# Agentic AI and the Industry Composition of Developer Work: Revised Results Draft*

Alexander Quispe Kevin Xu

July 2026

**Abstract**

Around first detectable Claude Code use, the sectoral reach of developers' public GitHub output expands sharply: monthly distinct industries nearly triple relative to baseline, and entries into never-seen sectors spike to 3.5 times the baseline flow. The expansion is not the mechanical footprint of the treatment-defining commit — 78 percent of the entry effect survives when outcomes are rebuilt from non-Claude commits only — and it is not a pure activity artifact: entries rise conditional on activity, per repository touched, and after residualizing on commits, repositories, and active days. Placebo adoption dates among later adopters show nothing. Consistent with a model in which agentic delegation lowers the fixed cost of entering unfamiliar domains, first entries concentrate among pre-adoption *specialists* — two to four times the generalist response at every activity level. A mechanical depletion simulation shows the subsequent negative entry flow is retiming (entries pulled forward), not exhaustion of the 19-sector menu; the cumulative sectoral footprint remains permanently higher (∼50 percent), and post-adoption entries persist nearly as long as calendar-matched organic entries. Estimates are event-time effects of adoption among eventual adopters; classifier validation by human audit remains outstanding.

*Revision responding to the internal referee report of July 2026. All estimates July 10, 2026, pipeline steps 33–60. The human classifier audit and the README-contamination reclassification (referee §4.5–4.6) were not executed in this round and are stated as limitations in Section 9.

# Contents

# 1 Data

## 1.1 Sample and treatment timing

The developer sample, treatment dating, and commit-level contribution histories are inherited from the companion language paper. The estimation sample contains 5,346 developers: 2,813 treated developers whose first Claude-co-authored commit — detected through the machine-readable `Co-Authored-By` trailer — falls in April–September 2025, and 2,533 later adopters (first Claude commit October 2025–January 2026) who serve as not-yet-treated comparisons before their own adoption. The sample is screened for automation accounts and for detectable use of competing agentic tools before adoption. The panel window covers 28 months, January 2024 through April 2026; after the pre-period activity filter, estimation uses 5,332 developers.

## 1.2 From commits to repositories

Because industry attaches to a repository, the unit of measurement is the developer–repository pair. Within the window, sampled developers author 2.76 million non-merge commits across 85,048 developer–repository pairs and 84,565 unique repositories (median 10 repositories per developer; 78 percent of pairs are own repositories; 4,907 repositories are first touched exactly in the adoption month, motivating the mechanical exclusions below).

## 1.3 Repository metadata and industry classification

Metadata (name, description, topics, README, creation date, fork status, primary language) were fetched for all repositories in one batched GraphQL pass; 99.8 percent resolve. Each repository is classified into one of 19 two-digit NAICS sectors by our fine-tuned RoBERTa-large classifier (6,588 GPT-labeled training repositories; held-out F1 0.863), consuming name/description/topics/README (first 5,000 characters). The sector distribution is concentrated but not degenerate (Professional, Scientific, and Technical Services 27.1 percent; Information 23.1; Education 10.3; Finance 8.6; Arts and Entertainment 6.4). Main outcomes count confident classifications (top-1 probability $\geq 0.5$; 63.3 percent of repositories); the unthresholded variant tracks the main results closely. For external entries we additionally fetched onboarding-friction fields (stars, forks, size, contributor pool, `CONTRIBUTING.md`, CI configuration) for the 18,471 external repositories.

### 1.4 Outcome construction

Outcomes are built on a balanced developer $\times$ month skeleton mirroring the companion paper: monthly distinct industries, newly-entered industries (sectors absent from the developer's entire observed history), cumulative industries, and commit-weighted industry entropy. The exploration ladder refines the entry flow: of 21,011 first developer–sector contacts, 17,747 occur through clean repositories (non-fork, not the first Claude repository), 11,122 meet a substantive bar ($\geq$5 clean commits or $\geq$2 active months within three months of entry), and 3,397 are confirmed by two or more distinct repositories.

## 2 Empirical design and estimand

We estimate group-time average treatment effects following Callaway and Sant'Anna (2021),

$$ATT(g,t) \;=\; \mathbb{E}[\,Y_t(g) - Y_t(0)\,|\,G = g\,]\,,$$

aggregated to event time, where treatment is *first detectable Claude Code adoption* and comparisons are not-yet-treated eventual adopters. The estimand is therefore the effect of adoption on the sectoral composition of public GitHub work *among eventual adopters*, relative to not-yet-treated adopters — not the population effect of agentic AI. Specification throughout: doubly robust estimation, varying base period, one month of anticipation (the trailer is a slightly late measure of first use), 1,000 bootstrap iterations.

