Technological Breakthroughs and the Progress of Science: Evidence from Early Computers*

Pedro Aldighieri† Franco Malpassi‡

Abstract

How do major technological shocks transform scientific research? We study the diffusion and impact of early digital computers at U.S. universities through 1970, using a novel database of 2,200 installations across 184 universities linked to publication records. Computer-referencing papers appear in the record immediately after campus installations, concentrated in subjects that had leaned on manual calculation before 1945. Conditional on observables, these papers are 20% more cited and nearly 50% more likely to reach the top 1%. The premium comes in part from computer papers being cited disproportionately as canonical sources of methods and concepts rather than as empirical findings; their reference lists pair recent work with an older, dormant tail. To identify causal effects, we exploit staggered university adoption and pre-computer differences in subjects' computational intensity in a triple-difference design. Within a decade of installation, numerically intensive areas gain in both output and citations per paper, and research shifts toward simulation, methods, and computational theory, away from empirics. The evidence is consistent with computers unlocking research paths bottlenecked by computation rather than by ideas. Unlike the slow, diffuse gains typical of general-purpose technologies, the returns arrive within years and concentrate precisely where that constraint had been binding – the pattern of a new tool meeting latent demand from problems scientifically mature but previously computationally intractable.

Keywords: digital computers, general-purpose technology, science of science, technology diffusion, research productivity, economic history.
JEL Classification: O31, O33, N72, I23.

*We thank Joseph Ferrie, Carola Frydman, Walker Hanlon, Ryan Hill, Benjamin Jones, and Joel Mokyr for their continued guidance and support. We also thank Netanel Ben-Porath, Wallace Inocêncio, Myera Rashid, Andrew Wooders, and participants at Northwestern's 501 Seminar, Economic History Lunch, and Applied Microeconomics Lunch for useful feedback. Thiago Catarino, Aster Li, Paulo Maya, and Jacob Siegel provided excellent research assistance. We acknowledge generous financial support from the Economic History Association's Early Stage Dissertation Grant, the Northwestern Center for Economic History, and computing resources from SambaNova and Groq via NSF's NAIRR Pilot.

†Department of Economics, Northwestern University. pedro.aldighieri@u.northwestern.edu
‡Department of Economics, Northwestern University. francomalpassi@u.northwestern.edu

1 Introduction

Scientific research and knowledge are widely regarded as key inputs to technological progress, which in turn is a main driver of long-run economic development (Jaffe, 1989; Mansfield, 1991). A large literature links basic research and new ideas to patented inventions (Ahmadpoor and Jones, 2017), industrial innovation (Mansfield, 1991), and productivity (Kantor and Whalley, 2019). A related line of work studies factors that shape the path of scientific discovery itself (Bryan and Williams, 2021).

The relationship between science and technology, however, is bi-directional. Technologies used as research tools feed back into science by changing what can be measured, computed, or inferred, altering the relative costs of tasks (Rosenberg, 1992; Mokyr, 2002; Krauss, 2026). This can affect the productivity of scientists, increasing the pace of scientific discovery and shifting its direction as fields benefit unevenly from technological advances. Despite the importance of technology for scientific research, empirical work studying how broad technological changes shape knowledge production remains scarce.

This paper studies the diffusion and effects of early digital computers on scientific research, from their first installations in universities in 1950 to 1970.[1] The adoption of high-speed digital computers represented a large discrete jump in computational power, which was a major constraint for research that relied on numerically intensive analysis, such as simulations. During this period, computers became available to researchers for the first time, although access depended on institutional affiliations and adoption was staggered (National Research Council, 1966). The early years of mainframe computing in US universities thus offer us a unique context to study the effects of an unevenly distributed technological shock on science.

To measure digital computer adoption and access, we construct the first comprehensive database of computer installations across US universities. We digitize and compile multiple historical surveys and archival sources, manually verifying 2,200 computer installations across all research-intensive US universities from 1950 up to 1971. We then combine

[1]For descriptive analysis, we sometimes extend our period to 1980, but there is no remaining variation in computer adoptions after 1969 in our sample.

university-level information on computer adoption with large-scale academic publication metadata from OpenAlex and SciSciNet to measure scientific outcomes. Finally, we use full text of papers to identify and understand computer usage.

We first document some facts on the adoption and diffusion of digital computers. Computers appear in the scientific record immediately after university installations, reaching about 10% of publication output at the university-year level within a decade. Early use is concentrated in the Physical Sciences and, to a lesser extent, the Social Sciences, with Life and Health Sciences lagging behind. Penetration across subjects is strongly predicted by pre-digital computational intensity, measured as the share of 1940–1944 papers featuring manual or mechanical calculation.

At the paper level, computer-related work carries sizable impact premia. Even after controlling for observables like year, subject, and author fixed effects, these papers earn roughly 20% more citations and are nearly 50% more likely to be in the top 1%; they are also about 21% more novel in their reference combinations by the Uzzi et al. (2013) metric. The premia are concentrated in papers that use computers *as tools* (e.g., for simulation or numerical analysis) rather than in papers that study or merely mention them. These papers are cited more as methods and canonical sources than as empirical findings, and their references pair younger work at the center-left of the age distribution with older, more dormant work in the tail. Together, these patterns suggest that computers often unlocked valuable research paths already understood in principle but that scientists lacked the tools to tackle.

At the researcher level, computer adopters are also positively selected: they are more experienced, publish more, receive 4.3× more citations, and use 30% more topics than non-adopters, even after controlling for observables.

To study the causal impacts of computer adoption, we exploit both variation in adoption timing across universities and across scientific subjects in their pre-computer era numerical intensity.[2] Between 1951 and 1969, 184 US research universities built university-wide computing centers, dramatically increasing the computational resources available to their

[2]We measure a field or subfield's numerical intensity by 1940-1944 reliance on manual or mechanical calculation as proxied by the share of papers with specific keywords.

researchers. Due to the high costs of purchasing and maintaining early mainframes, similar universities installed computers years apart, allowing us to compare researchers from early treated universities with those from not-yet-treated ones. At the same time, within treated universities, different subjects have different levels of exposure to this technological shock, explained by varying "computational demand" of research tasks across fields. These two sources of variation motivate a triple-difference design where we compare more vs. less exposed areas inside treated universities, relative to the same comparison inside not-yet-treated universities. We also exploit the variation in installation dates to implement a more traditional difference-in-differences event study design at the university level.

We find that, post installation, computers lead to a clear shift of scientific output towards compute-amenable areas like Numerical Analysis or Statistics. In the Physical Sciences, scientific output in more exposed subfields increases by about 15% after 10 years relative to unexposed ones; in the Social Sciences, the effect is 10% after 10 years. We observe even stronger effects on publication quality. Within a decade of adoption, papers in exposed Physical Sciences areas garner approximately 22% more citations than their unexposed counterparts; in the Social Sciences, this premium reaches 32%. These gaps represent a stark departure from the pre-computer era, where citation rates were indistinguishable across exposed and unexposed subjects.

The triple-differences design identifies relative effects within university-years. To recover level effects, we complement it with university-level difference-in-differences regressions, estimated subfield by subfield. The level effects are sizable. Computer adoption raised output, quality, and research breadth across subjects on average.

But the gains are higher in numerically intensive subfields. Subfield-specific estimates show that, within the Physical and Social Sciences, areas near the 90th percentile of pre-computer numerical intensity experienced publication gains of about 31%, compared with about 17% at the 10th percentile. The spread in other outcomes is similar: citations per paper rise by about 51% versus 24%, and topics per paper by 0.37 versus 0.17, for high- versus low-intensity subfields, respectively. That the magnitude of the effect scales smoothly with an exposure measure built from 1940–1944 publications – before digital computers existed

– supports the interpretation that computers mattered precisely where they relaxed a previously binding constraint.

Finally, research methodologies shift markedly following computer installations. The empirical share falls while methods and simulation rise, with imprecise gains in theory. The content underneath is consistent with computers relaxing binding computational constraints. Methods gains concentrate in algorithmic and numerical work; simulation in direct numerical experimentation; and the theory gains are associated with terms like "plasma" in physics, where governing equations had long been written down but required computers to actually solve. Even empirical work, though losing relative share, sees movements in a computer-intensive direction: toward measurement, monitoring, and numerical processing of observed systems rather than the regression-heavy style of later decades. These causal shifts in research composition line up with the paper-level descriptives: computer papers are more likely to function as methods and conceptual sources, and to draw on older but newly tractable lines of work.

