

The Solo-Authorship Reversal

Generative and Agentic AI in Academic Authorship

Hao Ding · Oxford-Man Institute of Quantitative Finance, University of Oxford · hao.ding@omi.ox.ac.uk

AGENTIC AI SUMMIT 2026

UC Berkeley · August 1–2



dinghao.co.uk

+2.8 pp

solo share in **writing-intensive** fields after ChatGPT
— DiD vs. lab fields ($p = 0.011$)

+6.7 pp

solo share in **coding-intensive** fields after the Codex agent, May 2025 — DiD ($p < 0.001$)

−0.47

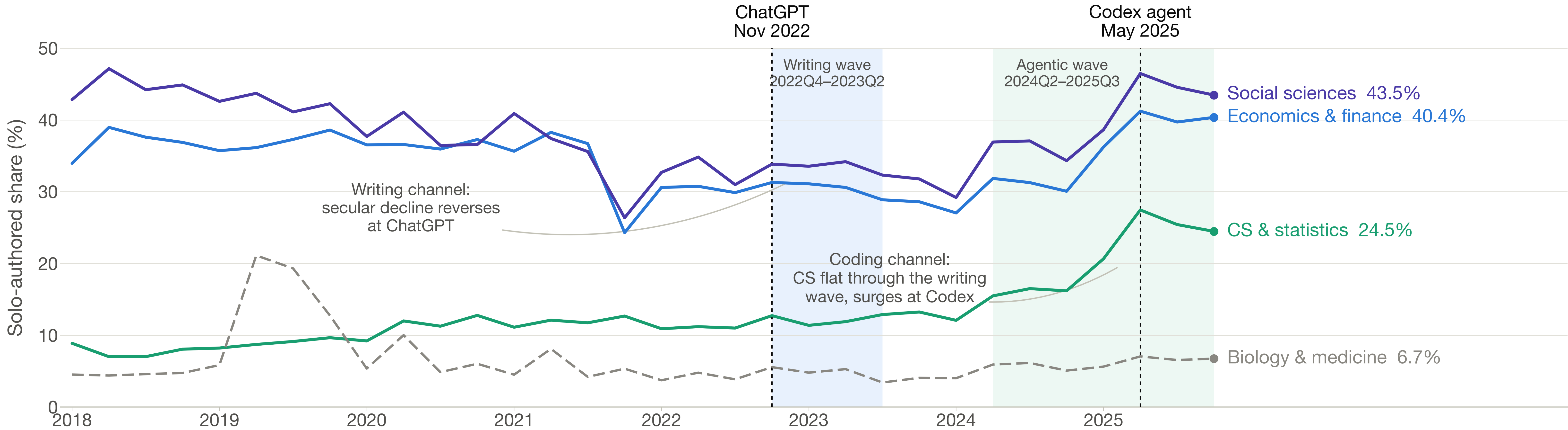
co-authors per economics paper vs. synthetic control (permutation $p < 0.001$)

$r = -0.86$

fields adopting AI writing **most** go solo **least** — substitution is task-specific, not adoption-driven

Two shocks, two task channels — the same substitution mechanism

Fifty years of team science went into reverse on a schedule set by what AI could do: writing-heavy fields turned at ChatGPT; coding-heavy fields waited for a coding agent.



Quarterly solo-authored share, pooled across arXiv, SSRN, bioRxiv/medRxiv, OSF and Scopus (950,053 papers in the four groups shown, 2018–2025). End-point labels are 2025Q4 values; annual means reproduce the paper's results table exactly (CS & statistics: 9.9% pre-ChatGPT → 15.1% in 2024 → 24.7% in 2025). Physics is omitted: its pooled series is distorted by a documented SSRN misclassification; arXiv-only physics appears in the wave table below.

WHY TEAMS — AND WHY NOW

AI competes with the marginal co-author

As knowledge accumulates, researchers specialize narrowly and must collaborate to assemble all the skills a paper needs — Jones's (2009) *burden of knowledge*. Solo authorship accordingly fell for five decades; in economics, from ~50% of papers in 1996 to ~25% by 2014 (Kuld & O'Hagan 2018). LLMs now supply, at near-zero marginal cost, exactly the tasks junior collaborators contributed: drafting, literature synthesis, turning results into prose — and, once agents matured, research code.