Four limitations qualify the estimand. Treatment timing is a lower bound (trailers can be disabled), attenuating estimates toward zero. Adoption is voluntary and possibly timed around project transitions; Section 4 shows the results survive the observable versions of this concern, but selection on unobserved opportunities cannot be excluded. The sample is not representative of all developers. And outcomes measure the industry orientation of public output, not labor-market mobility.

## 3 Main results

Table 1 and Figure 1 report the primary event studies. At adoption, monthly distinct industries rise by 0.71 against a pre-adoption mean of 0.36; never-seen-sector entries rise by 0.31 against a baseline flow of 0.12; industry entropy rises by 0.11 against 0.05; and the cumulative sectoral stock jumps by 0.48 and plateaus at roughly +0.66 — a permanent ∼50 percent expansion of lifetime sectoral reach — without reversion. Pre-adoption coefficients are economically negligible from $e = -6$ to $e = -2$; the small positive $e = -1$ coefficient is consistent with late measurement of first use and is handled by the anticipation allowance, and Section 4 shows the conclusions do not depend on it.

**Table 1:** Main results: industry outcomes around Claude Code adoption

| | Pre-mean | ATT($e$=0) | ATT($e$=1) | Avg $e$ 0–2 | Avg $e$ 1–4 | Ratio |
|---|---|---|---|---|---|---|
| Distinct industries | 0.363 | 0.710*** | 0.386*** | 0.429*** | 0.168*** | 1.96 |
| | | (0.020) | (0.021) | (0.020) | (0.022) | |
| Newly-entered industries | 0.122 | 0.308*** | 0.014 | 0.090*** | -0.063*** | 2.52 |
| | | (0.015) | (0.013) | (0.011) | (0.013) | |
| Cumulative industries | 1.297 | 0.481*** | 0.582*** | 0.563*** | 0.632*** | 0.37 |
| | | (0.017) | (0.021) | (0.020) | (0.027) | |
| Industry entropy | 0.048 | 0.111*** | 0.074*** | 0.075*** | 0.037*** | 2.32 |
| | | (0.006) | (0.006) | (0.005) | (0.005) | |

*Notes:* Callaway–Sant'Anna doubly robust estimates, not-yet-treated controls, anticipation of one month, 1,000 bootstrap iterations. Averages over event times are influence-function aggregations, not means of point estimates. Ratio is ATT($e$=0) divided by the pre-adoption mean. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

