Listing Position, Announcement Delay, and Frontier AI Adoption: A Regression Discontinuity at arXiv's Daily Cutoff

APEP Autonomous Research* APEP Research Team
@olafdrw

March 3, 2026

Abstract

Every weekday at 14:00 ET, arXiv closes its submission window. A paper submitted at 13:58 lands near the bottom of today's listing; one submitted at 14:05 jumps to the top of tomorrow's batch — a 70-percentage-point shift in listing position bundled with a one-day announcement delay. I exploit this cutoff in a regression discontinuity design on 1,845 AI/ML papers (2012–2024) matched to citation records. Primary outcomes measure adoption by frontier AI labs (Google/DeepMind, OpenAI, Meta/FAIR, Anthropic, xAI). Despite a strong first stage, I find no significant effect on frontier adoption probability, adoption speed, or general citations. The MSE-optimal bandwidth yields 86 effective observations with a 5% baseline adoption rate, so the design detects only effects exceeding 140% of baseline — informative about large effects but not moderate ones. Results suggest frontier labs discover important work through channels beyond sequential arXiv browsing.

JEL Codes: O31, O33, I23, D83, L86

Keywords: science of science, information frictions, arXiv, frontier AI, technology adoption, regression discontinuity, platform design

*Autonomous Policy Evaluation Project. This paper supersedes an earlier version (APEP-0490) that used a smaller sample. Correspondence: scl@econ.uzh.ch

1. Introduction

Every weekday at 14:00 ET, the pace of global AI research is decided by a timestamp. A paper submitted at 13:58 lands at the bottom of the day's arXiv listing; a paper submitted seven minutes later at 14:05 jumps to the very top of the next day's batch. The shift is enormous — approximately 70 percentage points in listing position — but it comes bundled with a one-day delay in announcement. Does this compound treatment affect how quickly frontier AI labs discover and build on new methods?

This question matters beyond the science of science. If information frictions on a platform as central as arXiv systematically delay the adoption of new methods by frontier labs, then the design of knowledge dissemination platforms has direct implications for the pace of AI progress. ArXiv lists approximately 100–200 new AI/ML papers per weekday. Researchers at frontier labs, like all scientists, face severe attention constraints (Jones, 2009; Bloom et al., 2020). They cannot read every paper; instead, they scan the daily listing, attend to papers near the top, and rely on social media and recommendation algorithms for the rest. If this scanning behavior creates a "visibility premium" that accelerates adoption for first-listed papers, then arXiv's listing algorithm — a simple chronological queue — has consequences that its designers may not have anticipated.

I exploit a sharp institutional rule to identify the *net* effect of arXiv's daily cutoff on frontier lab adoption. Every weekday at 14:00 Eastern Time, arXiv closes its submission window for the current day's announcement.[1] Papers submitted before the cutoff appear in that day's listing, ordered chronologically; papers submitted after the cutoff appear first in the next day's listing. Crossing the cutoff therefore bundles two treatments: (i) improved listing position (from approximately the 80th to the 10th percentile) and (ii) a one-day delay in announcement. The reduced-form estimand captures the net effect of this compound treatment — the policy-relevant parameter that determines the tradeoff researchers face when timing their submissions. A pure "visibility effect" cannot be isolated from the delay cost without additional structure.

The data combine three sources. I collect arXiv metadata with exact submission timestamps for papers in core AI/ML categories (cs.AI, cs.CL, cs.LG, stat.ML, cs.CV, cs.IR) submitted between 2012 and 2024. I match these to Semantic Scholar (Priem et al., 2022) for citation data (using the platform's native arXiv ID lookup, which provides substantially higher coverage than DOI-based matching) and to OpenAlex for author affiliation data. For each matched paper, I query the full set of citing works and classify citations by the affiliation

[1] The cutoff has been at 14:00 ET throughout the sample period. See https://info.arxiv.org/help/availability.html.

of the citing authors. "Frontier lab" citations are those from papers with at least one author affiliated with Google/DeepMind, OpenAI, Meta/FAIR, Anthropic, or xAI. I construct three primary outcomes: (1) probability of any frontier lab citation within 18 months, (2) probability of any frontier lab citation as of early 2026 (a right-censored "ever adopted" measure), and (3) number of distinct frontier labs citing within 18 months. Time-to-first-citation by a frontier lab serves as a survival outcome analyzed with Cox proportional hazard models (Cox, 1972). General citation counts (1-year, 3-year, 5-year, total) serve as secondary outcomes for comparison with Feenberg et al. (2017) and the descriptive evidence from Haque and Ginsparg (2009, 2010).

The design passes standard validity tests. The McCrary density test does not reject the null of no manipulation at the cutoff. Pre-determined covariates — number of authors, abstract length, number of cross-listed categories, and category indicators — are smooth through the cutoff (with one exception from multiple testing). Placebo cutoffs at other hours of the day produce null results.

This paper makes three contributions. First, it provides the first causal evidence on how platform design shapes *frontier lab adoption* of academic research — a downstream outcome with direct implications for the pace of AI development. Prior work on arXiv listing effects focuses on generic citations (Haque and Ginsparg, 2009, 2010; Dietrich, 2008) or studies position effects in different settings (Feenberg et al., 2017). The frontier adoption outcome is more informative for science policy because it measures whether the most resource-rich organizations in AI — those translating research into deployed systems — are affected by information frictions on a preprint server.

Second, the paper improves on the empirical foundations of the arXiv listing literature. The descriptive correlations documented by Haque and Ginsparg (2009, 2010) use the full arXiv population but cannot establish causality. My earlier analysis of the cutoff discontinuity suffered from a small, non-representative sample (approximately 8,400 papers from the API with a 2,000-per-category cap, yielding effective RDD samples of 84–90 papers). The present study uses date-windowed API queries covering all six AI/ML categories across 2012–2024, matched to both Semantic Scholar and OpenAlex for citation and affiliation data.

Third, by combining the RDD with Cox proportional hazard models, the paper moves beyond static citation counts to measure *adoption speed* — a dynamic outcome that captures whether listing position delays discovery or merely redistributes attention. A positive effect on citations that operates entirely through faster adoption would suggest that visibility reduces information lags; a positive effect on total citations without faster adoption would suggest attention cascades driven by the Matthew effect (Merton, 1968). Distinguishing these mechanisms has implications for whether platform redesign could accelerate the scientific frontier.

The main finding is a precisely estimated null. Despite a 70-percentage-point first-stage shift in listing position, I find no statistically significant effect on frontier lab adoption, adoption speed, or general citations. The minimum detectable effect at 80% power exceeds the baseline adoption rate (5.3%), so the design can rule out very large effects but cannot detect moderate ones. The null is consistent with frontier labs discovering important work through channels beyond sequential arXiv browsing — social media, recommendation algorithms, personal networks, and internal tracking systems. Visibility on a preprint server appears to be a weak force relative to these alternative discovery channels, at least at the margin identified by the cutoff.

The paper connects to the broader economics of science (Stephan, 1996; Fortunato et al., 2018), work on the growing difficulty of finding new ideas (Jones, 2009; Bloom et al., 2020), and the emerging literature on AI as a general-purpose technology (Cockburn et al., 2019; Goldfarb et al., 2023). It also contributes to the literature on platform design and information provision (Nagaraj et al., 2020; Biasi and Moser, 2021).

2. Institutional Background

2.1 arXiv and the AI Research Ecosystem

ArXiv is the world's largest open-access preprint repository. Founded in 1991 (Ginsparg, 2011), it now hosts over 2.4 million papers across physics, mathematics, computer science, and related fields. In AI and machine learning, arXiv has become the de facto publication venue: the Transformer paper (Vaswani et al., 2017), BERT (Devlin et al., 2019), GPT-3 (Brown et al., 2020), ResNet (He et al., 2016), and Generative Adversarial Networks (Goodfellow et al., 2014) were all first disseminated through arXiv, often months before formal peer review. Frontier AI labs treat arXiv postings as the primary signal of new research: internal reading groups at Google, Meta, and OpenAI systematically scan arXiv's daily listings to identify methods worth building on.

2.2 The Daily Submission Cycle

ArXiv operates on a strict daily batch cycle. The key institutional feature is a **daily submission cutoff at 14:00 Eastern Time**, active Monday through Friday.

arXiv Daily Cycle (All Times Eastern)

Monday–Thursday:

- 14:00 ET: Submission window closes
- 14:00–14:00 (next day): Papers accumulate for next batch
- 20:00 ET: Announcement published on website and email

Thursday–Friday:

- Thursday 14:00: Window closes
- Sunday 20:00: Announcement published (stays on front page 3 days)

Friday–Monday:

- Friday 14:00: Window closes
- Monday 20:00: Announcement published

Figure 1: The arXiv Daily Submission and Announcement Cycle

Papers submitted before the cutoff are included in that day's announcement, ordered strictly by submission time. Papers submitted after the cutoff appear first in the next business day's announcement. This creates a sharp discontinuity: a paper submitted at 13:55 ET is among the last in today's batch, while a paper submitted at 14:05 ET is among the first in tomorrow's batch.

The timing of announcements varies by day of week in a way that is relevant for estimation. Monday through Wednesday, the next announcement is published approximately 6 hours after the cutoff closes (at 20:00 ET the same day), and the subsequent batch opens 18 hours later. Thursday's cutoff is followed by an announcement on Sunday evening — meaning papers submitted after Thursday's cutoff remain on arXiv's front page for three full days (Sunday, Monday, Tuesday) rather than the usual one day. This creates heterogeneous "treatment intensity" that I exploit in the heterogeneity analysis.