Hypothesis. When AI can perform a collaborator's tasks, the marginal value of adding that human falls. Substitution should appear **field by field, task by task, timed to the capability that automates it** — not uniformly across science.

DATA

1.41M records, five platforms

Source	Papers	Role
SSRN	915k	robustness / within-treated
arXiv	185k	primary DiD sample
bioRxiv · medRxiv	210k	lab-science control
PsyArXiv · SocArXiv · OSF	74k	social-science treated
Scopus	30k	published cross-check

Source taxonomies harmonized into six field bins; field-specific 99th-percentile author-count caps; quarterly field-level panel.

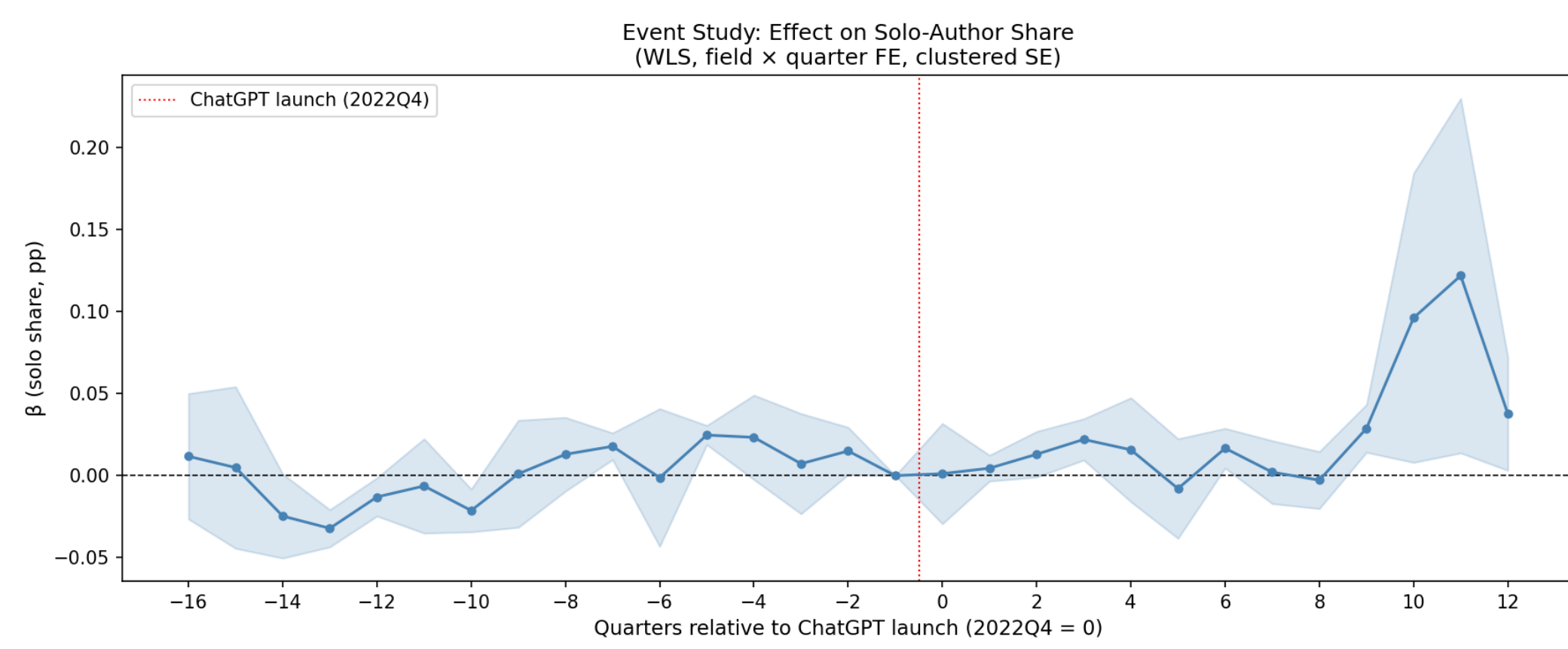
IDENTIFICATION

Field-blind shocks, exposure-based treatment

- **DiD + event study:** writing-intensive (econ · finance · soc sci · CS/stats) vs. lab-intensive fields (bio · med · physics · math) around ChatGPT — field & quarter FE, clustered SE.
- **Synthetic control** for arXiv economics (donors: math, physics, q-bio), permutation inference.
- **Second field-blind shock:** the Codex coding agent (May 2025) hits coding-intensive fields — free of any COVID-era confound.