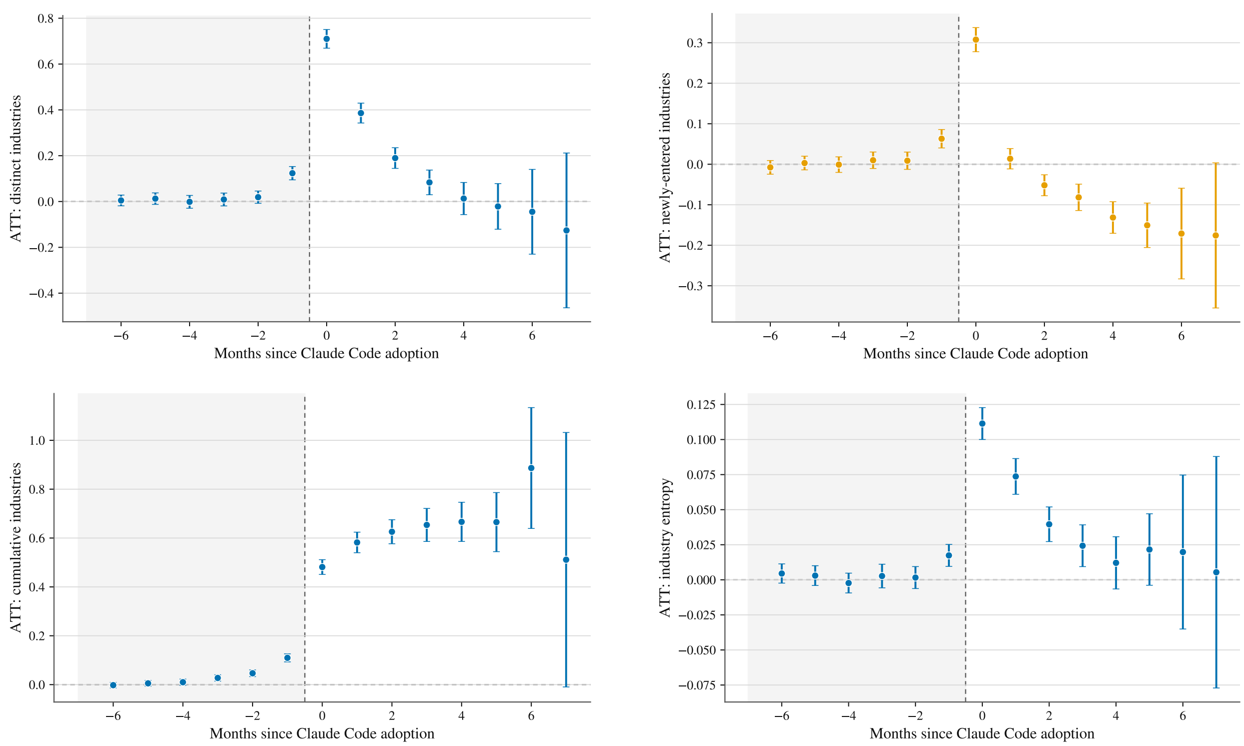



**Figure 1:** Event-study estimates, primary industry outcomes. Shaded region: pre-adoption; bars: 95% pointwise confidence intervals.

**Dynamics: a retimed burst with a permanent footprint.** The entry flow turns negative from $e = 2$ ($-0.05$ to $-0.17$ by $e = 6$). A mechanical simulation disciplines the interpretation (Figure 2): calibrating developer-specific entry hazards from pre-adoption behavior (mean hazard 0.008 per unseen sector-month over a mean 16.5 unseen sectors) and removing the observed adoption-month entries from each developer's reservoir, pure menu depletion generates a post-adoption flow deficit of only $-0.0015$ per month — one to three percent of the observed dip. The negative flow is therefore *not* exhaustion of the 19-sector menu; it is predominantly retiming: the adoption-month burst pulls forward entries that would otherwise have occurred over the following months, together with composition effects across cohorts at long horizons. The permanent interpretation accordingly rests on the cumulative stock, which remains $+0.63$ (s.e. 0.027) on average over $e = 1$–4 — after the adoption month is excluded entirely.

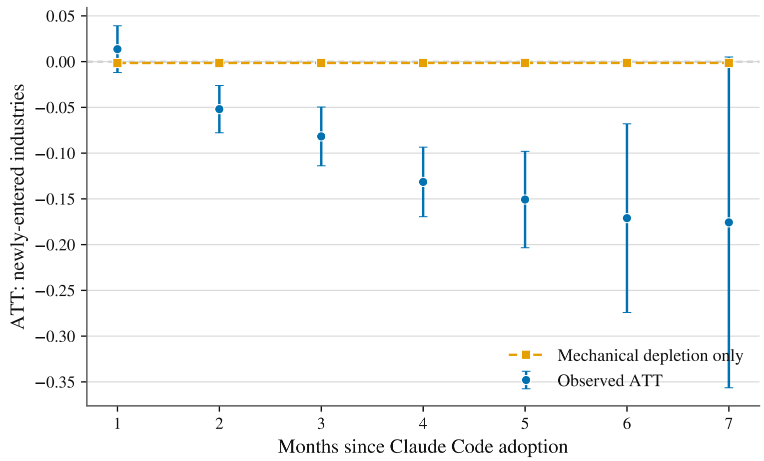



**Figure 2:** Observed post-adoption entry-flow ATTs versus the flow deficit implied by mechanical reservoir depletion alone, calibrated on pre-adoption entry hazards. Depletion accounts for 1–3 percent of the observed dip.

# 4 Robustness: mechanical exposure, activity, and timing

## 4.1 The effect appears in unassisted work

Treatment is dated by a commit that itself carries an industry. Table 2 rebuilds every outcome from purged commit streams. Using *only commits without a Claude co-author trailer*, never-seen-sector entries rise 0.239 (s.e. 0.013) at adoption — 78 percent of the baseline effect — and the cumulative stock remains elevated (avg $e$ 1–4: $+0.53$). Excluding every commit to the first Claude repository leaves an entry effect of 0.123 (s.e. 0.013) and a cumulative effect of $+0.46$. Figure 3 overlays the baseline and non-Claude event studies: adoption changes the sectoral composition of the developer's broader portfolio, not merely the treated commits.

**Table 2:** Mechanical-exposure robustness: outcomes rebuilt from purged commit streams

| ATT($e$=0), then avg $e$ 1–4 | Baseline (1) | Non-Claude commits (2) | Excl. first Claude repo (3) |
|---|---|---|---|
| Distinct industries | 0.710*** | 0.576*** | 0.318*** |
| | (0.020) | (0.019) | (0.019) |
| avg $e$ 1–4 | 0.168*** | 0.082*** | 0.137*** |
| | (0.022) | (0.022) | (0.023) |
| Newly-entered industries | 0.308*** | 0.239*** | 0.123*** |
| | (0.015) | (0.013) | (0.013) |
| avg $e$ 1–4 | -0.063*** | -0.074*** | -0.031*** |
| | (0.013) | (0.013) | (0.011) |
| Cumulative industries | 0.481*** | 0.410*** | 0.256*** |
| | (0.017) | (0.016) | (0.015) |
| avg $e$ 1–4 | 0.632*** | 0.527*** | 0.459*** |
| | (0.027) | (0.028) | (0.026) |
| Industry entropy | 0.111*** | 0.095*** | 0.060*** |
| | (0.006) | (0.006) | (0.005) |
| avg $e$ 1–4 | 0.037*** | 0.023*** | 0.029*** |
| | (0.005) | (0.006) | (0.005) |