This paper presents novel evidence on how major, systemic technological shocks are diffused across scientific research, and how this can affect researchers' choices, skewing their output towards specific areas and methods. Previous work on factors affecting the quality and direction of scientific research have focused on the effect of enhanced data access (Nagaraj and Tranchero, 2023), perceived societal needs (Hill et al., 2021), the provision of targeted funding (Myers, 2020), environment diversity (Truffa and Wong, 2024), or competition across fields (Borjas and Doran, 2012). By leveraging novel data in a unique historical context, we are able to focus on the understudied question of how does technology itself affect scientific progress and the direction of science. Boudou and Mckeon (2024) have studied the effects of supercomputers on research, capturing the effects of increases in compute within an already-digital context. Our setting instead captures a discontinuous change: the arrival of programmable computation itself.

More broadly, we contribute to the literature exploring the links between technology and science. As mentioned earlier, much work has been focused on the technological innovations and productivity gains stemming from scientific research (Jaffe, 1989; Gaetani and Bergolis, 2015; Kantor and Whalley, 2019; Liu, 2015; Mansfield, 1991). The literature on the reverse direction, looking into how technological innovations can influence basic research, is sparser, and focuses on either narrower technological shocks, such as internet-enabled communication technology on collaboration (Agrawal and Goldfarb, 2008), and patenting (Wernsdorf et al., 2022), or contexts less amenable to causal inference, such as AI adoption (Gao and Wang, 2023), or the relationship between patents and papers (Ahmadpoor and Jones, 2017). We provide causal evidence on a large-scale technological shock, as well as descriptive evidence of the adoption of a transformative technology by researchers. We leverage quasi-experimental variation to identify how computers shifted research direction, methods, and impact. Our analysis spans two decades, revealing both short and medium term productivity effects in scientific methodology.

Finally, we contribute to the literature on the diffusion of general-purpose technologies (David, 1990; Rosenberg and Trajtenberg, 2001; Comin and Hobijn, 2010, 2004) by examining computer technology, its diffusion and effects on knowledge production itself. Canonical work on GPTs emphasizes the slow unfolding and need for complementary investments, reorganization, and codified know-how (David, 1990; Bresnahan and Trajtenberg, 1995). Our setting provides a new lens: the diffusion of a GPT within science, where institutional access is discrete, timing is well documented, and the scientific record leaves a contemporaneous trace of use.[3] This lets us track precisely when, where, and how computers appeared in academic research, providing micro-level evidence on GPT adoption that broader economy-wide studies cannot capture.

Taken together, the evidence suggests that computers opened up new research paths, and that these paths were ones bottlenecked by tools rather than by ideas. Several features of our findings point in this direction. The gains concentrate in fields that had leaned hardest on manual calculation before 1945, precisely where the computational constraint had been binding. Within those fields, computer-intensive papers serve disproportionately as methods and conceptual sources rather than as empirical findings, the signature of work that encodes a

[3]The distinction between core-technology papers and tool-use papers established in Section 4.5 helps reconcile our descriptive findings with Moser and Nicholas (2004), who document that electricity patents are *less* cited than their counterparts. This pattern is consistent with a GPT's research value lying in its enabling effects across domains rather than in the core technology per se.

newly practical way of doing something rather than a new thing to know. Their reference lists pair recent work with an older, dormant tail, consistent with researchers picking up threads they understood in principle but had been unable to execute. Quantity and quality also rise together, with no trade-off of the kind we would expect if computers were merely lowering the cost of producing papers. Computers unlocked valuable research paths that had long been within intellectual reach but beyond computational grasp.

This pattern matters beyond the historical case. Standard GPT narratives emphasize slow, diffuse productivity gains that arrive only after decades of organizational adjustment. Our evidence shows that in scientific areas where the binding constraint has been tools rather than ideas, gains arrive quickly and concentrate precisely where latent demand had accumulated against that constraint. Problems scientifically mature but computationally blocked translated into published results within a decade of adoption. For transformative research tools like Artificial Intelligence, this implies scientific winners and losers sorted less by who adapts fastest than by whose binding constraints are the ones the new technology happens to relax.

The rest of the paper proceeds as follows. Section 2 reviews the historical context of early university computing. Section 3 describes our data and measurement strategies. Section 4 presents descriptive facts and analysis on diffusion, usage, and impact of digital computers. Section 5 lays out our identification strategy. Section 6 reports the main results on direction, quality, breadth, and methods. We conclude in Section 7.

2 Historical Background

> ”It would appear that we have reached the limits of what is possible to achieve with computer technology, although one should be careful with such statements, as they tend to sound pretty silly in five years.”
>
> – John von Neumann (1949)

Before the advent of digital computers, scientific research was severely constrained by computational power. Complex calculations required large teams of human computers and punched card tabulating machines working for months or years on single problems (Grier, 2013). These manual methods were not only slow but also error-prone – the 1836 Nautical Almanac had to include erratas of erratas of erratas (Swade, 2001). Even mechanical devices like differential analyzers offered limited relief: they required constant human intervention between operations and were special-purpose machines, hardwired for specific calculations.[4] Many calculations that are nowadays regarded as routine, such as running multivariate regressions, were simply impossible (Campbell-Kelly et al., 1980).

The ENIAC, completed in 1946, represented a discrete technological jump.[5] Unlike its mechanical predecessors, it operated 1,000 times faster (War Department, Bureau of Public Relations, 1946) and was general-purpose programmable. This breakthrough initiated an exponential increase in computing power, with Nordhaus (2007) documenting an approximately 100,000-fold increase in computations per second between 1945 and 1970.

In many cases, the bottleneck was not a lack of ideas but a lack of feasible computation. Researchers often knew which problems mattered and how to tackle them in principle, but could not execute the relevant calculations at useful scale. Numerical weather prediction is the clearest example: Richardson (1922) laid out the vision decades before Charney et al. (1950) made it operational on digital computers. The same pattern appears in early quantum chemistry, where digital computers made molecular-orbital calculations tractable for the first time (Pritchard and Sumner, 1954), and in X-ray crystallography, where the EDSAC computer at Cambridge enabled Fourier syntheses on a previously unattainable scale, underpinning Kendrew and Perutz's determination of the three-dimensional structure of proteins and their 1962 Nobel Prize in Chemistry (Kendrew, 1963).

[4]This meant, for example, that calculating ballistics tables and inverting matrices required different physical hardware, not just new instructions.

[5]The ENIAC's place as the very first digital computer is disputed, given other computation projects happening in parallel, but its completion date gives us an estimate of the very first computation technology (Picker, 2018).