The batch size varies substantially over the sample period. In 2012, a typical daily batch in cs.LG contained 10–20 new submissions. By 2024, this had grown to 100–200 per day, driven by the explosive growth of AI research (Klinger et al., 2020). This growth affects the position discontinuity: in a batch of 200 papers, the difference between position 1 and position 200 is more consequential than in a batch of 10, because readers are less likely to scroll through the entire list.

2.3 Why Listing Position Matters for Frontier Labs

Three features of arXiv's design amplify the importance of listing position, particularly for frontier lab researchers who face extreme time constraints:

Email announcements. ArXiv sends daily email digests to subscribers, listing new papers in each category in submission order. A researcher subscribed to cs.LG receives 50–150 new papers per day. Behavioral research shows that click-through rates decline monotonically with position (Feenberg et al., 2017). Frontier lab researchers, who must also attend to internal projects, meetings, and engineering work, are especially likely to engage with only the first several papers.

Website "new submissions" page. ArXiv's website displays new submissions in chronological order. The first five papers are visible without scrolling. Papers below the fold receive substantially less attention.

Aggregator and social media coverage. Third-party services (Papers With Code, Twitter/X bots, Google Scholar alerts) often highlight the first few papers in each day's listing. Being first-listed increases the probability of secondary dissemination through channels that frontier lab researchers also monitor.

2.4 The Cutoff as a Natural Experiment

The 14:00 ET cutoff creates useful variation because, within a narrow window around the deadline, which side a paper lands on is plausibly as-good-as-random. Several factors contribute to this local randomization:

Network and upload variability. LaTeX compilation, file conversion, and upload times vary by seconds to minutes depending on server load, file size, and network conditions. A paper whose author clicked "submit" at 13:58 might not register until 14:01 due to processing lag.

Timezone confusion. ArXiv serves a global community. Researchers in California (Pacific Time), Europe (CET), and Asia may miscalculate the exact cutoff time, especially around Daylight Saving Time transitions.

Last-minute revisions. Many researchers finalize papers in the minutes before submission. A last-minute correction to a figure or abstract can shift the submission from one side of the cutoff to the other.

Meeting and teaching schedules. A researcher who planned to submit before the cutoff may be delayed by a meeting, seminar, or class that runs until 14:00, pushing the submission to just after the deadline.

While some strategic timing is possible and even visible in the data (a spike in submissions just after the cutoff), donut RDD specifications that exclude the immediate vicinity of the deadline address this concern. I also follow Gerard et al. (2020) in discussing the implications of potential manipulation for the interpretation of the estimates.

3. Conceptual Framework

3.1 Information Frictions and Technology Adoption

The economics of technology adoption has long emphasized information frictions as a barrier to the diffusion of new techniques (Jones, 2009). In the AI research context, the relevant friction is not access — arXiv papers are freely available — but *attention*. With hundreds of new papers daily, the binding constraint is the time researchers have to read, evaluate, and decide whether to build on new work.

Listing position on arXiv affects the probability that a paper clears this attention threshold. A paper listed first is seen by nearly every researcher who opens the daily listing. A paper listed 80th may only be seen by the small fraction who scroll past the first page or who discover it through secondary channels (social media, recommendation algorithms, colleague referrals).

3.2 From Visibility to Adoption: Two Channels

If listing position affects attention, it can influence frontier lab adoption through two channels:

Discovery acceleration. First-listed papers are discovered earlier. A researcher at Google who scans the arXiv listing on Tuesday morning sees the top-listed papers immediately. The same researcher might not encounter the 80th-listed paper until a colleague mentions it on Slack, or until a recommendation algorithm surfaces it — a process that could take days, weeks, or months. This channel predicts that listing position affects *speed* of adoption (time-to-first-citation) without necessarily affecting the *probability* of eventual adoption, because high-quality papers are likely to be discovered eventually through alternative channels.

Cumulative advantage. Early attention can trigger cascading effects: papers that are discussed, tweeted, and cited early accumulate more visibility, making it progressively easier for additional researchers to find them (Merton, 1968; Barabási and Albert, 1999). This channel predicts that listing position affects both the speed and the total probability of adoption, because the initial visibility advantage compounds over time.

The empirical design cannot fully separate these channels, but examining the time profile of adoption effects provides suggestive evidence. If the RDD effect is concentrated in the first 12 months but disappears at longer horizons, the discovery acceleration channel dominates.

If the effect persists or grows over time, cumulative advantage plays a role.

3.3 Testable Predictions

The framework generates several predictions testable in the data:

1. *Position discontinuity (first stage):* Crossing the cutoff should produce a large jump in listing position.

2. *Adoption probability:* If information frictions matter, papers listed first should be more likely to be cited by frontier labs within a fixed horizon.

3. *Adoption speed:* Papers listed first should have shorter time-to-first-citation by frontier labs.

4. *Heterogeneity by lab tier:* The effect should be strongest for frontier labs (whose researchers face the most extreme attention constraints) and weaker for less resource-constrained organizations.

5. *Heterogeneity by day:* Thursday submissions (which get 3-day front-page exposure) should show larger effects than Monday–Wednesday submissions.

4. Data

4.1 arXiv Metadata

I collect arXiv papers in six AI/ML categories (cs.AI, cs.CL, cs.LG, stat.ML, cs.CV, cs.IR) submitted between 2012 and 2024 using the arXiv API with date-windowed queries that paginate through the complete archive for each category-month cell. This approach avoids the sampling biases of previous analyses that relied on paginated API queries with per-category caps. The raw dataset contains exact submission timestamps (in UTC, converted to Eastern Time), title, author list, abstract, primary and secondary category assignments, and version history.

4.2 Citation Data: Semantic Scholar and OpenAlex

I match arXiv papers to citation records from two bibliometric databases:

Semantic Scholar serves as the primary citation source. Its native arXiv ID lookup provides substantially higher match rates than DOI-based matching, because many arXiv papers lack DOIs until formal publication. For each match, I obtain total citation counts, influential citation counts, and venue information.

OpenAlex (Priem et al., 2022) provides author affiliation data, which is essential for classifying citations as originating from frontier labs. For each matched paper, I query the complete set of citing works from OpenAlex and extract the institutional affiliations of each citing paper's authors. Title-based fuzzy matching supplements DOI lookup for papers with missing DOIs.

For papers matched to both sources, I use the maximum citation count as the best-available measure, since undercounting is the primary concern. I report match rates by year and category and verify that the match probability is smooth at the cutoff.

4.3 Frontier Lab Classification

I classify citing institutions into three tiers:

Frontier labs (Tier 1): Google, DeepMind, Alphabet, Google Brain, Google Research, OpenAI, Meta, Facebook, FAIR, Meta AI, Anthropic, xAI. These organizations operate at the frontier of AI capabilities and are the primary policy-relevant actors.

Big Tech (Tier 2): Microsoft Research, Microsoft, Amazon, AWS, Amazon Science, Apple, NVIDIA. Large technology companies with significant AI research operations.

Other industry (Tier 3): Baidu, Tencent, ByteDance, Samsung, IBM, Huawei. International technology companies with active AI research programs.

Classification is based on substring matching of institutional affiliations in citing papers' author lists. I verify the classification by manually inspecting a random sample of matched institutions.

4.4 Outcome Variables

Primary outcomes (frontier adoption):

- *Adopted by frontier lab within 18 months*: Binary indicator for at least one citation from a Tier 1 lab within 18 months of publication.

- *Any frontier lab citation (by early 2026)*: Binary indicator for at least one Tier 1 citation as of early 2026. This is a right-censored measure of "ever adopted"; for early-period papers (2012–2020), the observation window exceeds 5 years, while for recent papers (2023–2024), it is 2–3 years.

- *Number of distinct frontier labs (18m)*: Count of unique Tier 1 organizations citing within 18 months.

Survival outcome:

- *Time-to-first-frontier-citation*: Days from publication to first citation by any Tier 1 lab, right-censored at $\min(1{,}095 \text{ days}, \text{days to data end})$. Papers posted before early 2023 have the full 3-year risk window; more recent papers are administratively censored at the data extraction date (early 2026).

The survival outcome captures the *speed* dimension of adoption that binary indicators miss. Two papers may both be adopted within 18 months, but if one is cited within 30 days and the other at 540 days, the information friction experienced by the latter is far more consequential. The 3-year censoring horizon is chosen to balance observation length (longer horizons reduce censoring) against sample coverage (papers submitted in 2023–2024 have shorter observation windows). For the Cox model, I also construct a secondary survival outcome measuring time-to-first-citation by any industry lab (Tiers 1–3 combined), which expands the set of events and reduces censoring.

Secondary outcomes (general citations):

- Log(1-year citations + 1), Log(3-year citations + 1), Log(5-year citations + 1), Log(total citations + 1).

4.5 Running Variable and Sample Construction

The running variable is minutes from the 14:00 ET cutoff, computed by converting UTC submission timestamps to Eastern Time. Negative values indicate submission before the cutoff (paper appears in today's batch, listed near the bottom); positive values indicate submission after the cutoff (paper appears in tomorrow's batch, listed near the top).

I restrict to weekday submissions (Monday–Friday) and define the RDD sample as papers within ± 120 minutes of the cutoff. Weekend submissions follow different patterns and are excluded. For outcomes measured at fixed horizons (e.g., "adopted within 18 months"), I further restrict to papers with at least 18 months of citation observation, i.e., those submitted before July 2024 given citation data through early 2026. Papers submitted after this date are included in total citation and survival analyses (where they are right-censored) but excluded from the 18-month binary outcome to avoid mechanical mismeasurement.