*Notes:* Column (2) rebuilds all outcomes using only commits without a Claude co-author trailer; column (3) excludes every commit to the repository of the developer's first Claude-co-authored commit. Estimation sample and specification identical to Table 1; the pre-activity filter uses baseline commits in all columns. Pre-adoption means (baseline/non-Claude/purged): distinct industries 0.363/0.362/0.340; newly-entered 0.122/0.122/0.117. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

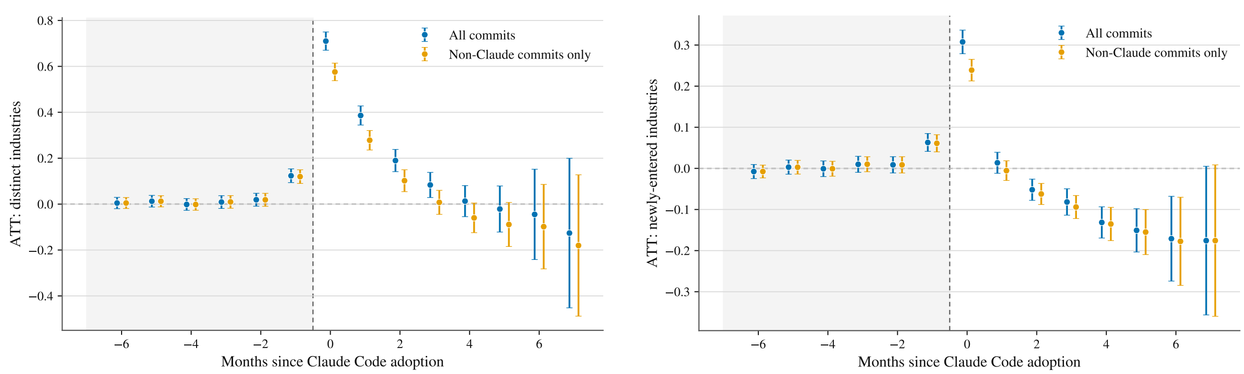



**Figure 3:** Baseline versus non-Claude-commit outcomes.

## 4.2 Not an activity artifact

Activity does surge at adoption (+4.1 active days, +1.09 newly touched repositories, +1,198 files changed at $e = 0$), so the central alternative — more activity mechanically reveals more sectors — must be confronted directly. Three results reject it as the full explanation. The probability of entering at least one new sector rises 24.0 percentage points against a baseline of 10.1 percent. Sectors per repository touched rise 0.267 against a baseline of 0.179 — the *marginal* repository is more sectorally diverse, not just more numerous. And after residualizing outcomes on $\log(1 + \text{commits})$, $\log(1 + \text{repositories})$, and $\log(1 + \text{active days})$ with developer and month fixed effects, never-seen-sector entries still rise 0.059 (s.e. 0.014) at adoption. The honest quantification: most of the raw *count* response is activity-scale, but the direction of marginal activity shifts toward unfamiliar sectors at fixed volume.

### 4.3 Timing, placebo, and cohorts

Shifting the adoption date one month earlier moves the spike with it (the shifted $e = 0$, i.e. the true $e = -1$, shows only $+0.07$). Assigning placebo adoption dates six months before later adopters' true adoption — estimated entirely within their pre-adoption periods — produces no positive effects ($-0.070$ and $-0.076$ for entries and distinct industries; the small negatives likely reflect the truncated placebo window). Leave-one-adoption-cohort-out estimates range 0.307–0.323, and per-cohort ATTs are reported in the appendix tables. Restricting aggregation to $e \geq 1$ (dropping both $e = -1$ and the adoption month) leaves distinct industries $+0.17$ and the cumulative stock $+0.63$ on average over $e$ 1–4.

## 5 The exploration ladder

Table 3 and Figure 4 re-estimate the entry margin under increasingly stringent definitions. Substantive entries — requiring the first sector contact to occur through a clean repository and at least five clean commits or two active months within three months — rise 0.078 (s.e. 0.010) against a baseline of 0.064; two-repo-confirmed entries rise 0.023 (s.e. 0.006) against 0.021. Both are roughly a doubling. The attenuation of the proportional effect from 3.5× (contact) to 2.1–2.2× quantifies the light-touch share of the raw spike: about half. The ladder builds the mechanical exclusions (forks, first Claude repository) into the variable definition itself.