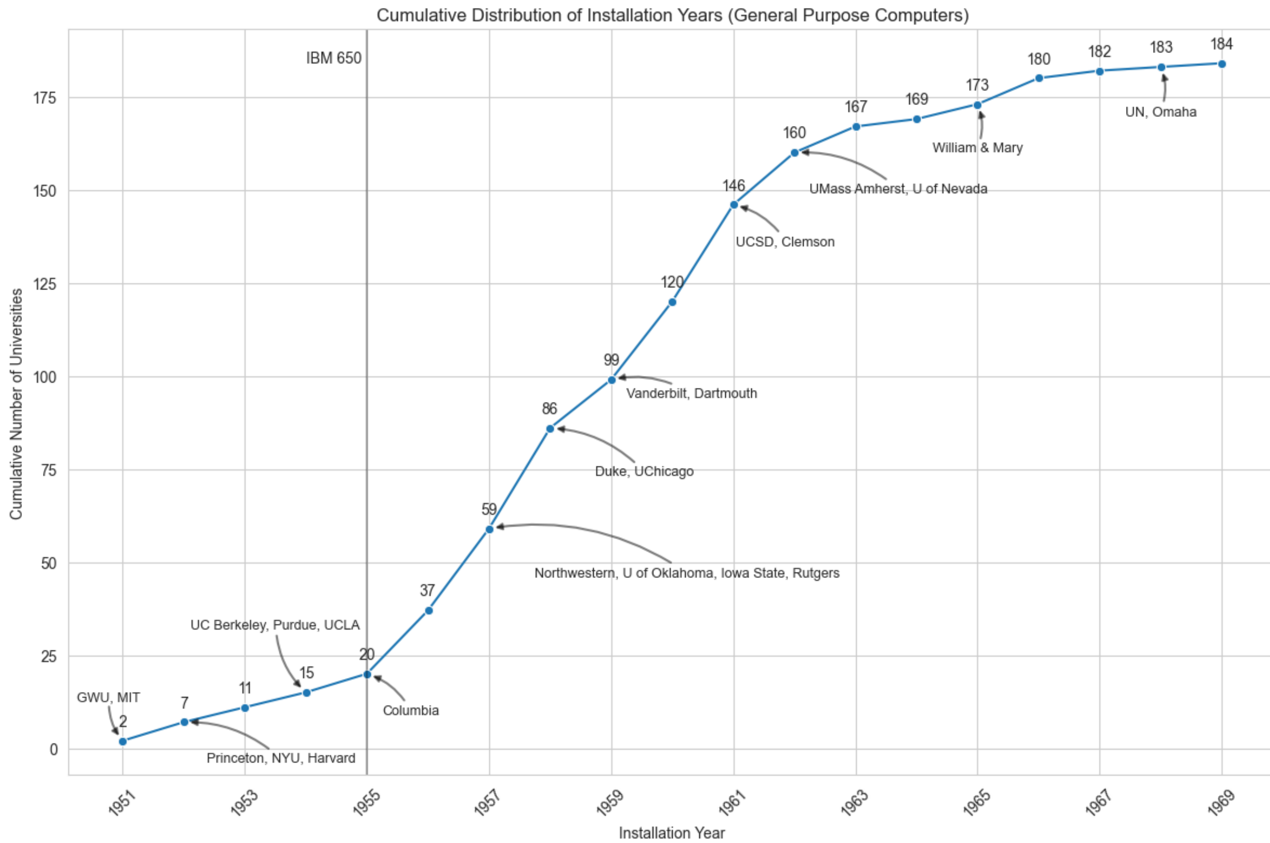


Figure 1: Digital Computer Adoption by US Universities, 1950-1970. $N = 184$. This figure excludes analog or electromechanical computers and non-general purpose digital computers, like the IBM 602. We also exclude computers that were built at universities but shipped elsewhere shortly thereafter like the UPenn's ENIAC.

2.1 Diffusion and Access of Digital Computers at Universities

Digital computers took over a decade to diffuse across US universities. Figure 1 displays the cumulative adoption in our sample; only by 1969 had all research-intensive institutions acquired at least one computer.[6] The main bottleneck was access and cost, not demand. Scientists recognized computers' transformative potential early on, but the UNIVAC I cost $1.5 million in 1951 (about $15 million in 2025 dollars) and required 382 square feet of space. Before the IBM 650's release in 1954, commercial availability was extremely limited and universities mostly built computers in-house.

Because computer machinery was so expensive, installations were often funded through equipment-specific channels – manufacturer donations,[7] NSF acquisition grants,[8] or dedi-

[6]For more details on the figure, refer to Section 4.1.

[7]Georgia Tech, UChicago, NYU, Notre Dame, Ohio University, the University of Southern California, the University of Maine, and the Case Institute of Technology all had computers donated by manufacturers.

[8]Some examples include: the University of Oklahoma, the University of Minnesota, Iowa State, Yale University, UNC Chapel Hill, and Cornell University (National Science Foundation, 1958, 1959).

cated fundraisers – rather than general research budgets. Due to costs, even similar installations often happened apart in time. For example, New York University's first computer was a UNIVAC I acquired through a joint venture with the Atomic Energy Commission in November 1952, whereas the University of Chicago's first computer was also a UNIVAC I, donated by Remington Rand and installed in May 1958. Once installed, universities housed computers in shared centers serving the entire campus (representing over 50% of installations in our sample, see Appendix E.1). The National Science Foundation reinforced this model by requiring university-wide access as a condition for funding support (Ceruzzi, 2003; National Research Council, 1966).

This institutional structure, combined with limited remote access until the late 1960s, meant researchers could effectively use computers only at their home institutions.[9] The 1950-1970 period thus represents a unique window where computer access was uneven across the US and sharply determined by institutional affiliation.

3 Data

Our analysis draws on two primary data sources: (1) a novel database of computer installations in US universities, and (2) scientific publication metadata from OpenAlex and SciSciNet.

3.1 Computer Installations Database

We constructed the first comprehensive database of computer installations in US higher education through 1971.

In total, we have collected, digitized, and processed 82 surveys from the period, yielding 18,282 computer snapshots.[10] The raw database covers over 1,180 US universities out of approximately 3,000 institutions in the period, spanning all 50 states, the District of Columbia,

[9]Several universities accessed UCLA's Western Data Processing Center via telephone or mail, but Oregon State records show outside institutions had less than one hour daily for data transmission and no direct computer access. Faculty had to travel to UCLA to debug programs, severely limiting usefulness.

[10]There are in total 24 survey sources, many of which were repeated over time, totaling 82 survey-year pairs.

and Puerto Rico, including all US doctoral-granting institutions of the period.

For the analysis in this paper, we have consolidated and verified installation data for 184 universities, comprising exactly 2,200 computer installations, from 1951 to 1971. These universities include all of the most research intensive universities – R1 and R2 – as measured by the Carnegie Commission on Higher Education (1973).[11]

To fill in gaps and adjudicate a few cases of conflicting information, we supplement the surveys with university-sourced data, which includes online information, digital archives, and archival materials digitized upon request. For 74% of these installations, we have direct documentation of installation dates. Where precise installation dates are unavailable, we used the date of first survey appearance as a conservative estimate.

To validate our installation dates, we cross-referenced our database against mentions of specific computer models in the scientific literature. For example, mentions of Northwestern's IBM 650, or the University of Illinois' ILLIAC I in published papers closely track their university installation dates, as it was common for authors to mention using computers in the early days of computing, sometimes by name (see Appendix A.4). This validation confirms the accuracy of our installation database and shows that researchers began using and citing these machines immediately upon installation.

Each observation in the final computer installations dataset is a computer linked to a university. In Appendix A, we provide more details on available variables (A.1), sources (A.2), and universities in sample (A.3).

3.2 Publication Data

We retrieve publication and citation metadata from OpenAlex (Priem et al., 2022), an open source database that succeeded Microsoft Academic Graph and provides the most comprehensive coverage among scientific publication metadata repositories. The data source has

[11] R1 universities are the 50 leading universities in terms of federal financial support for academic science, measured over at least two of the three academic years: 1968-69, 1969-70, and 1970-71. They must have awarded at least 50 Ph.Ds in the academic year 1969-70. R2 universities are among the top 100 institutions in terms of federal financial support for academic science in the same period and likewise must have awarded at least 50 PhDs.

been used extensively in the economics of science literature, e.g. Azoulay and Greenblatt (2025); Schmallenbach et al. (2024); Shvadron et al. (2025).

OpenAlex uses a 4-tier subject classification hierarchy (domains, fields, subfields, and topics) that we leverage in our analysis. There are four domains (Physical Sciences, Life Sciences, Health Sciences, and Social Sciences), divided into 26 fields, 252 subfields, and 4,516 topics.[12]

We apply a series of consistent filters and aggregation rules in our analysis. First, we restrict all samples to papers published in English and after 1940. The language restriction is necessary because our keyword-matching strategy relies on English terms, while the 1940 cutoff provides a decade-long pre-computer baseline period for comparison.

For paper-level descriptives, we focus on works for which the full text is available for keyword search via n-grams. This allows us to capture trends up to 1970 or 1980, depending on the horizon of interest. Within these bounds, our dataset includes 8.9 million papers before 1970 and 16.7 million before 1980, with searchable full text available for roughly 40% of works in both slices.[13]

For author- and university-level analysis, we further restrict to papers with an affiliation to universities in our installations database between 1940 and 1970. This yields about 650,000 papers, of which 73% have searchable full text.[14]

We merge OpenAlex metadata with SciSciNet to expand the set of outcomes available (Lin et al., 2023).[15] We include citations within 5 and 10 years of publication, the disruption measure by Park et al. (2023), the atypicality measure by Uzzi et al. (2013), patent citations to papers, and NSF grants awarded to papers.

[12]For more details, see `https://help.openalex.org/hc/en-us/articles/24736129405719-Topics`.

[13]For more details on the full text sample, see Section E.2.

[14]Out of OpenAlex papers between 1940-1970, only 10% have affiliation information. This is a result of both improper metadata cataloging and old papers not reporting affiliations of researchers. Therefore, the universities in our sample cover more than half of all published research for which an affiliation can be found.

[15]SciSciNet is a data lake seeded with OpenAlex data, but enriched for the purposes of science of science research.

3.2.1 Identifying computer use in articles

We use two complementary approaches to identify whether research papers made use of computers.

N-gram search in OpenAlex corpus. As previously mentioned, OpenAlex enables n-gram search from paper full text across roughly 40% of works in our sample period. We search for computer-related keywords such as "digital computer," "electronic computer," and "high-speed computing device," (see Appendix A.5 for keywords). It was common for early computer users to state explicitly that results were obtained using a digital computer, making keyword detection reliable. This approach flags 2.27% of searchable papers as using or mentioning computers – or 4.86% when including "computer" as a standalone word as well. As shown in Sections 4 and A.4, the timing of these mentions closely tracks university computer installations.