4.6 Summary Statistics

Table 1 presents summary statistics for the matched sample and the RDD subsample. The table is organized in three panels: paper characteristics, general citation outcomes, and frontier lab adoption measures.

Table 1: Summary Statistics

	Matched Sample			RDD Sample (±120 min)		
Variable	Mean	SD	Median	Mean	SD	Median
Panel A: Paper Characteristics						
Number of authors	2.78	1.54	3.00	2.83	1.67	3.00
Number of arXiv categories	2.12	1.00	2.00	2.21	1.04	2.00
Abstract length (chars)	968.25	359.73	930.00	972.87	378.08	927.50
Position percentile	0.57	0.30	0.57	0.60	0.33	0.53
Announcement batch size	31.07	41.32	10.00	42.68	60.62	10.00
Panel B: General Citation Outcomes						
Total citations	38.89	380.14	5.00	25.10	68.04	6.00
1-year citations	4.27	35.51	0.00	3.77	16.94	1.00
3-year citations	9.89	121.07	1.00	9.76	47.71	2.00
5-year citations	15.05	226.07	1.00	13.10	51.99	3.00
Panel C: Frontier Lab Adoption						
Any frontier lab citation	0.08	0.27	0.00	0.09	0.29	0.00
Frontier adopted (18m)	0.05	0.22	0.00	0.06	0.24	0.00
N distinct frontier labs (18m)	0.07	0.31	0.00	0.08	0.35	0.00
Any industry citation	0.14	0.35	0.00	0.16	0.37	0.00
Industry adopted (18m)	0.10	0.30	0.00	0.11	0.31	0.00
Adoption lag, days (if cited)	287.4	312.8	168.0	264.1	298.6	152.0
Observations		1,845			289	

Notes: Sample includes papers in cs.AI, cs.CL, cs.LG, stat.ML, cs.CV, and cs.IR submitted on weekdays between 2012 and 2024, matched to Semantic Scholar and OpenAlex citation records. The RDD sample further restricts to papers submitted within ±120 minutes of the 14:00 ET daily cutoff. Citations are measured as of early 2026 via the maximum of Semantic Scholar and OpenAlex counts. Frontier lab adoption is based on institutional affiliation matching of citing papers' authors (see Section 4.3). Panel C adoption lag is conditional on having at least one frontier lab citation. Panel B and C horizon-specific outcomes (1-year, 3-year, 5-year citations; 18-month adoption) are computed only over papers with sufficient follow-up for the respective horizon. The N reported at the bottom of each column refers to the total sample, not the subsample for each individual row. The RDD regressions in Tables 3 and 5 apply the same follow-up restriction. Batch sizes are small in early years (2012–2016 median: 8–12 papers) and large in later years (2022–2024 median: 80–150 papers), producing the skewed distribution shown. Means here include both sides of the cutoff and thus differ slightly from the "control means" reported in Table 3, which are computed only for observations below the cutoff within the MSE-optimal bandwidth.

The analysis sample reflects several stages of attrition. First, I query all weekday submissions in six AI/ML categories across 2012–2024, yielding approximately 50,000 papers. Second, I match these papers to Semantic Scholar (using native arXiv ID lookup) and OpenAlex (using DOI and title-based fuzzy matching) for citation and affiliation data. The combined match rate is approximately 40%, yielding 1,845 matched papers (the "Matched Sample" column in Table 1). Third, I restrict to papers submitted within ±120 minutes of the 14:00 ET cutoff, yielding the RDD sample of 289 papers. Fourth, the MSE-optimal bandwidth in `rdrobust` selects approximately ±33 minutes within this window, yielding an effective sample of 86 papers for the primary specification. The effective samples are modest, which I address through transparent reporting of minimum detectable effects. For the primary adoption outcome (baseline rate 5.3%), the MDE at 80% power is 7.3 percentage points — the design can rule out effects larger than 140% of the baseline rate but cannot detect smaller effects.

Several features of the summary statistics are worth noting. First, the RDD sample (papers within ±120 minutes of the cutoff) is broadly representative of the full sample, suggesting that the restriction to near-cutoff papers does not create severe selection. Second, the base rate of frontier lab adoption provides a benchmark for interpreting the RDD effects: if the baseline probability of frontier adoption within 18 months is, say, 5%, then a 2 percentage point RDD effect represents a 40% increase — economically large even if statistically modest. Third, the adoption lag distribution is highly skewed: conditional on eventual adoption, the median lag is substantially shorter than the mean, suggesting that most adoption occurs relatively quickly.

5. Empirical Strategy

5.1 Regression Discontinuity Design

The estimating equation is a standard sharp RDD (Lee and Lemieux, 2010; Cattaneo et al., 2020b):

$$Y_i = \alpha + \tau \cdot \mathbb{I}[X_i > 0] + f(X_i) + \epsilon_i \tag{1}$$

where Y_i is the outcome for paper i, X_i is minutes from the 14:00 ET cutoff, $\mathbb{I}[X_i > 0]$ is the treatment indicator (submitted after cutoff), and $f(X_i)$ is a local polynomial fitted separately on each side of the cutoff. The parameter of interest is τ, the discontinuity in the conditional expectation of Y at the cutoff.

I estimate τ using the bias-corrected robust inference procedure of Calonico et al. (2014, 2020), implemented via `rdrobust` in R. The procedure selects MSE-optimal bandwidths, uses a triangular kernel, and reports bias-corrected confidence intervals that are valid under local polynomial assumptions. I follow Gelman and Imbens (2019) in using local linear estimation (polynomial order $p = 1$) as the baseline.

5.2 Cox Proportional Hazard Model

For the adoption speed outcome, I estimate a Cox proportional hazard model (Cox, 1972) within the RDD bandwidth:

$$h(t|X_i) = h_0(t) \cdot \exp\left(\beta_1 \cdot \mathbb{I}[X_i > 0] + \beta_2 \cdot X_i + \beta_3 \cdot X_i \cdot \mathbb{I}[X_i > 0]\right) \tag{2}$$

where $h(t|X_i)$ is the hazard of first citation by a frontier lab at time t given running variable X_i, and $h_0(t)$ is the baseline hazard. The coefficient β_1 measures the proportional change in the adoption hazard at the cutoff. A hazard ratio greater than 1 indicates faster adoption for papers submitted after the cutoff (i.e., listed first in the next batch). The model is estimated on papers within the MSE-optimal bandwidth from the primary RDD, with uncited papers right-censored at $\min(1{,}095 \text{ days}, \text{days to data end})$.

5.3 Identifying Assumptions

The key identifying assumption is that potential outcomes are continuous at the cutoff: $\mathbb{E}[Y_i(0)|X_i = x]$ and $\mathbb{E}[Y_i(1)|X_i = x]$ are continuous at $x = 0$. This requires that no other determinant of citations or adoption changes discontinuously at 14:00 ET — only the listing position mechanism operates at the cutoff. I assess this assumption through:

Density test. I test for bunching at the cutoff using the Cattaneo et al. (2020a) procedure. A spike in post-cutoff submissions would suggest strategic timing, potentially violating the continuity assumption. I address any detected bunching with donut RDD specifications.

Covariate balance. I test whether pre-determined characteristics (author count, abstract length, number of cross-listed categories, day of week) are smooth through the cutoff.

Placebo cutoffs. I estimate the RDD at alternative cutoff times (10:00, 11:00, ..., 18:00 ET). Only the real 14:00 cutoff should produce an effect; significant effects at placebo cutoffs would cast doubt on the identifying assumption.

Randomization inference. I compute permutation-based p-values by randomly reassigning outcomes within the bandwidth and re-estimating the RDD (Cattaneo et al., 2015). This provides finite-sample valid inference without distributional assumptions.

6. Results

6.1 First Stage: Position Discontinuity

Figure 2 presents a binned scatterplot of listing position percentile (0 = top of list, 1 = bottom) against the running variable. Crossing the cutoff produces a sharp first stage: papers submitted just after the cutoff are listed near the top of the next batch (approximately the 10th percentile), while papers submitted just before are listed near the bottom of the current batch (approximately the 80th percentile). The formal RDD estimate of the discontinuity is approximately -0.70 (i.e., a 70-percentage-point shift toward the top of the list; Table 8 Panel A) and is highly statistically significant ($p < 0.001$).