**Table 3:** The exploration ladder: event-study estimates under increasingly stringent definitions of new-industry entry

| | Contact (1) | Substantive (2) | Two-repo (3) | Subst. own repo (4) | Subst. joined (5) |
|---|---|---|---|---|---|
| $e = -3$ | 0.010 | 0.002 | -0.002 | -0.000 | 0.002 |
| | (0.010) | (0.007) | (0.004) | (0.007) | (0.002) |
| $e = -2$ | 0.009 | 0.013* | 0.009** | 0.011 | 0.002 |
| | (0.009) | (0.007) | (0.004) | (0.006) | (0.003) |
| $e = -1$ | 0.063*** | 0.018** | 0.009** | 0.018** | -0.000 |
| | (0.012) | (0.007) | (0.005) | (0.008) | (0.003) |
| $e = 0$ | 0.308*** | 0.078*** | 0.023*** | 0.068*** | 0.010*** |
| | (0.015) | (0.010) | (0.006) | (0.010) | (0.004) |
| $e = 1$ | 0.014 | 0.023** | -0.001 | 0.021** | 0.002 |
| | (0.013) | (0.010) | (0.006) | (0.009) | (0.003) |
| $e = 2$ | -0.052*** | -0.018* | -0.012** | -0.019** | 0.002 |
| | (0.014) | (0.009) | (0.006) | (0.009) | (0.004) |
| $e = 3$ | -0.082*** | -0.040*** | -0.029*** | -0.035*** | -0.005 |
| | (0.016) | (0.012) | (0.007) | (0.012) | (0.004) |
| $e = 4$ | -0.132*** | -0.070*** | -0.038*** | -0.061*** | -0.009* |
| | (0.020) | (0.015) | (0.009) | (0.014) | (0.005) |
| Pre-adoption mean | 0.122 | 0.064 | 0.021 | 0.052 | 0.012 |
| ATT($e$=0) / pre-mean | 2.52 | 1.22 | 1.13 | 1.31 | 0.85 |

*Notes:* Callaway–Sant'Anna doubly robust event-study estimates (not-yet-treated controls, one month of anticipation, 1,000 bootstrap iterations); standard errors in parentheses. Contact entry (col. 1) counts any confidently classified commit in an industry absent from the developer's entire observed history. Substantive entry (col. 2) additionally requires the first contact to occur through a non-fork repository other than the developer's first Claude repository, and at least five clean commits or two active months in the industry within three months of entry. Two-repo entry (col. 3) further requires at least two distinct repositories classified into the entered industry. Columns 4–5 split substantive entries by whether the entry repository is owned by the developer or an external project. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