- **Wave-split DiD** separates the writing wave from the agentic wave; in-space placebos throughout.

WAVE 1 · WRITING CHANNEL — CHATGPT, NOV 2022

The decline turns within a year



Quarterly event-study coefficients on solo share (WLS, field and quarter FE, clustered SE; preprint main sample: arXiv + bioRxiv/medRxiv + PsyArXiv/SocArXiv). Pre-trends flat for 16 quarters; the effect builds after launch and accelerates in the agentic period.

- Baseline DiD: **+2.81 pp** ($p = 0.011$); economics alone **+4.18 pp**; biology & medicine, where lab work resists substitution: cumulative **+1.0 pp**.
- Within-treated SSRN DiD: **+3.66 pp** ($p = 0.014$); interrupted time series **+1.12 pp/quarter** ($p = 0.006$).
- SSRN economics & finance solo share: **38% (2018) → 30% (2022) → 40% (2025)** — the first documented uptick in decades.

WAVE 2 · CODING CHANNEL — CODEX AGENT, MAY 2025

Agentic coding does it again, on cue

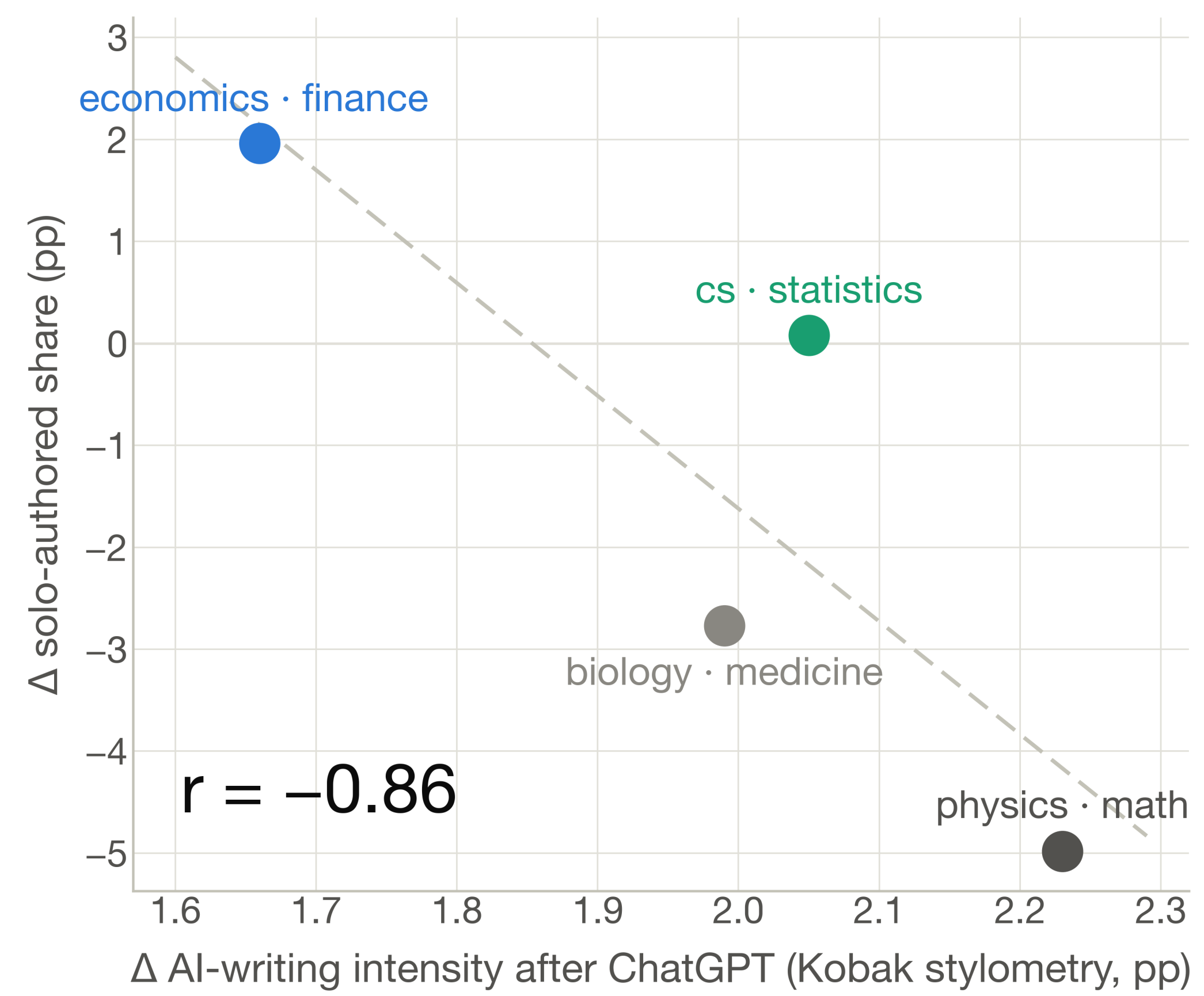
arXiv field	GenAI wave 22Q4–24Q1	Agentic wave 24Q2–25Q3
Economics	+1.8 pp	+1.9 pp
Quant finance	+0.3 pp	+1.2 pp
Statistics / ML	−0.5 pp	+3.3 pp
Computer science	+0.8 pp	+0.5 pp
Physics	+2.3 pp	+1.1 pp
Mathematics	−4.0 pp	−6.9 pp

