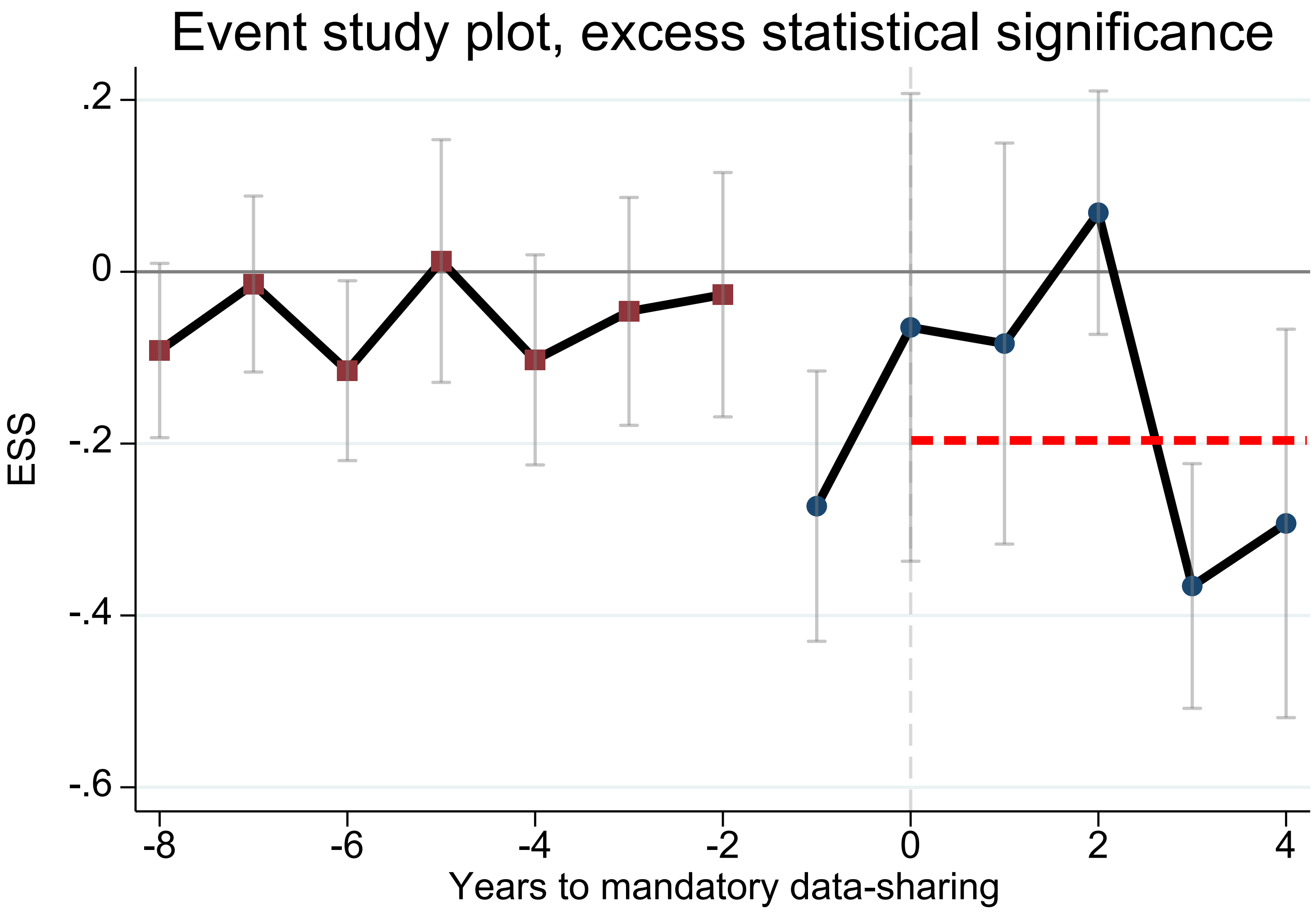
Event study dynamic effects



Pre-trends p-value = 0.737

Lags p-value = 0.038

Notes: Black bold lines graph the event study coefficients for years -1 to 4 and pre-trends coefficients for years -2 to -8. Vertical bars are 90% confidence intervals. Data-sharing introduced in event year 0. Event year -1 allows for anticipation. Event year 4 affected by journal compositional effects. Dashed line denotes the static DD coefficient. Pre-trends *p*-value is a joint test for all seven pre-trends coefficients. Lags *p*-value is a joint test for all data-sharing coefficients.



Pre-trends p-value = 0.378

Lags p-value = 0.000

Notes: Black bold lines graph the event study coefficients for years -1 to 4 and pre-trends coefficients for years -2 to -8. Vertical bars are 90% confidence intervals. Data-sharing introduced in event year 0. Event year -1 allows for anticipation. Event year 4 affected by journal compositional effects. Dashed line denotes the static DD coefficient. Pre-trends *p*-value is a joint test for all seven pre-trends coefficients. Lags *p*-value is a joint test for all data-sharing coefficients.