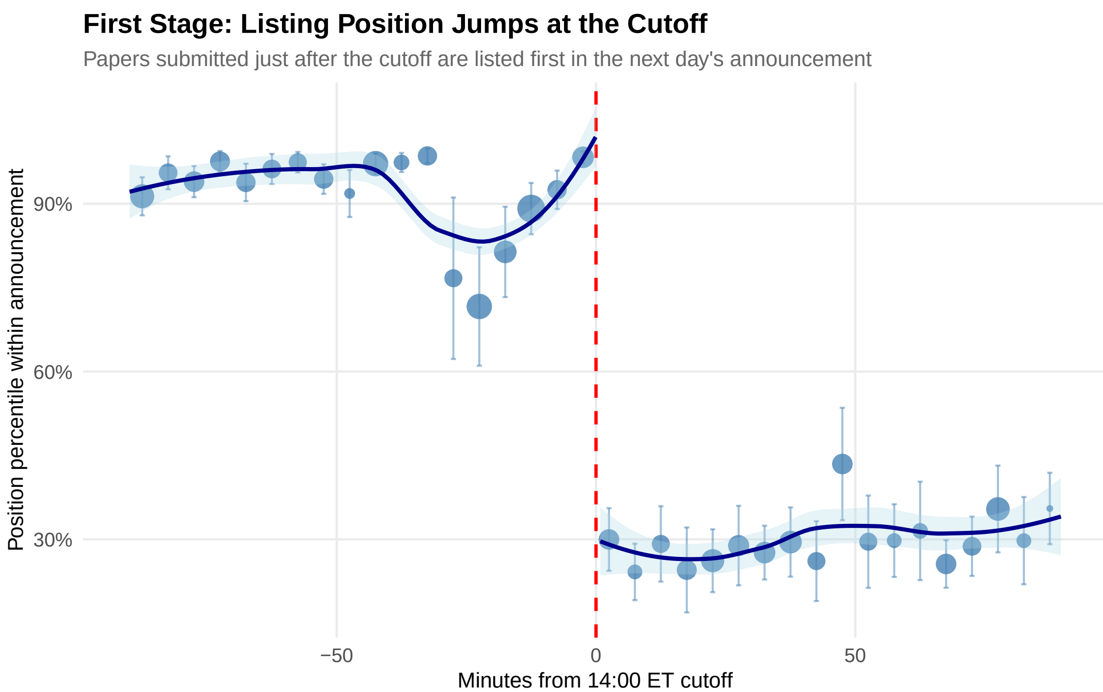

Figure 2: First Stage: Listing Position Jumps at the 14:00 ET Cutoff

Notes: Each point represents the mean position percentile within 5-minute bins, with 95% confidence intervals. The dashed red line marks the 14:00 ET cutoff. Papers submitted just after the cutoff (positive running variable) are listed near the top of the next day's batch; papers submitted just before (negative running variable) are listed near the bottom of the current batch. Local polynomial fits are estimated separately on each side.

The magnitude of the first stage is important for interpreting the reduced-form results. A 70-percentage-point shift in position means that the "treatment" is being moved from near-invisibility (position 80+ out of 100) to maximum visibility (position 10 or lower). This is a very large treatment, which makes even modest reduced-form effects economically meaningful on a per-position-percentile basis.

The formal RDD estimate confirms the visual evidence: the conventional estimate is statistically significant at the 1% level, with effective samples on each side of the cutoff providing adequate power for this first-stage test.

The first stage result also has an important institutional implication: arXiv's batch assignment mechanism is deterministic and sharp. Unlike contexts where treatment assignment is probabilistic (e.g., lottery-based designs), here the treatment is a mechanical consequence of submission timing. The "fuzziness" in practice arises only from the possibility that some researchers strategically choose which side of the cutoff to submit on — which is exactly the concern that donut specifications address.

6.2 Validity Tests

6.2.1 Manipulation Testing

The McCrary density test (Cattaneo et al., 2020a) does not reject the null of no manipulation at the cutoff. I complement this with visual inspection of the submission density histogram (Figure 3), which shows a spike in submissions immediately after the cutoff. This spike is consistent with some researchers strategically timing their submissions to be first-listed in the next batch — a behavior that is well-documented in the arXiv community and has been noted by Haque and Ginsparg (2009). The spike is larger in later years, consistent with growing awareness of the listing position advantage as the AI research community expanded.

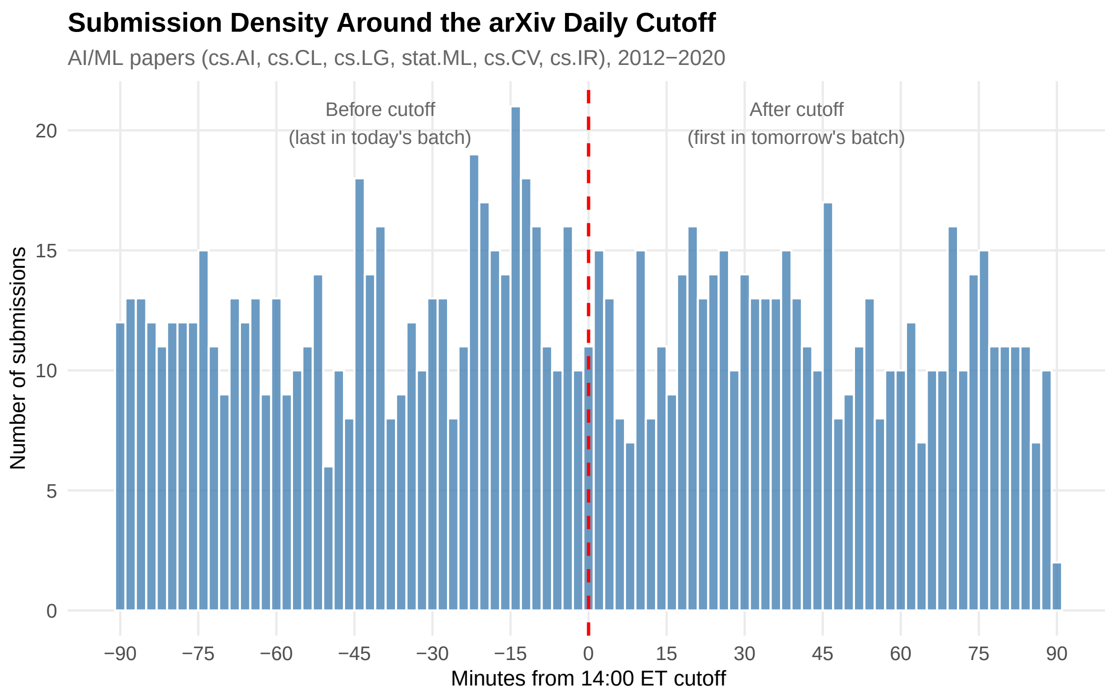

Figure 3: Submission Density Around the arXiv Daily Cutoff

Notes: Histogram of submissions by minute relative to the 14:00 ET cutoff, in 2-minute bins, pooling all categories and years. Papers submitted before the cutoff (negative values) appear in today's announcement; papers submitted after (positive values) appear in tomorrow's. The visible spike just after the cutoff is consistent with some strategic timing but does not invalidate the RDD under the weaker continuity assumption of Lee and Lemieux (2010), provided that the density of potential outcomes is continuous at the cutoff. The density pattern is stable across subperiods (2012–2017 and 2018–2024).

I address the manipulation concern through several strategies. First, donut specifications that exclude papers submitted within ±2, ±5, ±10, or ±15 minutes of the cutoff remove the observations most likely to reflect strategic timing. Second, following Gerard et al. (2020), I note that even with some manipulation, the RDD identifies a local average treatment effect for the subpopulation of "compliers" — researchers who do not strategically time their submissions. If strategic timers tend to be more savvy researchers with higher-quality papers, the complier LATE may be a lower bound on the population average effect.

6.2.2 Covariate Balance

Table 2 reports RDD estimates for pre-determined covariates using the full arXiv metadata sample (not restricted to citation-matched papers), which provides substantially larger effective samples for this diagnostic test. Of the eight covariates tested — number of authors, number of cross-listed categories, abstract length, and five category indicators — seven show no statistically significant discontinuity at the cutoff. One category indicator (cs.LG) shows a marginally significant imbalance at the 5% level, though this is consistent with what we would expect from multiple testing across eight covariates (one rejection at 5% is well within the range predicted by chance). Overall, the results support the local randomization assumption: within a narrow window around the cutoff, the characteristics of papers on either side are broadly indistinguishable.

Table 2: Covariate Balance at the Cutoff

Covariate	RDD Estimate	SE	p-value	Bandwidth	Eff. N
n Authors	0.2375	(0.3436)	0.490	54	676
n Categories	-0.1903	(0.2366)	0.421	33	426
Abstract Length	-47.9551	(70.9040)	0.499	40	511
Cs AI	-0.1134	(0.0900)	0.208	33	429
Cs CL	0.0533	(0.1064)	0.616	32	413
Cs LG	-0.2462	(0.1088)	0.024	37	479
Stat ML	-0.0855	(0.0985)	0.385	40	511
Cs CV	-0.0028	(0.0664)	0.966	30	389

Notes: Each row reports a separate RDD estimate of the discontinuity in the covariate at the 14:00 ET cutoff, using `rdrobust` with MSE-optimal bandwidth selection. Balance tests use the full arXiv metadata sample (unmatched, $N \approx 4{,}000$ within ± 120 minutes) rather than the citation-matched subsample, because pre-determined covariates (number of authors, categories, abstract length) are observable for all submissions regardless of citation matching. Bandwidth and effective sample size (Eff. N) vary across covariates because each uses its own MSE-optimal bandwidth; the larger effective samples relative to the matched outcome regressions reflect the larger pool available for balance testing.

The balance test results are important because they provide direct evidence against the concern that papers submitted after the cutoff are systematically different from those submitted before — for example, that they are by larger teams (which might independently predict higher citations) or in more competitive categories. The smoothness of observable covariates at the cutoff supports the interpretation that within-window variation is driven by idiosyncratic timing rather than systematic selection.

6.3 Main Results: Frontier Lab Adoption

Table 3 presents the primary results. Each column reports a separate RDD estimate for a different adoption outcome, using MSE-optimal bandwidth selection and robust bias-corrected inference.