Full text classification with Large Language Models (LLMs). We complement the n-gram approach by analyzing 1.3 million full text papers obtained directly from publishers.[16] We first screen papers using an extended keyword list for high recall (see Appendix A.5), then use LLMs to classify whether papers use or mention computers in their research. Within the sample of papers with full text, LLMs identify 3.45% of papers as using or mentioning computers.

Both methods perform very well on a hand-coded validation sample of 200 papers ($F_1 = 0.928$ for keywords, 0.967 for LLMs),[17] with similar detection rates. Check Appendix A.5 for details.

[16] We obtained authorization via Northwestern library from several publishers to download paper full text for research purposes.

[17] The F_1 is a standard measure to assess quality of labeling. The score is the harmonic mean of precision (correctly labeled positives divided by all labeled positives) and recall (correctly labeled positives divided by all ground truth positives), ranging from 0 to 1. Scores above 0.9 are considered excellent, indicating that the method correctly identifies most relevant papers while producing few false positives and negatives.

4 Descriptive Analysis

4.1 Digital Computers at Universities

Figure 1 shows the cumulative distribution of digital computer adoptions by US universities from 1950 to 1970. The pattern reveals an S-shaped diffusion curve typical of technology adoption, with a slow start in the early 1950s, accelerating adoption in the late 1950s and early 1960s, and saturation by 1969.

The IBM 650 was a landmark, serving as the first computer for 49 universities (27% of our sample). The intensity of computer adoption varied substantially across institutions, with universities having a median of 8 computer installations through 1971, ranging from 1 to 83 installations. Importantly for our identification strategy, computer centers housed most computers (50.5% of all installations), providing university-wide access rather than department-specific exposure. External computers comprised only 1.3% of devices universities had access to, reinforcing that computer access was determined at the institutional level. See Appendix E.1 for additional details on manufacturers, models, and departmental distributions.

4.2 Field Distribution of Computer-Intensive Papers

Computer adoption varied markedly across academic disciplines. Figure 2 shows that Physical Sciences led computer intensive research, accounting for over 70% of computer-related papers compared to 40% of all searchable papers. Within Physical Sciences, engineering and computer science dominated adoption, while in mathematics, numerical analysis and probability and statistics were the primary users (see Appendix E.2).

At the field level, Mathematics, Decision Sciences, and Engineering were early adopters, though Mathematics' computer usage was relatively low by 1980 despite its early start.

In Appendix E.2, we display more results at field and subfield level, and also show results are robust to using LLM classifications instead.

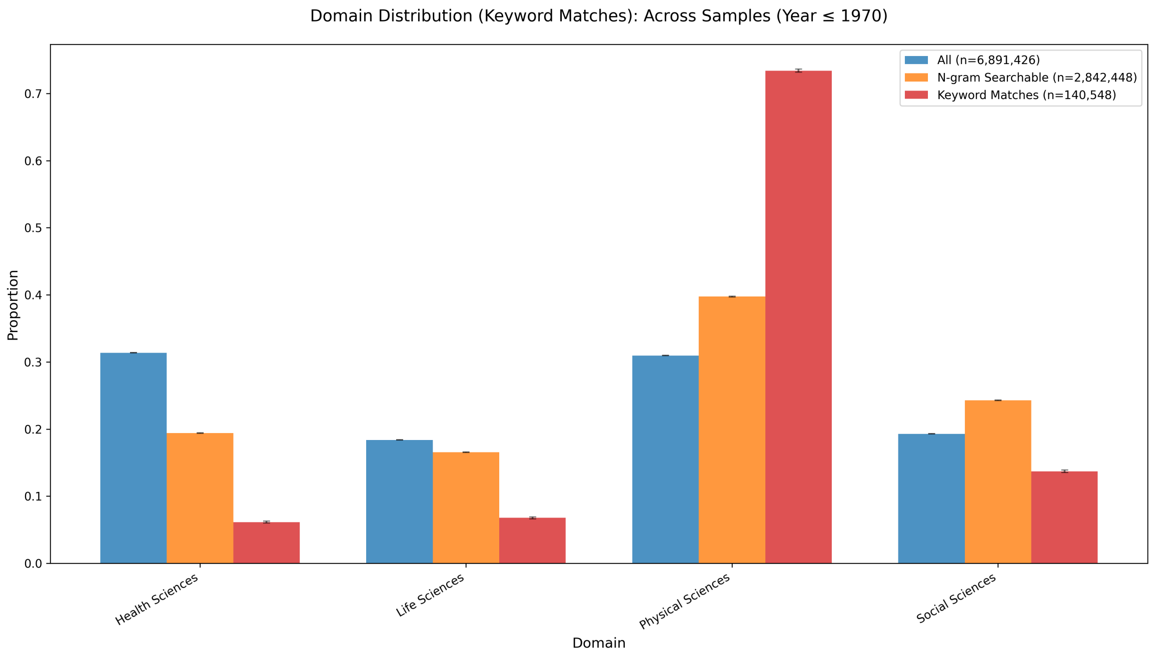


Figure 2: Distribution of Computer Papers Across Domains. The plot compares the distribution of papers by domain in three slices: (a) all papers; (b) the papers we can search full text n-grams for in OpenAlex; and (c) the papers that were flagged as using or mentioning computers from keywords. Bars of the same color correspond to each slice and add to 1. Plot includes "computer" as a keyword.

4.3 Taxonomies and Trajectories of Computer Usage in Research

To characterize the ways in which computers appeared in the scientific literature, we employ large language models (LLMs) to classify computer usage and mentions from the full text of papers into four broad categories as a starting point. Among papers classified as computer use or mention, the categories are: (i) *Computer as a Tool* (62.7%), where machines are used instrumentally e.g. for analysis or simulation; (ii) *Computer as an Object of Study* (2.6%), investigating the computer itself; (iii) *Hardware or Software Papers* (5.5%); and (iv) *Other Mentions* (29.2%) for references to computers do not fit the previous categories, usually mentions in passing to computers.

We then complement this initial approach with a data-driven procedure inspired by Tamkin et al. (2024). We extract computer usage descriptions from papers using an LLM, then cluster their embeddings to generate usage categories.

In Figure 3, we show the evolution of tasks within computer papers from 1950 to 1970. Computers were primarily used for numerical computation (22.2% of papers), dynamic mod-

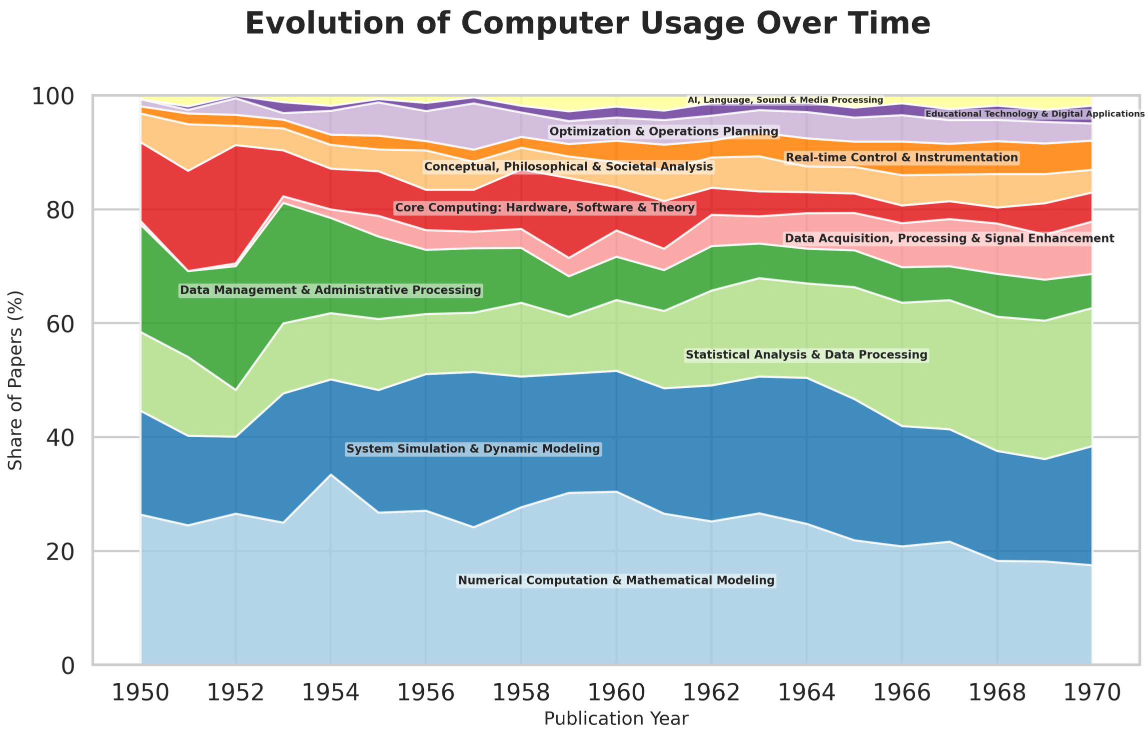


Figure 3: Evolution of computer-usage categories among papers flagged as mentioning or using computers, 1950–1970. Categories come from the LLM-plus-clustering taxonomy described in Section 4.3 and Appendix C. Shares are computed each year over the subset of computer-flagged papers and normalized to 100%. $N = 30{,}959$ clustered research-article records. Series refers to the pooled full-text sample.

elling and simulation (21.3%), and statistical analysis (19.3%). Statistical Analysis & Data Processing grows in importance over time. The relative stability of usage categories across this period is noteworthy, given the nascent state of digital computing technology. Also of note is the immediate and sustained presence of simulation and dynamic modeling applications. These are computational approaches that, while theoretically well-established, had been practically infeasible with manual or mechanical calculation methods prior to the advent of digital computers.