Δ solo share vs. pre-LLM baseline within arXiv, by wave (selected fields; full heatmap in the paper). Statistics/ML activates **only** when agentic coding arrives.

- Across all platforms, coding-fields DiD after Codex: **+6.70 pp** ($p < 0.001$); CS/stats solo share **15.1% → 24.7%** in twelve months.
- CS shows the **largest cumulative rise of any field (+14.7 pp)** — after showing **no response** to the writing wave.
- Two field groups moving exactly at their task-specific shocks, 30 months apart, is hard to reconcile with COVID recovery or secular trends.

MECHANISM

Substitution for tasks — not AI enthusiasm



Change in field-level AI-writing intensity after ChatGPT (Kobak et al. 2025 stylometry) vs. change in solo share. Physics and biology are among the heaviest AI-writing adopters — and go solo least; the effect runs through whose **tasks** get automated, not who uses AI.

- Author-level stayer panel: writing that turns most AI-like after ChatGPT comes with teams **0.115 co-authors smaller per SD** ($p = 0.031$).
- **Productivity paradox:** AI-exposed authors publish **3.3% fewer papers per year** ($p < 0.001$) — consistent with reallocation toward fewer, more ambitious solo projects, not an output boom.

ROBUSTNESS

Five proxies, one answer

Kobak writing stylometry · author-level GitHub coding intensity · field-level coding classification · staggered university enterprise-ChatGPT adoption · the Codex break — five independent identification proxies all point to task-specific substitution. In-space placebo permutations, platform-composition controls, and a documented SSRN misclassification fix (349,942 papers) are handled in the paper.

What this means for the agentic era

Agents restructure labor — now.
Authorship, the credential currency of science, moved within months of each capability release.

Disclosure needs teeth.

A solo paper in 2025 is not the 2015 credential; AI-contribution statements should be as standard as data-availability statements.

The apprenticeship pipeline is exposed.

AI absorbs the tasks PhD students and RAs learn on — writing-intensive departments should monitor demand now.

The coding channel will deepen.

CS in 2025 looks like economics in late 2022 — its reversal has just begun, and coding agents are still maturing.