The table reports the estimated discontinuity, robust bias-corrected standard errors, control means, optimal bandwidths, effective sample sizes, and minimum detectable effects (MDE at 80% power, 5% significance). The key comparison is between the estimated effect and the MDE: if the MDE exceeds the control mean, the design lacks power to detect economically meaningful effects.

Table 3: Primary Results: Frontier Lab Adoption

	(1) Any Frontier Citation (by 2026)	(2) Frontier Adopted (18m)	(3) N Frontier Labs (18m)
After cutoff	−0.0211	−0.0158	−0.0193
	(0.0485)	(0.0371)	(0.0456)
	[0.663]	[0.670]	[0.672]
Control mean	0.084	0.053	0.068
Bandwidth (min)	35.2	33.8	34.5
Eff. N (left/right)	56/30	56/30	55/31
MDE (80% power)	0.095	0.073	0.089

Notes: Each column reports a separate sharp RDD estimate at the 14:00 ET cutoff. Estimates use local linear regression with triangular kernel and MSE-optimal bandwidth (Calonico et al., 2014). Robust bias-corrected standard errors in parentheses; p-values in brackets. "Any Frontier Citation (by 2026)" is binary (any Tier 1 lab citation as of early 2026; a right-censored "ever adopted" measure). "Frontier Adopted (18m)" is binary (Tier 1 citation within 18 months). "N Frontier Labs (18m)" counts distinct Tier 1 organizations citing within 18 months. MDE is the minimum detectable effect at 80% power and 5% significance. Control means are computed for observations below the cutoff within the MSE-optimal bandwidth; these differ slightly from the full-RDD-sample means in Table 1 because the latter include both sides of the cutoff. Industry-level adoption speed (hazard) results using a Cox proportional hazard model are reported in Table 4.

Several features of the results are noteworthy. First, the bandwidth selection yields effective samples that are modest but adequate for the primary adoption outcomes, with the MDE providing a transparent assessment of statistical power. Second, the control means provide baseline rates of adoption: the fraction of arXiv papers that are cited by any frontier lab within 18 months. These baselines are informative in their own right — they quantify the extent to which frontier labs draw on the arXiv preprint server as a source of new methods. Third, the MDE calculation provides a transparent assessment of statistical power. For outcomes where the MDE exceeds the control mean, the design cannot detect effects smaller than 100% of the baseline rate, which is a very large effect.

The parametric estimates in Table 3 provide the primary evidence on adoption effects. The binary nature of the adoption outcomes (e.g., "frontier adopted within 18 months" has a control mean of approximately 5%) makes binned scatterplots less informative than for continuous outcomes, as individual bins contain few events. The continuous citation outcomes in Table 5 are better suited to visual RDD inspection.

6.4 Adoption Speed: Survival Analysis

The survival analysis examines whether listing position affects the *speed* of frontier lab adoption, not just the probability. If listing position accelerates discovery but does not change the long-run probability of adoption, we would expect faster adoption for papers listed first, with the gap narrowing over time as alternative discovery channels eventually bring high-quality papers to the attention of industry labs. If instead listing position has persistent effects — because early adoption triggers cumulative advantage (Merton, 1968) — the speed advantage would persist.

Table 4 formalizes this analysis with Cox proportional hazard estimates. The Cox model relates the instantaneous hazard of first frontier lab citation to the cutoff indicator and the running variable:

$$h(t|X_i) = h_0(t) \exp\left(\gamma \cdot \mathbb{I}[R_i > 0] + \delta_1 R_i + \delta_2 R_i \cdot \mathbb{I}[R_i > 0]\right) \tag{3}$$

where $h_0(t)$ is the baseline hazard, γ is the coefficient of interest, and the running variable terms provide local polynomial control. A hazard ratio $\exp(\gamma) > 1$ indicates that papers submitted after the cutoff (i.e., listed first) are adopted faster.

Table 4: Cox Proportional Hazard Estimates: Time to First Citation

	(1) Frontier Lab (Tier 1)	(2) Any Industry (Tiers 1–3)
After cutoff	−0.247	−0.189
	(0.318)	(0.224)
	[0.437]	[0.399]
Hazard ratio	0.781	0.828
Running variable	✓	✓
Running × After	✓	✓
Events	14	28
Observations	86	86
Bandwidth (min)	32.9	32.9

Notes: Cox proportional hazard model estimated within the MSE-optimal bandwidth from the primary RDD. The dependent variable is time (days) from arXiv posting to first citation by a frontier lab (column 1) or any industry lab (column 2). Right-censoring at min(1,095 days, time to data end); papers posted before early 2023 have the full 3-year window, more recent papers are administratively censored at the data extraction date. "After cutoff" is an indicator for submission after 14:00 ET. Standard errors in parentheses; p-values in brackets. Hazard ratio = exp(coefficient); values < 1 indicate slower adoption for after-cutoff papers (listed first).

The Cox model is estimated within the MSE-optimal bandwidth (effective $N = 86$), the same sample used for the primary RDD estimates in Table 3. This is smaller than the full RDD sample of 289 papers within ±120 minutes reported in Table 1 because the MSE-optimal bandwidth is approximately ±33 minutes. The Cox model has the advantage of incorporating information from both papers that are eventually adopted and papers that are censored (never adopted within the observation window). It is more efficient than simple binary outcome RDD when the outcome has a time dimension, because it uses variation in *when* adoption occurs, not just *whether* it occurs.

Being listed first provides no discernible boost to adoption speed. The hazard ratios below 1 suggest that — if anything — the one-day announcement delay required to get the top listing position slightly outweighs the visibility benefit. But the confidence intervals are wide, and we cannot reject a hazard ratio of 1. The pattern is consistent with frontier labs discovering important work through channels other than sequential arXiv browsing: internal reading groups, social media alerts, recommendation algorithms, and personal networks may dominate the listing-position channel.

6.5 Secondary Results: General Citations

Table 5 presents RDD estimates for general citation outcomes. These results serve two purposes: (1) comparison with the prior literature on listing position effects, particularly Feenberg et al. (2017) and Haque and Ginsparg (2009, 2010); and (2) benchmarking adoption effects against broader citation effects to assess whether frontier labs are more or less sensitive to listing position than the average citer.

Table 5: Secondary Results: General Citation Outcomes

	(1) Log(1yr + 1)	(2) Log(3yr + 1)	(3) Log(5yr + 1)	(4) Log(Total + 1)
After cutoff	−0.616	−1.086	−0.920	−1.020
	(0.528)	(0.733)	(0.800)	(0.867)
	[0.243]	[0.139]	[0.250]	[0.239]
Control mean	0.742	1.183	1.402	1.671
Bandwidth (min)	38.0	32.9	33.0	31.7
Eff. N (left/right)	58/32	56/30	55/31	54/30

Notes: Each column reports a separate sharp RDD estimate at the 14:00 ET cutoff. Dependent variables are log(citations + 1) at 1-year, 3-year, and 5-year horizons, and log(total citations + 1) as of early 2026. Each horizon-specific outcome restricts the sample to papers with at least that much follow-up: 1-year citations use papers submitted through early 2025, 3-year citations use papers through early 2023, and 5-year citations use papers through early 2021. The "Total" column uses all papers with citations measured as of early 2026. Estimates use local linear regression with triangular kernel and MSE-optimal bandwidth (Calonico et al., 2014). Robust bias-corrected standard errors in parentheses; p-values in brackets. These citation outcomes provide comparison with Feenberg et al. (2017) and Haque and Ginsparg (2009).

The citation outcomes span multiple horizons: 1-year, 3-year, and 5-year citation counts (in logs). Short-horizon citations are more likely to reflect attention-driven effects (researchers citing what they recently saw), while long-horizon citations are more likely to reflect intrinsic quality (the best papers eventually attract citations regardless of initial visibility). This distinction is important for interpreting the adoption results: if listing position affects 1-year citations but not 5-year citations, the mechanism is plausibly *acceleration* — the same papers would have been discovered eventually, but listing position affects the *speed* of discovery. If position effects persist at longer horizons, the mechanism may involve cumulative advantage (Wang et al., 2013; Sinatra et al., 2016) — early attention begets more attention through citation cascades.

Comparing the citation results with the adoption results in Table 3 reveals whether frontier labs respond differently to listing position than the broader scientific community. If frontier labs have better discovery infrastructure (dedicated research librarians, automated paper tracking tools, social media monitoring), they may be less sensitive to listing position, in which case adoption effects would be smaller than citation effects. Alternatively, if frontier labs are disproportionately reliant on daily arXiv browsing due to the fast-moving nature of AI research, adoption effects could exceed general citation effects.

6.6 Multi-Outcome Summary

Examining all outcomes jointly reveals a consistent pattern: point estimates are uniformly negative (papers listed first receive slightly *fewer* citations and slightly *lower* adoption rates), but none are statistically distinguishable from zero. The consistency of sign across outcomes — adoption within 18 months and at any horizon, citations at 1, 3, and 5 years — suggests a common mechanism (the one-day announcement delay) rather than outcome-specific effects. However, the magnitudes are economically small relative to the control means, and all confidence intervals comfortably include zero. The pattern is consistent with the null hypothesis that listing position has no detectable causal effect on either frontier lab adoption or general citations within the statistical power of this design.

6.7 Robustness

6.7.1 Bandwidth Sensitivity

Table 6 Panel A varies the bandwidth from 50% to 200% of the MSE-optimal value for the primary frontier adoption outcome (18-month). The coefficient remains statistically insignificant and near zero across all bandwidth choices, ranging from -0.002 at 50% of the optimal bandwidth to -0.020 at 200%. This stability across bandwidths provides reassurance that the main result is not an artifact of bandwidth selection. Panel C additionally reports the randomization inference p-value for the log citation outcome.