Channels

	t-statistic < |5|	t-statistic < |10|	t-statistic < |20|	All journals	Top five journals	Non-top five journals
	(1)	(2)	(3)	(4)	(5)	(6)
t-value (1)	-0.593 (0.244)	-1.138 (0.332)	-1.562 (0.495)			
Excess statistical significance (2)	-0.060 (0.089)	-0.084 (0.079)	-0.107 (0.076)			
Significant (3)				-0.201 (0.072)	0.151 (0.084)	-0.347 (0.093)
Barely significant (4)				0.059 (0.045)	0.067 (0.044)	0.037 (0.054)
Number of observations	17,142	18,941	19,609	19,965	10,822	16,183

Notes: Each cell reports the ATT from the BJS DD imputation estimator. The dependent variable is the *absolute* value of the reported *t*-statistic in row (1) and ESS in row (2), respectively. Columns (1), (2), and (3) limit the sample to estimates with *t*-statistics less than |5|, |10|, and |20|, respectively. Columns (4)–(6) report results from a linear probability model. Rows (3) and (4) report results where the dependent variable is whether an estimate is reported to be statistically significant at the 5% level and "barely" statistically significant (*t*-statistic between 1.96 and 2.58), respectively. Clustered standard errors at the journal article level are reported in parentheses.

Comparison of estimates (no anticipation)

	t-statistics	ESS
OLS – biased	-0.462 (1.150)	-0.060 (0.065)
OLS – ‘stacked data’	-1.325 (1.002)	-0.139 (0.093)
OLS – ‘stacked data’ (with endogenous sampling)	-2.612 (1.097)	-0.185 (0.080)
Imputation	-1.842 (0.784)	-0.110 (0.075)
Imputation (with endogenous sampling)	-3.135 (0.677)	-0.211 (0.055)

Standard errors in brackets

Some positive signs

- Changing rules (and norms) can make a difference to science:

  - Mandating data-sharing improves credibility of economics, at least in short-run

  - 8 Health economics journals, Editorial Statement on Negative Findings decreased the extent of publication bias (Blanco-Perez & Brodeur *EJ*, 2020)

No improvement among the top 5

- Very small reduction in *t*-statistics
- ESS *increased* post data-sharing

References

Blanco-Perez, C. & A. Brodeur. 2020. "Publication Bias and Editorial Statement on Negative Findings." *Economic Journal,* 130, 1226–1247.

Borusyak, K., X. Jaravel & J. Spiess. 2021. "Revisiting Event Study Designs: Robust and Efficient Estimation." Manuscript. May 19, 2021.

Brodeur, A., M. Lé, M. Sangnier & Y. Zylberberg. 2016. "Star Wars: The Empirics Strike Back." *American Economic Journal: Applied Economics,* 8, 1–32.

Fanelli, D., R. Costas & J.P.A. Ioannidis. 2017. "Meta-Assessment of Bias in Science." *Proceedings of the National Academy of Sciences*, 114, 3714–3719.

Goodman-Bacon, A. 2018. "Difference-in-Differences with Variation in Treatment Timing." National Bureau of Economic Research. Working Paper 25018.

Ioannidis, J.P.A. 2005. "Contradicted and Initially Stronger Effects in Highly Cited Clinical Research." *JAMA,* 294, 218–228.

Ioannidis, J.P.A. 2012. "Scientific Inbreeding and Same-Team Replication: Type D Personality as an Example." *Journal of Psychosomatic Research*, 73, 408–410.

Paldam, M., 2018. A model of the representative economist, as researcher and policy advisor. *European Journal of Political Economy,* 54, 5–15.

Stanley, T.D. & H. Doucouliagos. 2014. "Meta-Regression Approximations to Reduce Publication Selection Bias." *Research Synthesis Methods*, 5, 60–78.

Stanley, T.D. & H. Doucouliagos. 2017. "Neither fixed nor random: Weighted least squares meta-regression analysis," Res*earch Synthesis Methods*, 8, 19-42.

Stanley, T.D., H. Doucouliagos, J.P.A. Ioannidis & E.C. Carter. 2021. "Detecting Publication Selection Bias Through Excess Statistical Significance." *Research Synthesis Methods,* 12, 776–795.

Sun, L. & S. Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics*, 225, 175–199.

Tullock, G., 2002. People are people: the elements of public choice. In: Tullock, G., Seldon, A., Brady, G.L. (Eds.), Government Failure: A Primer in Public Choice. CATO Institute, Washington DC, pp. 3–16.