In Appendix C, we provide further details on the classification procedure and breakdown of usage by domain.

4.4 Correlates of Computer Diffusion

To better understand the drivers of computer diffusion across fields, we examine whether early reliance on intensive calculation practices predicts subsequent adoption. We construct field- and subfield-level measures of computational intensity by flagging papers between 1940-1944 that mention terms related to calculation by hand, proto-calculators, classical numerical methods, and instruments (see Appendix A.5 for the complete keyword list). Fields where researchers were already engaged in laborious numerical work, such as solving systems of equations, performing matrix operations, or using mechanical calculators, should be more likely to adopt electronic computers once available, as computers directly substitute for these manual methods.

Figure 4 plots the relationship between the share of 1940-1944 papers in a field (left plot) or subfield (right plot) flagged with numerically intensive keywords and the share of papers mentioning or using computers in 1960-1964. The regression fit is strong ($R^2 = 0.720$ and $R^2 = 0.359$ at the field and subfield levels, respectively) with a significant positive slope. At the subfield level, the relationship is still strong, though more noisy because subfields in the Life and Health Sciences display little variation in levels of both pre-computer calculation intensity and computer penetration relative to the Physical and Social Sciences.

To better understand the subfield-level patterns, we examine cross-sectional variation in adoption within each domain. Figure 5 relates the share of papers using computer keywords in 1960-1964 to the share of calculation-intensive keywords in 1940-1944 at the subfield level. The association is stronger in the Physical and Social Sciences, but much weaker in the Life and Health Sciences, which were also slower to incorporate computers.

These results are robust to varying the base and target periods, including "computer" as a keyword, and using LLM classification instead of n-gram matching.

4.5 Citations and Research Impact

In this section, we show computer-intensive papers had greater impact, breadth, and novelty along several measures.

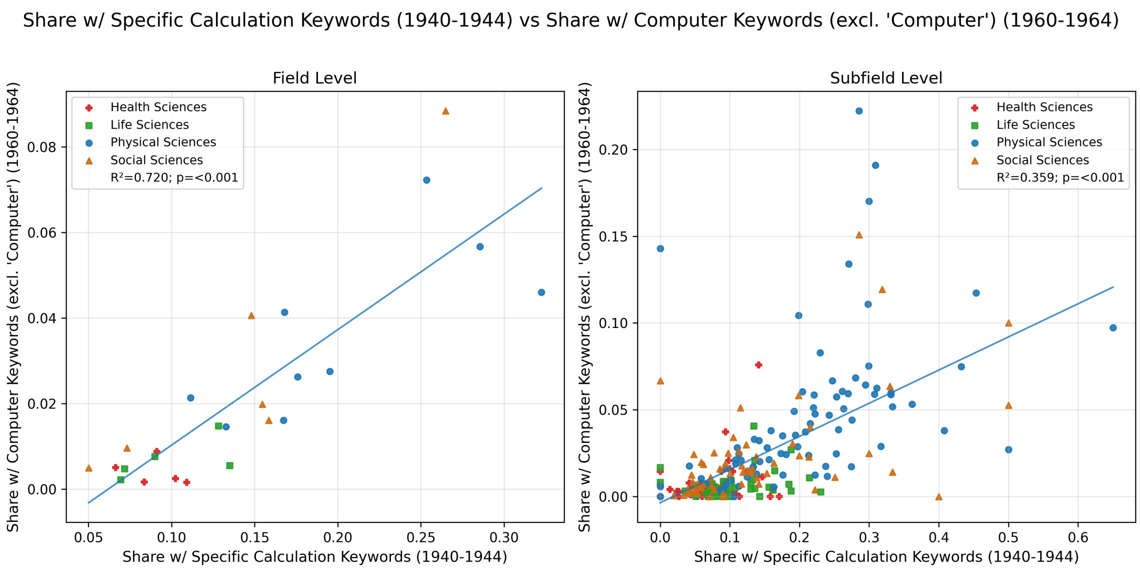


Figure 4: Computer diffusion across OpenAlex fields (left panel, $N = 26$) and subfields (right panel, $N = 252$). Each dot is a subject; y-axis is the 1960–1964 share of papers mentioning computer-related keywords, x-axis is the 1940–1944 share of papers containing calculation-specific keywords (see Section 4.4 and Appendix A.5). Red line shows the fitted OLS relationship; shaded band is the 95% CI. Subjects with high pre-digital numerical intensity show higher subsequent computer uptake, particularly among Physical and Social Sciences subfields.

In Table 1, we regress several citation outcomes on a computer-keyword indicator from OpenAlex full text n-grams (see Section 3.2.1). Regressions control for author, year, university and subject fixed-effects, in addition to the number of authors and the number of NSF grants awarded to the paper.[18]

Our results show computer-intensive papers display roughly a 20% citation premium, and a nearly 50% higher chance of being a top 1% paper by citations.[19] Papers are already recognized by citations within five-years of publication, though magnitudes are somewhat smaller. Appendix E.3 shows that these citation premia remain when we additionally partial out semantic embeddings from titles and abstracts together with SPECTER embeddings, which proxy for a paper's position in citation-neighborhood space.[20]

[18]To apply author fixed effects, we explode the dataset to author-paper pairs and then weight OLS by the inverse number of authors per paper.

[19]Top percentile by citations refer to percentiles within subfield and year.

[20]We use dense title and abstract embeddings as flexible controls for paper semantics. SPECTER is a paper-level embedding trained so documents with similar citation neighborhoods lie nearby in vector space; controlling for it absorbs variation associated with citation-network position and nearby intellectual neighborhoods, not just textual similarity.