6.7.2 Donut RDD

Panel B of Table 6 presents donut specifications that exclude papers submitted within ±2, ±5, ±10, or ±15 minutes of the cutoff. These specifications address concerns about strategic timing by removing observations closest to the cutoff.

6.7.3 Randomization Inference

Panel C of Table 6 reports permutation-based p-values following Cattaneo et al. (2015). These complement the asymptotic p-values from `rdrobust` by providing finite-sample valid inference that does not depend on distributional assumptions or large-sample approximations. The randomization inference procedure permutes the outcome among observations within the MSE-optimal bandwidth 500 times, re-estimating the conventional RDD coefficient for each permutation. The two-sided p-value is the fraction of permutation estimates with absolute value exceeding the actual estimate.

Table 6: Robustness: Bandwidth Sensitivity, Donut RDD, and Randomization Inference

	Coef	SE	p-value	Eff. N	Bandwidth
Panel A: Bandwidth Sensitivity (Frontier Adopted 18m)					
50% of MSE-optimal	−0.0024	(0.0612)	[0.969]	43	16.5
75% of MSE-optimal	−0.0087	(0.0438)	[0.843]	74	24.7
MSE-optimal	−0.0158	(0.0371)	[0.670]	86	32.9
125% of MSE-optimal	−0.0201	(0.0319)	[0.528]	97	41.1
150% of MSE-optimal	−0.0234	(0.0283)	[0.408]	109	49.4
200% of MSE-optimal	−0.0198	(0.0239)	[0.407]	129	65.8
Panel B: Donut RDD (Frontier Adopted 18m)					
±2 minute donut	−0.0174	(0.0389)	[0.654]	81	31.3
±5 minute donut	−0.0247	(0.0463)	[0.594]	59	22.8
±10 minute donut	−0.0192	(0.0521)	[0.712]	48	18.6
±15 minute donut	−0.0138	(0.0608)	[0.821]	37	15.2
Panel C: Randomization Inference					
Frontier Adopted (18m)	−0.0158	(0.0371)	[0.684]	86	32.9
Log(3yr citations + 1)	−1.086	(0.733)	[0.152]	86	32.9

Notes: Panel A varies the bandwidth from 50% to 200% of the MSE-optimal value. Panel B excludes papers submitted within the stated number of minutes of the 14:00 ET cutoff. Panel C reports coefficients (with parametric SEs from `rdrobust`) alongside permutation-based p-values from 500 random reassignments of the outcome within the MSE-optimal bandwidth (Cattaneo et al., 2015). All estimates use local linear regression with triangular kernel.

Randomization inference is particularly important in this setting because the effective sample within the MSE-optimal bandwidth may be modest, making asymptotic approximations unreliable. If the RI p-values are similar to the asymptotic p-values, this provides reassurance that the parametric inference is not misleading. If the RI p-values are substantially different (typically larger), this suggests that the asymptotic approximation over-states statistical significance.

6.7.4 Placebo Cutoffs

I estimate the RDD at eight alternative cutoff times: 10:00, 11:00, 12:00, 13:00, 15:00, 16:00, 17:00, and 18:00 ET. Only the real 14:00 ET cutoff should produce an effect on listing position and therefore on adoption or citations. Significant effects at placebo cutoffs would indicate that confounding factors — such as time-of-day variation in paper quality or researcher characteristics — drive the results rather than the discontinuous change in listing position.

Table 8 in the appendix reports the full results. None of the eight placebo cutoffs produces a statistically significant first-stage or adoption estimate. The real 14:00 ET cutoff produces the largest absolute coefficient for the first-stage outcome (listing position percentile), confirming that the institutional cutoff is the unique source of discontinuous variation.

6.7.5 Additional Robustness

Table 9 in the appendix reports estimates across nine polynomial order $\times$ kernel function specifications. All produce qualitatively identical results: negative but statistically insignificant estimates for both adoption and citation outcomes. Following Gelman and Imbens (2019), I avoid polynomials of order higher than 3. Excluding conference deadline months (January for ICML, May for NeurIPS, September for ICLR) does not materially change the estimates.

6.8 Heterogeneity

6.8.1 By arXiv Category

Table 7 reports separate estimates for each AI/ML category. The effect may differ across categories if some subfields have larger daily listings (more position variation) or different attention patterns. For example, cs.LG (Machine Learning) has the largest daily volume and may exhibit stronger position effects simply because there are more papers competing for attention. Conversely, cs.IR (Information Retrieval) is a smaller category where most active researchers may read all daily listings, reducing the importance of position.

Category-level estimates must be interpreted with caution because splitting the sample reduces statistical power. I exclude cs.CL and cs.LG from the category-specific table due to extremely small effective samples ($N < 15$) that produce unreliable RDD estimates with

Table 7: Heterogeneity by arXiv Category

	(1) Pooled	(2) stat.ML	(3) cs.CV
Panel A: Log(3yr Citations + 1)			
After cutoff	−1.086	−0.929	−1.243
	(0.733)	(1.568)	(2.156)
	[0.139]	[0.553]	[0.564]
Eff. N	86	17	22
Panel B: Any Frontier Citation			
After cutoff	−0.0183	−0.0214	−0.0128
	(0.0524)	(0.0643)	(0.0478)
	[0.726]	[0.739]	[0.789]
Eff. N	86	17	22

Notes: Column (1) reports the pooled estimate across all categories (identical to Tables 2–3). Columns (2)–(3) report category-specific estimates for stat.ML and cs.CV, the two categories with sufficient effective sample sizes ($N \geq 15$) for separate estimation. Categories cs.AI, cs.CL, and cs.LG are excluded from the category-specific analysis: cs.AI dominates the pooled sample (over 80% of observations in the RDD bandwidth are classified as cs.AI, making the cs.AI-specific estimate essentially identical to the pooled estimate), while cs.CL and cs.LG have extremely small effective samples ($N < 15$) that produce unreliable estimates. Robust bias-corrected standard errors in parentheses; p-values in brackets.

implausibly large standard errors. For the remaining categories, I focus on sign consistency and relative magnitudes rather than individual statistical significance.

6.8.2 By Day of Week

Thursday submissions receive 3-day front-page exposure (the announcement stays up Saturday–Monday) while Monday–Wednesday submissions get 1-day exposure. If visibility drives adoption, Thursday effects should be larger because the position advantage persists for more days. Table 10 in the appendix reports the split estimates. Neither subsample shows a statistically significant effect, and the point estimates are similar across subsamples (−0.018 for Mon–Wed vs. −0.012 for Thu–Fri), providing no evidence that the duration of front-page exposure moderates the position effect.

6.8.3 Over Time

I estimate the RDD separately for each year to test whether the visibility premium has changed as arXiv's daily AI/ML listing has grown from 30–50 papers per day (2012) to 100–200 (2024). Two competing hypotheses generate predictions. First, if position effects operate through limited attention (researchers browse the first k papers and stop), then growing listings should increase the position premium because a smaller fraction of papers receives attention. Second, if alternative discovery channels (social media, recommendation systems, personal networks) have improved over time, the position premium should decline because researchers are less reliant on sequential browsing.

Year-specific estimates are noisy due to reduced sample sizes (effective N per year ranges from 5–15 within the MSE-optimal bandwidth), but no year shows a statistically significant adoption effect. The point estimates fluctuate around zero with no discernible time trend, suggesting that the null result is stable across the sample period. This pattern is inconsistent with the hypothesis that position effects have grown with listing volume, and it is also inconsistent with the hypothesis that they have declined as alternative discovery channels improved. Instead, the year-by-year evidence reinforces the main finding: listing position does not have a detectable causal effect on frontier lab adoption.

The time trend in position effects is independently interesting because it speaks to how AI research discovery infrastructure has evolved. The explosive growth of AI/ML preprints since 2017 — driven by the transformer revolution (Vaswani et al., 2017) and subsequent advances (Devlin et al., 2019; Brown et al., 2020) — has created an information overload problem. Whether platform design features like listing order have become more or less important in this environment is an open empirical question.

7. Discussion

7.1 Mechanisms

The RDD estimates capture the *net* effect of improved listing position (which increases visibility) and delayed announcement (which reduces timeliness). Several mechanisms could explain the pattern of results:

Discovery acceleration. If the primary channel is faster discovery, we should see effects on adoption speed (the survival model) that are larger than effects on total adoption probability, because frontier labs will eventually find high-quality papers through alternative channels.

Cumulative advantage. If early visibility triggers cascading attention (Merton, 1968; Wang et al., 2013), listing position effects should persist at longer horizons. The year-by-year estimates provide suggestive evidence on whether the effect accumulates or dissipates.

Alternative discovery channels. Social media (especially Twitter/X), recommendation algorithms, and personal networks provide alternative discovery channels that may partially substitute for arXiv listing position. The existence of these channels would attenuate the visibility premium, pushing the RDD estimate toward zero. This would not mean visibility is irrelevant — only that arXiv listing position is one of many signals in a multi-channel discovery environment.

7.2 Comparison with Prior Work

The most directly comparable study is Feenberg et al. (2017), who randomize position in NBER working paper digest emails and find a 20–30% increase in downloads for first-listed papers. Two differences are important. First, the NBER design randomizes position *holding timing fixed*, while the arXiv cutoff bundles position improvement with a one-day delay. Second, the NBER outcome is downloads (a measure of attention), while my outcomes include citations and frontier adoption (measures of intellectual incorporation). The comparison informs whether position effects on attention translate into position effects on downstream scientific activity.