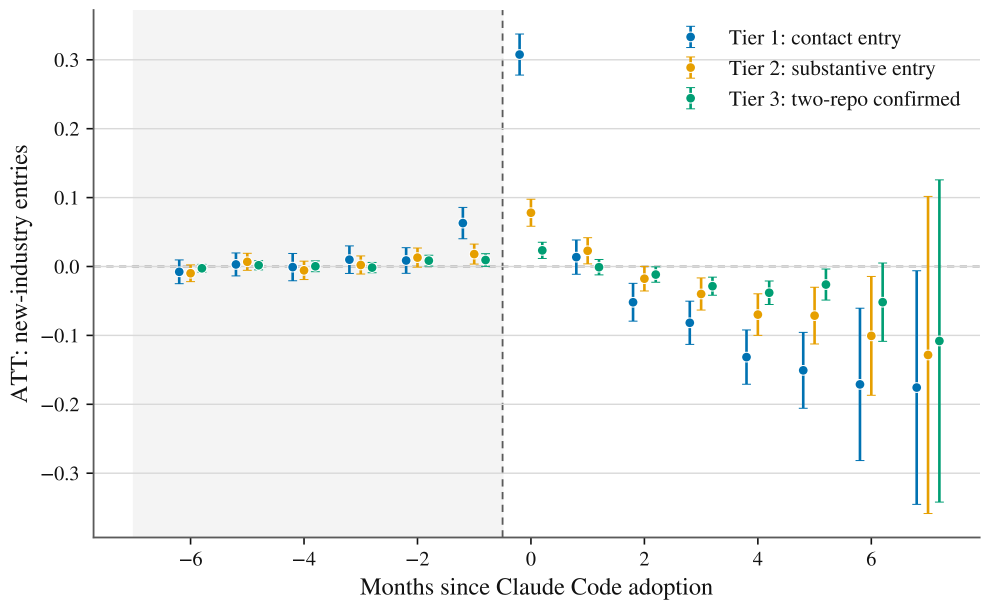



**Figure 4:** Event studies for the exploration ladder.

# 6 Mechanisms

## 6.1 Specialists respond most

The delegation model's sharpest cross-sectional prediction is that entries concentrate among developers with narrow pre-adoption portfolios, who have both headroom and the most to gain from delegated execution. Table 4 and Figure 5 report the double sort. At every activity level, specialists (at most one pre-adoption sector) respond two to four times more strongly than generalists: 0.334 versus 0.115 among high-activity developers; 0.484 versus 0.234 among low-activity developers. A uniform activity shock would not produce this gradient. (Developers with no classified pre-adoption work show the largest raw response, 0.746, as headroom mechanics predict; we do not lean on this cell because its baseline is undefined.)

**Table 4:** Specialist–generalist double sort: newly-entered industries at adoption

| Cell | Treated $N$ | Pre-mean | ATT($e$=0) | (s.e.) | Ratio |
|---|---|---|---|---|---|
| Specialist × high activity | 155 | 0.049 | 0.334*** | (0.049) | 6.8 |
| Specialist × low activity | 665 | 0.048 | 0.484*** | (0.028) | 10.0 |
| Generalist × high activity | 1,136 | 0.211 | 0.115*** | (0.023) | 0.5 |
| Generalist × low activity | 549 | 0.129 | 0.234*** | (0.034) | 1.8 |
| No classified pre-adoption work | 303 | 0.000 | 0.746*** | (0.039) | — |

*Notes:* Specialists have at most one confidently classified pre-adoption sector; generalists two or more; activity split at the median of positive pre-adoption commit counts. Each cell estimated separately with the paper specification. Ratio is ATT($e$=0) over the cell's pre-adoption mean. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

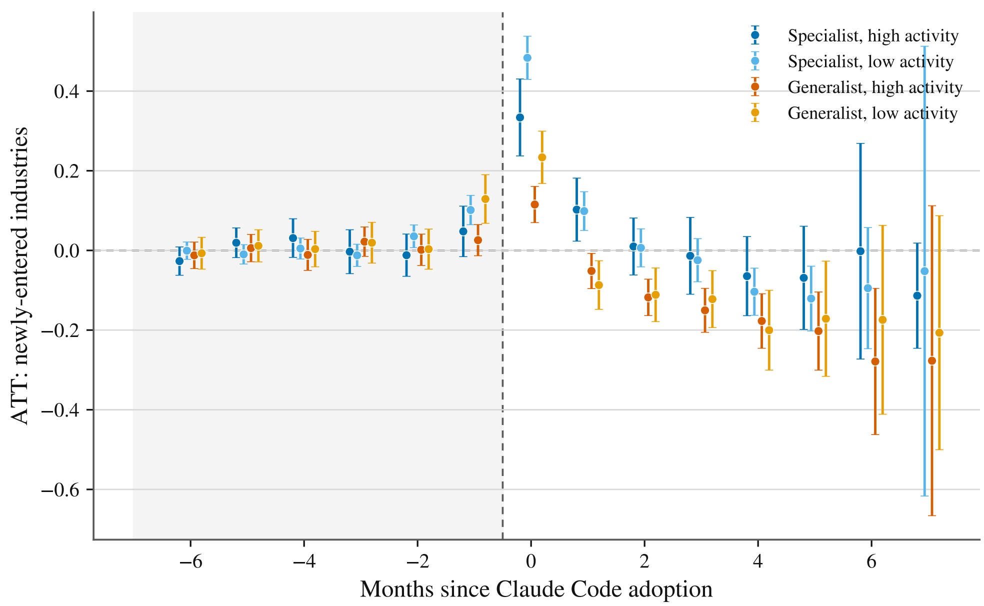



**Figure 5:** Newly-entered industries by pre-adoption specialization and activity.

## 6.2 Distance, re-entry, and external onboarding

Table 5 decomposes the entry response. *Distant versus adjacent*: splitting entries at the median co-occurrence distance between the entered sector and the developer's history, both margins respond strongly (0.146 and 0.141 at $e = 0$); proportionally the tilt is toward distant entries (4.6× versus 3.9× baseline), but the difference is modest — adoption widens exploration across the distance distribution rather than only at the far end. *Re-entries into already-seen sectors* also rise (0.145 at $e = 0$, 3.9× a 0.038 baseline): the tool evidently lowers *re-activation* costs as well as first-entry costs. This tempers a pure unfamiliarity-cost reading; the discriminating evidence for the delegation mechanism is therefore the specialist gradient above, which a generic reactivation-plus-activity shock would not generate. *External onboarding*: substantive entries through external repositories nearly double at adoption on both sides of the friction split (contributor pool above/below median: +0.0050 and +0.0054 against baselines of 0.0066 and 0.0068), with no differential by friction. Two caveats: external entries are rare, and the external repositories developers actually join are predominantly small (median 0–1 stars, 3 mentionable users), so the high-friction margin of large-OSS onboarding is thin in this sample.