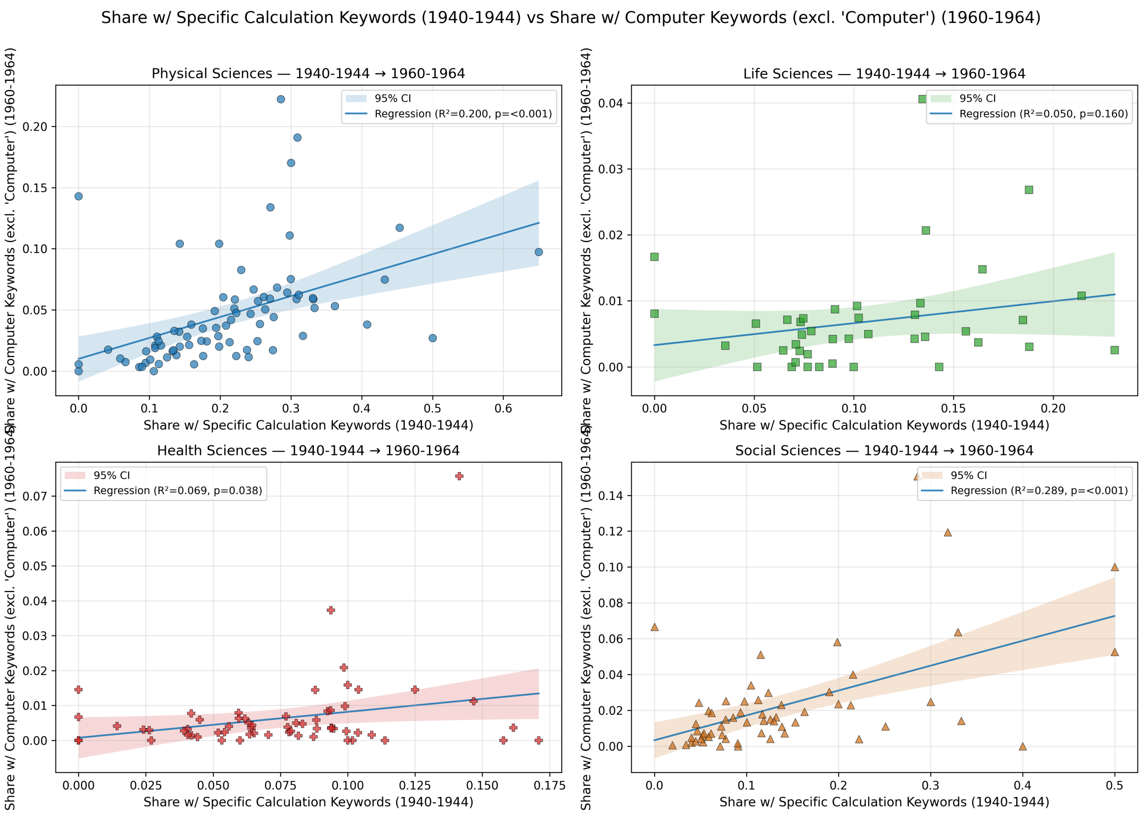


Figure 5: Subfield-level OLS of the 1960–1964 share of papers mentioning computer-related keywords on the 1940–1944 share of calculation-specific keywords, split by domain ($N = 252$ subfields across Physical, Life, Health, and Social Sciences). Each dot is a subfield; red line is the within-domain fitted relationship with 95% CI. The pre-digital numerical-intensity gradient is steepest in the Physical and Social Sciences and weakest in the Life and Health Sciences.

Table 1: Citation Outcomes

	Log cites	Top 10%	Top 1%	FWCI	C5	Cite pct.
Computer-Keyword Flag	0.191***	0.046***	0.009***	0.409***	1.122***	0.035***
	(0.004)	(0.001)	(0.001)	(0.018)	(0.042)	(0.001)
Number of Authors	0.114***	0.021***	0.002***	0.343***	0.834***	0.021***
	(0.002)	(0.000)	(0.000)	(0.007)	(0.020)	(0.000)
NSF Grants (paper)	0.294***	0.091***	0.007	0.583***	1.001***	0.049***
	(0.026)	(0.012)	(0.005)	(0.100)	(0.213)	(0.004)
Observations	3,903,691	3,903,526	3,903,526	3,902,497	3,903,691	3,903,526
R^2	0.610	0.428	0.291	0.385	0.435	0.612
Mean of Dep Var	1.827	0.191	0.019	1.950	3.880	0.568
Author/Year/Univ FE	Yes	Yes	Yes	Yes	Yes	Yes
Subject FE	Topic	Topic	Topic	Topic	Topic	Topic

Notes: Paper-author observations, weighted by the inverse number of authors. Treatment is an indicator for computer-related keywords in the paper's full text. Controls are the number of authors and the number of NSF grants awarded to the paper. Fixed effects are author, publication year, university, and primary topic. Standard errors are clustered at the paper level (OpenAlex Work ID). Sample restricted to papers with at least one in-sample university affiliation and searchable full text in OpenAlex. Field-Weighted Citation Impact (FWCI) normalizes citation counts by field, year, and type where 1.0 indicates average performance. C5 refers to citations accrued within 5 years of publication. Citation percentile is also calculated within year and field. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Computer papers are also about 21% more novel or atypical in their journal combinations within the references by the Uzzi et al. (2013) metric.[21] They also use more concepts and topics, and have a more spread out conceptual and topical focus as measured by HHI of concepts and topic relevance scores. These breadth effects are modest, generally ranging from about 1% to 4% of the mean. Appendix E.3 provides details and additional results on these novelty and breadth measures. Further results in Appendix E.3 show that the citation premia and concept-based novelty advantages are concentrated among papers using computers as tools, following the categorization we introduced in Section 4.3.

4.6 Why Are Computer Papers Cited More?

To understand why computer papers receive more citations, we analyze the function their incoming citations serve. Using open-source citing papers, we recover short LLM descriptions

[21]The measure assigns each cited-journal pair an atypicality score relative to a baseline network of random citations. We use the 10th percentile of the within-paper atypicality z-score distribution, so the measure captures the paper's 10% most atypical reference combinations.

of why each cited paper is used.[22] We then cluster them into five buckets: technical evidence and protocols, established findings as evidence, representative literature precedent, formal method or theory foundation, and canonical sources for concepts and systems.[23] Appendix D details the pipeline, sample construction, and provides illustrative examples.

Figure 6 shows clear compositional differences. Relative to matched controls, computer papers are cited less often as established findings and more often as representative precedent, formal methods, and especially canonical sources. The distinction matters substantively: an established finding is invoked as evidence for a factual claim, while a canonical source marks the recognized origin of a concept or system that later work builds on, regardless of whether the original version remains in use.

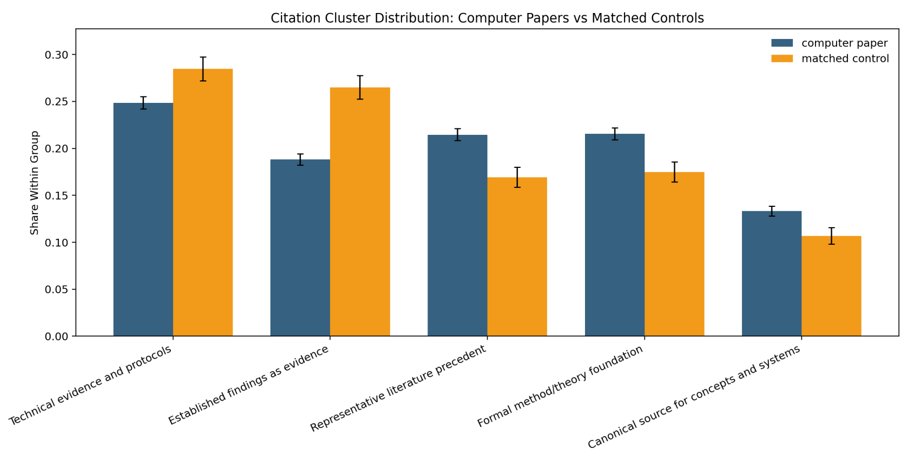


Figure 6: Citation functions of incoming citations to computer papers and matched controls. Bars report the share of usable classified citation links in each citation-function bucket. Buckets are constructed from embeddings of short LLM descriptions of why the cited paper is used. The sample contains 20,910 usable citation links, of which 16,114 point to computer papers and 4,796 to matched controls.

This reallocation toward high-premium citation roles accounts for part of the citation premium. A descriptive accounting exercise in Appendix D.3 shows that adding citation-

[22]Incoming citations are drawn from the full available record through 2025, not restricted to our analysis window. This lets us observe the long-run function a paper serves in the literature, including uses that emerge decades after publication.

[23]Formal methods and canonical sources are related but distinct. A *formal method* citation invokes a named estimator, framework, or derivation that the citing paper's analysis directly uses (e.g., Wilkinson's numerical methods when solving eigenvalue problems). A *canonical source* citation invokes the recognized origin of a concept, system, or language for provenance (e.g., Hoare on monitors, or the TAXIS data model), often without using the original version directly. Appendix D.4 provides illustrative examples.

function composition and an excess-breadth measure to the baseline regression attenuates the FWCI premium by about 10%, with a Shapley decomposition assigning roughly 90% of that explained portion to composition rather than breadth. Computer papers are not cited more because they play many roles at once; they are cited more because they disproportionately play especially high-premium roles, above all as canonical sources and formal methods. This dovetails with the broader methodology results in Section 6.4, where empirical work loses share and methods, simulation, and theory gain ground.

4.7 Reference Age and Sleeping Beauties

A recurring theme in early computer use is that researchers could act on ideas that had existed in principle for years but had been bottlenecked by access to computation, like weather prediction (see Section 2 for historical examples). We quantify this unlocking margin by studying the age profile and dormancy of the references cited by computer papers.

Panel A of Table 2 shows a distinctive pattern. Computer papers cite younger work at the left and center of the age distribution: both the minimum and median cited reference are younger. At the same time, they reach further back in the right tail: the oldest cited reference is significantly older. The omitted 90th-percentile cited-reference age measure is also positive and significant, so this right-tail pattern is not driven solely by the single oldest citation. This is close in spirit to the "hotspot" pattern documented by Mukherjee et al. (2017), where work is more likely to become highly cited when it combines recent references with a long tail of older knowledge. In our setting, that long right tail is especially suggestive of older ideas whose implementation had been constrained by computational bottlenecks.