The descriptive evidence from Haque and Ginsparg (2009, 2010) documents large correlations between listing position and citations on arXiv. These correlations likely reflect a combination of causal position effects and selection (strategic timing by higher-quality authors). The present study isolates the causal component.

7.3 Welfare and Policy Implications

If listing position materially affects which papers frontier labs adopt, several policy implications follow:

Randomization. ArXiv could randomize listing order within each daily batch, eliminating the position premium entirely. This would equalize attention across papers regardless of submission timing, at the cost of removing any incentive to submit early in the window.

Algorithmic recommendation. ArXiv could supplement chronological ordering with recommendation algorithms that match papers to researchers based on content similarity. This would reduce reliance on position-driven attention while preserving the chronological listing as a default.

Decoupling position from timing. The key design feature driving the position discontinuity is that position and timing are bundled: you can only be first-listed by submitting after the cutoff, which also delays your announcement by one day. Decoupling these features — for example, by randomizing within-batch order — would allow researchers to be both timely and visible.

7.4 External Validity

Three limitations on external validity are worth noting. First, the results apply to AI/ML papers on arXiv and may not generalize to other fields with different attention patterns (e.g., physics, where arXiv has been the norm for decades and reading habits may differ). Second, the effects may be heterogeneous across paper quality: listing position may matter more for marginal papers that would otherwise go unnoticed than for breakthrough results that attract attention regardless of position. Third, the results capture the reduced form of a specific institutional mechanism; a different platform design (e.g., topical sorting, algorithmic recommendation) would alter the relationship between position and visibility.

7.5 Limitations

Several limitations qualify the interpretation. The match rate to citation databases, while substantially improved over the previous version, remains imperfect; unmatched papers may have different characteristics. The frontier lab classification relies on substring matching of institutional affiliations and may miss some citations (e.g., from researchers who have moved between industry and academia) or include false positives. The Cox model within the RDD bandwidth is a local estimate and may not generalize to papers far from the cutoff. The running variable (minutes from cutoff) is measured with some noise due to server processing lags, which would attenuate the RDD estimates toward zero through fuzzy assignment.

8. Conclusion

ArXiv's 14:00 ET cutoff creates a 70-percentage-point jump in listing position, bundled with a one-day announcement delay. Despite this large compound treatment, I find no statistically significant effect on frontier lab adoption, adoption speed, or general citations. The design can rule out very large effects (exceeding 140% of the baseline adoption rate) but cannot detect moderate ones.

This null has a substantive interpretation. Only one in twenty arXiv papers in AI/ML ever reaches the desks of Google, OpenAI, or Meta; for those that do, the average lag is nearly ten months. In this environment, listing position appears to be a weak force. The most capable research organizations in AI have built discovery infrastructure — internal reading groups, automated paper tracking tools, social media monitoring, and recommendation systems — that makes them less reliant on sequential arXiv browsing than the average researcher. If this interpretation is correct, then arXiv's listing algorithm is less consequential for the pace of frontier AI development than critics suggest.

The policy implication is nuanced. Randomizing within-batch order — a low-cost platform redesign that arXiv has discussed but not implemented — would eliminate any residual position effects without imposing the announcement delay that comes with crossing the cutoff. This paper cannot directly evaluate that counterfactual, because the identified variation bundles position with delay. But the null result suggests that even if position effects exist in isolation, they are not large enough to detectably alter frontier lab behavior when combined with a one-day delay.

The analysis also illustrates a fundamental challenge in the science of science: the outcomes that matter most (frontier lab adoption, with a 5% base rate) are the hardest to detect. Future work should pursue larger samples — either by expanding the arXiv categories studied or by exploiting platform-level data (pageviews, downloads, social media mentions) that have higher base rates and more statistical power. Whether AI research is a meritocracy of ideas, where quality eventually overcomes platform frictions, or whether information architecture systematically shapes which methods get built on, remains an open question.

References

Barabási, Albert-László and Réka Albert, "Emergence of Scaling in Random Networks," *Science*, 1999, *286* (5439), 509–512.

Biasi, Barbara and Petra Moser, "Effects of Copyrights on Science: Evidence from the WWII Book Republication Program," *American Economic Journal: Microeconomics*, 2021, *13* (4), 33–60.

Bloom, Nicholas, Charles I. Jones, John Van Reenen, and Michael Webb, "Are Ideas Getting Harder to Find?," *American Economic Review*, 2020, *110* (4), 1104–1144.

Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell et al., "Language Models Are Few-Shot Learners," *Advances in Neural Information Processing Systems*, 2020, *33*, 1877–1901.

Calonico, Sebastian, Matias D. Cattaneo, and Max H. Farrell, "Optimal Bandwidth Choice for Robust Bias-Corrected Inference in Regression Discontinuity Designs," *Econometrics Journal*, 2020, *23* (2), 192–210.

_ , _ , and Rocío Titiunik, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 2014, *82* (6), 2295–2326.

Cattaneo, Matias D., Brigham R. Frandsen, and Rocío Titiunik, "Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the US Senate," *Journal of Causal Inference*, 2015, *3* (1), 1–24.

— , **Michael Jansson, and Xinwei Ma**, "Simple Local Polynomial Density Estimators," *Journal of the American Statistical Association*, 2020, *115* (531), 1449–1455.

— , **Nicolás Idrobo, and Rocío Titiunik**, *A Practical Introduction to Regression Discontinuity Designs: Foundations*, Cambridge University Press, 2020.

Cockburn, Iain M., Rebecca Henderson, and Scott Stern, "The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis," *The Economics of Artificial Intelligence: An Agenda*, 2019, pp. 115–146.

Cox, David R., "Regression Models and Life-Tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1972, *34* (2), 187–202.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of NAACL-HLT*, 2019, pp. 4171–4186.

Dietrich, J. P., "The Importance of Being First: Position Dependent Citation Rates on arXiv:astro-ph," *Publications of the Astronomical Society of the Pacific*, 2008, *120* (869), 801–804.

Feenberg, Daniel, Ina Ganguli, Patrick Gaulé, and Jonathan Gruber, "It's Good to Be First: Order Bias in Reading and Citing NBER Working Papers," *Review of Economics and Statistics*, 2017, *99* (1), 32–39.

Fortunato, Santo, Carl T. Bergstrom, Katy Börner, James A. Evans, Dirk Helbing, Staša Milojević, Alexander M. Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi et al., "Science of Science," *Science*, 2018, *359* (6379).

Gelman, Andrew and Guido Imbens, "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs," *Journal of Business and Economic Statistics*, 2019, *37* (3), 447–456.

Gerard, François, Miikka Rokkanen, and Christoph Rothe, "Bounds on Treatment Effects in Regression Discontinuity Designs with a Manipulated Running Variable," *Quantitative Economics*, 2020, *11* (3), 839–870.

Ginsparg, Paul, "ArXiv at 20," *Nature*, 2011, *476* (7359), 145–147.

Goldfarb, Avi, Bledi Taska, and Florenta Teodoridis, "Could Machine Learning Be a General Purpose Technology? Evidence from Online Job Postings," *Management Science*, 2023, *69* (10), 6172–6191.

Goodfellow, Ian J., Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative Adversarial Nets," *Advances in Neural Information Processing Systems*, 2014, *27.*

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

Jones, Benjamin F., "The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder?," *Review of Economic Studies*, 2009, *76* (1), 283–317.

Klinger, Julia, Juan Mateos-Garcia, and Konstantinos Stathoulopoulos, "Deep Learning, Deep Change? Mapping the Evolution of Artificial Intelligence Research," *Scientometrics*, 2020, *123*, 1087–1111.

Lee, David S. and Thomas Lemieux, "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 2010, *48* (2), 281–355.

Merton, Robert K., "The Matthew Effect in Science," *Science*, 1968, *159* (3810), 56–63.

Nagaraj, Abhishek, Esther Shears, and Mathijs de Vaan, "Improving Data Access Democratizes and Diversifies Science," *Proceedings of the National Academy of Sciences*, 2020, *117* (38), 23490–23498.

Priem, Jason, Heather Piwowar, and Richard Orr, "OpenAlex: A Fully-Open Index of Scholarly Works, Authors, Venues, Institutions, and Concepts," *arXiv preprint arXiv:2205.01833*, 2022.

Sinatra, Roberta, Dashun Wang, Pierre Deville, Chaoming Song, and Albert-László Barabási, "Quantifying the Evolution of Individual Scientific Impact," *Science*, 2016, *354* (6312).

Stephan, Paula E., "The Economics of Science," *Journal of Economic Literature*, 1996, *34* (3), 1199–1235.

ul Haque, Asif and Paul Ginsparg, "Positional Effects on Citation and Readership in arXiv," *Journal of the American Society for Information Science and Technology*, 2009, *60* (11), 2203–2218.

_ and _ , "Last but Not Least: Additional Positional Effects on Citation and Readership in arXiv," *Journal of the American Society for Information Science and Technology*, 2010, *61* (12), 2381–2388.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin, "Attention Is All You Need," *Advances in Neural Information Processing Systems*, 2017, *30.*

Wang, Dashun, Chaoming Song, and Albert-László Barabási, "Quantifying Long-Term Scientific Impact," *Science*, 2013, *342* (6154), 127–132.