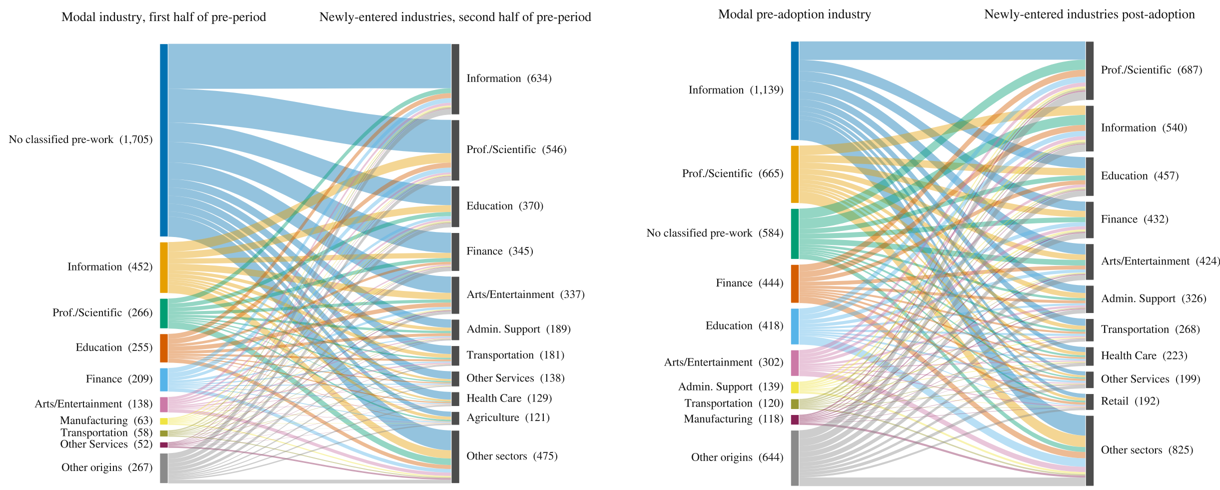
**Table 5:** Mechanism decompositions: entry margins at adoption ($e = 0$)

| Entry margin | Pre-mean | ATT($e$=0) | (s.e.) | Ratio |
|---|---|---|---|---|
| Distant-sector entries | 0.0320 | $0.1460^{***}$ | (0.0096) | 4.6 |
| Adjacent-sector entries | 0.0358 | $0.1412^{***}$ | (0.0086) | 3.9 |
| Re-entries, already-seen sectors | 0.0376 | $0.1449^{***}$ | (0.0102) | 3.9 |
| External entries, high-friction repos | 0.0066 | $0.0049^{**}$ | (0.0021) | 0.7 |
| External entries, low-friction repos | 0.0068 | $0.0056^{**}$ | (0.0025) | 0.8 |

*Notes:* Distant/adjacent split entries at the median pre-adoption co-occurrence distance between the entered sector and the developer's sector history at entry (entries with no prior history excluded from the split). Re-entries are sectors already in the developer's history, dormant for at least three months. External-entry friction split at the median contributor pool of external entry repositories. All rows use the paper specification. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

# 7 Where developers go: flows beyond baseline churn

Figure 6 maps exploration descriptively (79 percent of treated developers with confident post-adoption work enter at least one never-seen sector; 4,573 entries), against the placebo flows within the pre-period. Restricting to developers with a classified base and normalizing by window length, 73.9 percent explore post-adoption (0.167 entries per developer-month) versus 61.6 percent (0.126) in the placebo — despite the post measure facing a harder novelty bar. Table 6 makes the comparison analytical: row-normalized transition rates minus their placebo counterparts identify the over-represented destinations — Arts/Entertainment developers into Finance, Professional/Scientific into Health Care, Education into Finance, and Information into Health Care and Administrative Support, among others.