Panel B of Table 2 reinforces that interpretation. Relative to other papers, computer papers cite references with substantially higher Sleeping Beauty scores: about 20% higher on average, 11% higher for the median-age cited reference, 32% higher for the oldest cited reference, and 30% higher among references in the oldest-age decile.[24] In other words, computer pa-

[24]The Sleeping Beauty score B of Ke et al. (2015) is constructed by drawing a linear reference line from a paper's publication year to its citation peak and integrating, year by year, the gap between that line and the paper's actual annual citations (normalized by each year's count). The resulting area captures under-performance relative to a smooth growth path: higher B means the paper's citations sat well below the line for longer before

pers disproportionately reach back to ideas that had lain relatively dormant before drawing renewed attention. This does not imply that computer papers themselves waited unusually long to be recognized. Rather, it suggests that computers helped researchers revive older lines of work that had remained underused because the relevant calculations, simulations, or data processing tasks had previously been too costly.

Appendix E.5 shows that these patterns likewise remain similar under the same semantic and citation-neighborhood embedding controls.

4.8 Author-Level Patterns

Computer adopters were positively selected researchers.[25] Table 3 shows that computer adopters were positively selected along several dimensions: they received 4.3 times more citations and used 30% more topics,[26] even after controlling for publication counts, affiliation, cohort, and field. While adopters published 4 times more papers, this partly reflects mechanical association.

The results also hold at the intensive margin – more computer-related papers lead to better outcomes – and even before exposure to computers via their affiliations, though magnitudes are much smaller than when looking over the whole period. See Appendix E.4 for more details.

5 Empirical Strategy

While the previous sections documented patterns in computer adoption and its impacts on research, these relationships are possibly endogenous even after controlling for observables. We therefore employ two complementary identification strategies to estimate causal effects.

First, we implement a difference-in-differences (DiD) design exploiting the staggered tim-

rising sharply, consistent with an initial period of dormancy followed by belated recognition.

[25]We proxy computer adoption by keyword mentions in papers. Since we cannot search keywords in all papers, some authors classified as non-adopters may have used computers, biasing results downward.

[26]Topics are the lowest level subject classification OpenAlex assigns to papers. See Section 3.2.

Table 2: Reference Age and Dormancy Outcomes

Panel A: Reference age				
	Mean	Min	Median	Max
Computer-Keyword Flag	0.025	-0.217***	-0.099***	1.300***
	(0.022)	(0.014)	(0.021)	(0.066)
Number of Authors	-0.188***	-0.100***	-0.186***	-0.314***
	(0.005)	(0.003)	(0.005)	(0.015)
NSF Grants (paper)	0.450**	-0.261**	0.190	2.944***
	(0.202)	(0.114)	(0.206)	(0.703)
Observations	2,912,619	2,912,619	2,912,619	2,912,619
R^2	0.564	0.545	0.551	0.490
Mean of Dep Var	8.554	2.705	7.134	20.159
Author/Year/Univ FE	Yes	Yes	Yes	Yes
Subject FE	Topic	Topic	Topic	Topic
Panel B: Reference dormancy (Sleeping Beauty scores)				
	Mean	Median Ref.	Oldest	Age90 Mean
Computer-Keyword Flag	14748.189***	5949.303*	47497.033***	42901.972***
	(2791.209)	(3497.185)	(9005.499)	(7858.461)
Number of Authors	-91.527	-308.707	1448.036	2225.190*
	(425.293)	(548.370)	(1454.504)	(1194.364)
NSF Grants (paper)	-10498.526	-45371.170**	-59951.894	-38000.606
	(15767.920)	(22087.269)	(42935.123)	(36637.594)
Observations	2,911,144	2,900,333	2,905,141	2,906,534
R^2	0.519	0.468	0.423	0.440
Mean of Dep Var	72464.983	52424.594	148906.728	141450.568
Author/Year/Univ FE	Yes	Yes	Yes	Yes
Subject FE	Topic	Topic	Topic	Topic

Notes: Paper-author observations, weighted by the inverse number of authors. Treatment indicates computer-related keywords in the paper's full text. Controls: number of authors and NSF grants citing the work. Fixed effects: author, publication year, university, and primary topic. Standard errors clustered at the paper level (OpenAlex Work ID). Sample restricted to papers with at least one in-sample university affiliation and searchable full text in OpenAlex. Panel A outcomes summarize the age distribution of cited references. Panel B outcomes report the mean cited-reference Sleeping Beauty score B, the B score of the median-age cited reference, the B score of the oldest cited reference, and the mean B score among references in the oldest-age decile, following Ke et al. (2015).

	(1) Log Works	(2) Log Cites	(3) H-Index	(4) # Topics	(5) # Affiliations
Computer Adopter	1.364***	1.463***	7.752***	5.247***	1.820***
	(0.0165)	(0.120)	(1.254)	(0.294)	(0.166)
Number of Works		0.00536**	0.0577**	0.0120**	0.00617**
		(0.00176)	(0.0185)	(0.00399)	(0.00204)
Observations	316,970	316,970	316,970	316,970	316,970
R^2	0.309	0.437	0.533	0.292	0.327
Mean of Dep Var	2.732	5.158	12.16	16.77	3.525
Affiliation FE	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes
Subject FE	Topic	Topic	Topic	Topic	Topic

Table 3: Author pooled regressions

Notes: Computer Adopter is an indicator for authors with at least one paper flagged for computer-related keywords. Outcomes are measured over the author's entire OpenAlex-observed career: log total works, log total citations, H-index, number of distinct OpenAlex topics, and number of distinct affiliations. Columns (2)–(5) control for the author's total number of works. Affiliation is the author's modal affiliation across publications; cohort is the year of first publication in the dataset; subject fixed effects use the modal OpenAlex topic (lowest classification level). Standard errors clustered at the affiliation level. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

ing of computer adoption across universities. This university-year level analysis relies solely on temporal variation in adoption dates. Second, we extend this to a triple difference-in-differences (DDD) specification that incorporates pre-existing, (sub)field-level variation in computational demand within university-year cells, measured before computers were invented. This additional source of variation relaxes the parallel trends assumption required for the DiD approach and strengthens our causal identification.

University-level DiD. For a given university-year pair, we estimate the model:

$$Y_{ut} = \sum_{k \neq -1} \gamma_k \, \mathbf{1}_t\{t - G_u = k\} + \lambda_t + \alpha_u + \varepsilon_{ut}, \tag{1}$$

Where $Y_{u,t}$ denotes a the outcome of interest for university u in year t, such as total papers in Physics and Astronomy published. We define treatment ($k = 0$) as the year in which a university installs a computer for the first time (G_u). We include a full set of event dummies and their coefficients, $\sum_{k \neq -1} \gamma_k \, \mathbf{1}\{t - G_u = k\}$, all relative to the treatment year, along with university and year fixed effects (α_u, λ_t). The γ_k coefficients are the coefficients of interest, representing the average effect of adopting computers k years ago on the outcome Y. We normalize the effect $k = -1$ to zero, following standard practice (Borusyak et al., 2024).

We estimate the model using modern staggered-adoption DiD estimators that aggregate cohort-by-time 2×2 (treated cohort g vs. not-yet-treated at t) into event-time effects.[27] Our main estimator follows Callaway and Sant'Anna (2021), with de Chaisemartin and D'Haultfoeuille (2020) as a robustness check.

Identification requires parallel trends in research outcomes between treated and untreated universities, no anticipatory effects, and no cross-university spillovers (Callaway and Sant'Anna, 2021; Borusyak et al., 2024).

The historical context of early computing supports these assumptions. First, spillovers were minimal: as discussed in Section 2 and evidenced in Section 4.1, researchers could effectively use computers only at their home institutions, with external access severely limited. Any sharing or cross-institution collaboration that did occur would bias our estimates toward zero. Second, while computer acquisitions were sometimes announced a year before installation, researchers could not use them until they were physically installed and operational.[28]

The main residual concern is that computer installations may have coincided with broader research investment pushes, since universities acquiring computers might be attracting other forms of federal science funding in this period. Two features of the setting mitigate this. First, as discussed in Section 2.1, installations were often funded through equipment-specific channels rather than general research budgets, weakening the link between computer adoption and general university funding shocks. Second, we reserve the DiD itself for applications where generic university-level shocks are least likely to drive the outcome. When

[27] Our setting features staggered adoption with heterogeneous and steeply increasing post-adoption effects, which cause a standard Two-Way Fixed Effects estimator to deliver inconsistent and biased results. In such designs, TWFE event study coefficients are non-convex averages of 2×2 comparisons that (i) contaminate treated cohorts with already-treated units and (ii) place negative weights on some cohort-time effects, often attenuating or even reversing the true dynamics (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021). Consistent with this, our TWFE estimates flip the effects. Appendix F shows the Goodman-Bacon decomposition for a simple pre/post case and the resulting bias. We therefore report modern estimators that only compare a treated cohort to not-yet-treated units and aggregate cohort-time ATTs into event-time effects.