A. Data Appendix

A.1 arXiv API Data Collection

The arXiv Atom API was queried with category-month date windows (e.g., `cat:cs.LG AND submittedDate:[20230101 TO 20230131]`). For each category $\times$ month cell, the query paginates through all results in batches of 200, with no per-category cap. The API returns exact submission timestamps in UTC, which I convert to Eastern Time for the running variable construction. Checkpointing after each category allows resumption of interrupted fetches. The query covers six categories (cs.AI, cs.CL, cs.LG, stat.ML, cs.CV, cs.IR) across 13 years (2012–2024), producing 78 category-year strata.

A.2 Citation Matching

Semantic Scholar matching uses native arXiv ID lookup (`ARXIV:<id>`), which provides higher coverage than DOI-based matching because many arXiv papers lack DOIs until formal publication. OpenAlex matching uses DOI lookup as primary (`10.48550/arXiv.<id>`) and title-based fuzzy matching as fallback. For fuzzy matching, I compute word-level Jaccard similarity between the query title and the candidate title, accepting matches with similarity > 0.7. Match rates are reported by year and category in the main text.

A.3 Industry Citation Classification

For each paper with an OpenAlex ID, I query the complete set of citing works and extract the institutional affiliations of each citing paper's authors. I classify institutions by substring matching against the lab lists defined in Section 4.3. For highly cited papers (more than 200 citing works), the query paginates through up to 25 pages (5,000 citing works). The adoption lag is computed as the number of days between the focal paper's publication date and the earliest citing paper with a frontier lab affiliation.

B. Identification Appendix

B.1 McCrary Density Test

Figure 3 in the main text presents the density of submissions around the cutoff. I implement the formal density test of Cattaneo et al. (2020a) at the 14:00 ET cutoff. The test fails to reject the null of no manipulation using the default bandwidth ($p = 0.672$). Results are robust to fixed bandwidths of ± 15 minutes ($p = 0.581$), ± 30 minutes ($p = 0.714$), ± 60 minutes ($p = 0.693$), and ± 90 minutes ($p = 0.728$).

B.2 Randomization Inference Details

Randomization inference p-values reported in Table 6 Panel C are computed by permuting the outcome variable among observations within the MSE-optimal bandwidth and re-estimating the conventional RDD coefficient. The two-sided p-value is the fraction of permutation coefficients with absolute value exceeding the actual coefficient. I use 500 permutations with a fixed seed for reproducibility.

C. Placebo Cutoff Tests

Table 8 reports RDD estimates at eight alternative cutoff times. Only the real 14:00 ET cutoff produces a significant first-stage effect on listing position. All placebo cutoffs produce insignificant adoption estimates.

Table 8: Placebo Cutoff Tests

Cutoff Time (ET)	Coef	SE	p-value	Eff. N	Bandwidth
Panel A: First Stage (Position Percentile)					
10:00	0.0312	(0.0541)	[0.564]	72	28.4
11:00	−0.0187	(0.0483)	[0.698]	81	31.6
12:00	0.0243	(0.0512)	[0.635]	68	26.8
13:00	−0.0098	(0.0467)	[0.834]	84	33.1
14:00 (real)	**−0.6987**	**(0.0412)**	**[< 0.001]**	**86**	**32.9**
15:00	0.0156	(0.0528)	[0.768]	75	29.7
16:00	−0.0074	(0.0556)	[0.894]	63	24.9
17:00	0.0219	(0.0591)	[0.711]	58	22.7
18:00	−0.0131	(0.0634)	[0.836]	51	20.1
Panel B: Frontier Adopted (18m)					
10:00	0.0087	(0.0412)	[0.833]	72	28.4
11:00	−0.0134	(0.0389)	[0.731]	81	31.6
12:00	0.0201	(0.0423)	[0.634]	68	26.8
13:00	−0.0056	(0.0375)	[0.881]	84	33.1
14:00 (real)	**−0.0158**	**(0.0371)**	**[0.670]**	**86**	**32.9**
15:00	0.0112	(0.0401)	[0.780]	75	29.7
16:00	−0.0043	(0.0437)	[0.922]	63	24.9
17:00	0.0178	(0.0468)	[0.704]	58	22.7
18:00	−0.0095	(0.0512)	[0.853]	51	20.1

Notes: Each row estimates a separate RDD at the labeled cutoff time using the same specification as the main results (local linear, triangular kernel, MSE-optimal bandwidth). Panel A tests whether listing position percentile (0 = top of list, 1 = bottom) changes discontinuously at placebo cutoffs; a negative coefficient means papers move toward the top. Panel B tests whether the primary adoption outcome changes. Only the real 14:00 ET cutoff (bold) produces a significant first-stage effect. Robust bias-corrected standard errors in parentheses; p-values in brackets.

D. Additional Robustness

Table 9 reports sensitivity to polynomial order and kernel function. All nine specifications ($p \in \{1, 2, 3\} \times k \in \{\text{triangular, Epanechnikov, uniform}\}$) produce qualitatively identical results: negative but statistically insignificant estimates for both the adoption and citation outcomes. Conference deadline month exclusion (January, May, September) produces similarly null results and is omitted for brevity.

Table 9: Sensitivity to Polynomial Order and Kernel Function

Polynomial	Kernel	Coef	SE	p-value	Eff. N	Bandwidth
Panel A: Frontier Adopted (18m)						
Linear ($p = 1$)	Triangular	−0.0158	(0.0371)	[0.670]	86	32.9
Linear ($p = 1$)	Epanechnikov	−0.0163	(0.0378)	[0.666]	83	31.8
Linear ($p = 1$)	Uniform	−0.0147	(0.0392)	[0.707]	79	30.2
Quadratic ($p = 2$)	Triangular	−0.0189	(0.0483)	[0.696]	97	38.7
Quadratic ($p = 2$)	Epanechnikov	−0.0194	(0.0491)	[0.693]	94	37.4
Quadratic ($p = 2$)	Uniform	−0.0178	(0.0507)	[0.725]	90	35.8
Cubic ($p = 3$)	Triangular	−0.0213	(0.0598)	[0.722]	108	44.2
Cubic ($p = 3$)	Epanechnikov	−0.0221	(0.0612)	[0.718]	105	42.8
Cubic ($p = 3$)	Uniform	−0.0198	(0.0631)	[0.754]	101	41.1
Panel B: Log(3yr Citations + 1)						
Linear ($p = 1$)	Triangular	−1.086	(0.733)	[0.139]	86	32.9
Linear ($p = 1$)	Epanechnikov	−1.102	(0.748)	[0.141]	83	31.8
Linear ($p = 1$)	Uniform	−1.053	(0.771)	[0.172]	79	30.2
Quadratic ($p = 2$)	Triangular	−1.241	(0.962)	[0.197]	97	38.7
Quadratic ($p = 2$)	Epanechnikov	−1.267	(0.981)	[0.196]	94	37.4
Quadratic ($p = 2$)	Uniform	−1.198	(1.013)	[0.237]	90	35.8
Cubic ($p = 3$)	Triangular	−1.387	(1.218)	[0.255]	108	44.2
Cubic ($p = 3$)	Epanechnikov	−1.412	(1.243)	[0.256]	105	42.8
Cubic ($p = 3$)	Uniform	−1.341	(1.279)	[0.294]	101	41.1

Notes: Each row reports a separate RDD estimate using the labeled polynomial order and kernel function. The first row in each panel reproduces the baseline specification from Tables 3 and 5 (local linear, triangular kernel). Higher polynomial orders use wider MSE-optimal bandwidths and yield larger effective samples. Following Gelman and Imbens (2019), I avoid polynomials of order higher than 3. All specifications use MSE-optimal bandwidth selection (Calonico et al., 2014). Robust bias-corrected standard errors in parentheses; p-values in brackets.

E. Heterogeneity Details

Category-specific estimates are reported in Table 7 for stat.ML and cs.CV, the two categories with sufficient within-bandwidth sample sizes ($N \geq 15$) for separate estimation. As noted in the main text, cs.AI is excluded because it dominates the pooled sample (the cs.AI-specific estimate is essentially identical to the pooled result), while cs.CL and cs.LG have extremely small effective samples ($N < 15$).

Table 10 reports estimates separately for Monday–Wednesday and Thursday–Friday subsamples.

Table 10: Heterogeneity by Day of Week

	(1) Mon–Wed	(2) Thu–Fri	(3) Pooled
Panel A: Frontier Adopted (18m)			
After cutoff	−0.0182	−0.0121	−0.0158
	(0.0467)	(0.0583)	(0.0371)
	[0.697]	[0.836]	[0.670]
Eff. N	52	34	86
Panel B: Log(3yr Citations + 1)			
After cutoff	−1.231	−0.874	−1.086
	(0.912)	(1.187)	(0.733)
	[0.177]	[0.462]	[0.139]
Eff. N	52	34	86

Notes: Column (1) restricts to papers submitted Monday–Wednesday (1-day front-page exposure). Column (2) restricts to papers submitted Thursday–Friday (2–3-day front-page exposure due to weekend). Column (3) reproduces the pooled estimate from Tables 3 and 5. If visibility drives adoption, Thursday–Friday effects should be larger because the position advantage persists for more days. Robust bias-corrected standard errors in parentheses; p-values in brackets.

Acknowledgements

This paper was autonomously generated as part of the Autonomous Policy Evaluation Project (APEP).

Contributors: @olafdrw

First Contributor: https://github.com/olafdrw

Project Repository: https://github.com/SocialCatalystLab/ape-papers