(a) Placebo: within the pre-period

(b) Around adoption

**Figure 6:** Industry exploration flows, treated developers. One flow unit is one developer–sector entry.

**Table 6:** Top over-represented post-adoption sector transitions relative to baseline churn

| Origin | Destination | Entries | Post share | Δ vs placebo (pp) |
|---|---|---|---|---|
| Arts/Entertainment | Finance | 38 | 0.126 | +3.9 |
| Prof./Scientific | Health Care | 38 | 0.057 | +3.8 |
| Finance | Admin. Support | 33 | 0.074 | +2.6 |
| Education | Finance | 48 | 0.115 | +2.5 |
| Arts/Entertainment | Prof./Scientific | 46 | 0.152 | +2.2 |
| Prof./Scientific | Transportation | 45 | 0.068 | +1.5 |
| Information | Health Care | 59 | 0.052 | +1.4 |
| Information | Admin. Support | 88 | 0.077 | +1.3 |
| Prof./Scientific | Admin. Support | 60 | 0.090 | +1.1 |
| Prof./Scientific | Retail | 33 | 0.050 | +0.8 |

*Notes:* Origins are modal pre-adoption sectors of treated developers with classified base-period work (first-appearance flows excluded). Shares are row-normalized transition rates; the placebo uses the midpoint split of the pre-adoption window (Figure 3a). Cells with fewer than 30 post-adoption entries excluded.

# 8 Persistence

Do adoption-era entries stick? Figure 7 compares survival of substantive entries under a fixed two-month detection window, with two corrections to the naive comparison: pre-adoption entries are kept only if their detection window closes *before* adoption (so treatment cannot prop up measured persistence), and a calendar-matched benchmark uses later adopters' entries in the same calendar months as the treated post-adoption entries. Against the calendar-matched benchmark — the cleanest comparison — post-adoption entries survive at 56.8 versus 59.6 percent after one month and 22.8 versus 26.5 percent after six: a gap of roughly three to four percentage points. Adoption-era entries are marginally more experimental than organic entries under identical calendar conditions, but the great majority of the additional exploration produces engagements with the same qualitative persistence profile — consistent with the permanent rise in the cumulative sectoral stock.

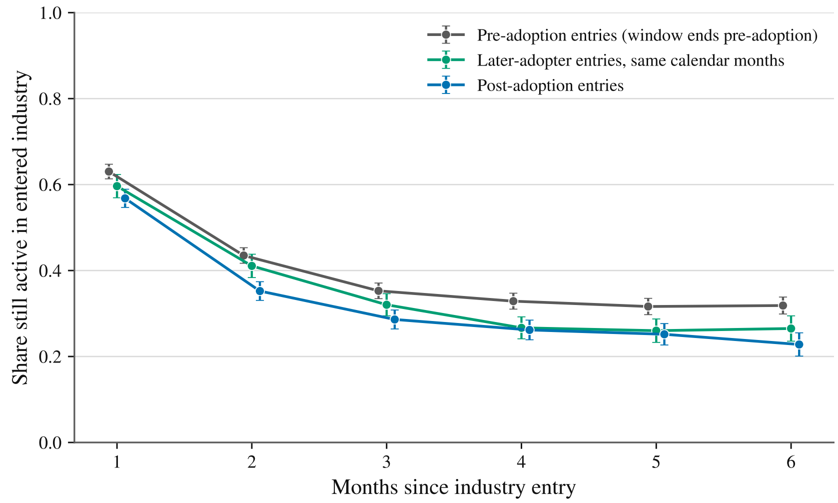



**Figure 7:** Survival of substantive new-sector entries with censoring and contamination corrections. Fixed two-month detection windows; bars are 95% confidence intervals.

# 9 Limitations and outstanding work

Three measurement limitations remain open and are stated rather than resolved. First, the classifier has not yet been validated by human audit on this population; its held-out F1 (0.863) is computed against GPT-generated labels, and 36.7 percent of repositories fall below the confidence threshold. The two-repo ladder tier and the threshold-stability results mitigate but do not replace an audit. Second, repository metadata were fetched after treatment; post-adoption READMEs may be agent-written, and the planned metadata-only reclassification has not been run. Third, diff-based volume controls (lines changed) and full collaboration networks are unavailable by construction of the commit pipeline; files changed and owner counts are the available proxies. Additional caveats: the placebo coefficients are slightly negative; the unfamiliar-language share of external entries is confounded by history length and is reported only descriptively; and dynamic coefficients at long horizons mix cohort composition.

## Summary of the revision

Relative to the referee report: the estimand is stated (§2); the mechanical-exposure suite is complete and the effect survives in unassisted work (§4); the activity-shock alternative is confronted with conditional, per-repository, and residualized outcomes (§4); timing, placebo, and cohort checks are reported (§4); the specialist double sort and the distance decomposition are core results (§6); the depletion interpretation was tested by simulation and *revised* — the post-adoption flow deficit is retiming, not menu exhaustion (§3); flows are benchmarked against placebo churn analytically (§7); and survival is corrected for censoring and treatment contamination (§8). Outstanding: the human classifier audit and README-contamination reclassification (§9).

## References

Callaway, B., and P. H. C. Sant'Anna (2021): "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 225(2), 200–230.