[28] To the extent that researchers strategically moved to universities because of computer installations, this would represent an interesting partial-equilibrium effect of early technology adoption, though general-equilibrium effects would be smaller.

estimating effects subject by subject and recovering the cross-subfield *gradient* with respect to pre-computer numerical intensity, a university-wide funding windfall can lift the levels of all subfields but would need to track that gradient to generate it spuriously; concretely, a confounding shock would have to be systematically correlated with subjects' 1940–1944 reliance on manual and mechanical calculation, a pre-determined characteristic with no obvious reason to co-move with contemporaneous federal science funding. Similarly, when estimating effects on the within-university composition of research methodologies, general funding shocks are unlikely to reallocate output toward simulation and numerical methods specifically unless they do so through the computational margin we are studying.

Triple-differences (DDD). To relax the parallel-trends requirement across treated and control universities, we exploit additional cross-subject variation. Fields differed sharply in their pre-computer computational demands – Numerical Analysis and Crystallography leaned heavily on slow mechanical or human calculation, while the Arts and Humanities did not – making researchers in the former much more exposed to the arrival of computing than those in the latter.

By comparing outcomes not only across universities but also *within* university across areas that differ in their numerical intensity, we can partial out any university-wide shock to treated universities.

The triple differences design formalizes this idea and exploits it to further identify the causal effect of computers on scientific research. For each subject area s, we first compute the share of "numerically intensive" papers in 1940-1944 (π_s^p), as defined in Section 4.4. We then create an indicator variable, E_s for whether a field is above or below the within-domain median of π^p.[29]

For a given university-year-subject triplet, we present the following regression specification that conceptually represents our triple-differences design (Olden and Møen, 2022):

[29]Let π_s^p be the base-period share of papers in subject s that contain the calculation-specific keywords defined in Section 4.4 (see Appendix A.5). We define the within-domain, above-median exposure indicator as

$$E_s^p = \mathbf{1}\Big\{ \pi_s^p \geq \operatorname{Med}\{\pi_{s'}^p : s' \in d(s)\}\Big\}.$$

In the baseline we use p = 1940-1944 and set $E_s \equiv E_s^{1940\text{-}1944}$.

$$Y_{ust} = \sum_{k \neq -1} \beta_k \, \mathbf{1}\{t - G_u = k\} \cdot E_s \; + \; \mu_{ut} \; + \; \eta_{st} \; + \; \delta_{us} \; + \; \varepsilon_{ust}. \tag{2}$$

Where we include university-by-year fixed effects (μ_{ut}), subject-by-year fixed effects (η_{st}), and university-by-subject fixed effects (δ_{us}). Because μ_{ut} absorb all university-year shocks (including the un-interacted event-time path), identification of $\{\beta_k\}$ comes entirely from *within* (u, t) contrasts between $E_s = 1$ and $E_s = 0$, net of discipline-wide shocks in η_{st}. The coefficients β_k therefore trace the dynamic *triple-difference* effect of computer adoption for numerically intensive subjects relative to non-intensive ones, normalized to zero at $k = -1$.

As before, we rely on difference-in-differences estimators by Callaway and Sant'Anna (2021) and de Chaisemartin and D'Haultfoeuille (2020). Operationally, our actual estimation is done in a two-step fashion: we first compute the within university difference in average outcomes for "numerically intensive" subjects vs. non-intensive ones, and we then estimate a standard DiD model using this difference as the outcome, following Olden and Møen (2022).[30]

Identification again rests on standard difference-in-differences conditions: (i) no anticipation; (ii) no cross-university spillovers; and (iii) parallel trends. The DDD version requires that, absent adoption, the *within-university* gap (exposed minus unexposed) would evolve in parallel across treated and not-yet-treated universities. University-wide shocks such as a funding windfall are differenced out unless they differentially shifted compute-amenable subjects (as pre-classified using 1940–44 numerical intensity) relative to less amenable ones *within the same university-year*.

We define exposure E_s using the subject-level frequency of manual and mechanical calculation in 1940–1944 rather than any contemporaneous measure. A contemporaneous measure would classify fields by how much they used computers after adoption, making the

[30]Formally, we define the average outcome among exposed subjects minus the average among unexposed subjects *inside* (u, t) as:

$$Y^*_{ut} \equiv \bar{Y}^{E=1}_{ut} - \bar{Y}^{E=0}_{ut}, \tag{3}$$

We then estimate an event study for Y^*_{ut} using Callaway and Sant'Anna (2021); de Chaisemartin and D'Haultfoeuille (2020), exactly as in (1) but with the outcome replaced by Y^*_{ut}. This yields the DDD dynamics while preserving the clean 2×2 comparison logic under staggered timing. Standard errors are clustered by university.

exposure a function of uptake itself. This creates a specific confounding channel: any force that caused a subject to rise in prominence in the 1950s or 1960s – a funding windfall, a policy push, the arrival of star researchers – could raise both its computer use and its research output, and a post-treatment measure would assign such a field to the high-exposure group and attribute its boom to computational intensity. The pre-1945 measure, by contrast, is fixed before digital computers existed and cannot be rewritten by any post-war shock. It captures a subject's latent demand for computation, measured at the field level across all universities rather than within specific institutions.

The remaining concern is that pre-1945 intensity could still proxy for persistent subject characteristics correlated with postwar shocks – most notably federal science funding flowing to quantitative disciplines. The η_{st} fixed effects absorb the national subject-year component of such shocks, but not potential university-specific differences in how funding flowed across subjects. To narrow the comparison, we estimate specifications within domains, so that the gradient comes from contrasts among subjects that share broader disciplinary trajectories and are therefore less susceptible to domain-level confounders. A related caveat is that identification relies on the 1940–1944 measure being predictive of later computer penetration *within* each domain (Section 4.4). Where this predictor is weak, first-stage power is limited and we should not expect sizable DDD effects; this is particularly relevant for the Life and Health Sciences, where the relationship between early numerical intensity and subsequent computer adoption is attenuated.

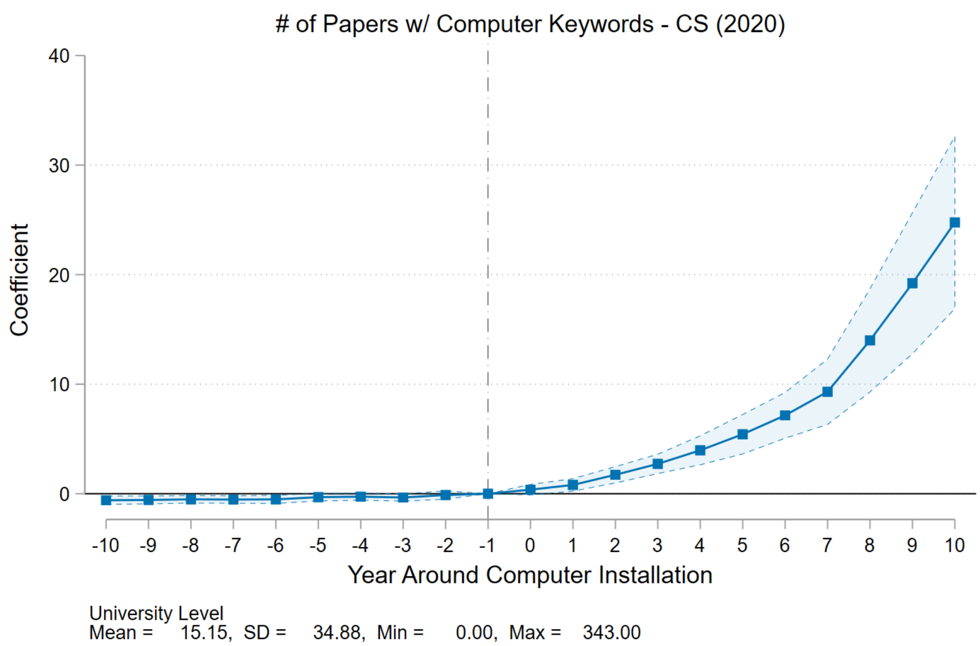
6 Results

This section presents our main causal findings. We first document computer uptake at treated institutions, then examine effects on the direction, quality, scope, and novelty of research, and finally turn to shifts in research methodologies.

6.1 Computer Usage

We first examine post-installation uptake using a university-year panel. Figure 7 plots event study coefficients from equation 1, estimated with Callaway and Sant'Anna (2021) to address treatment effect heterogeneity. Computer usage ticks up immediately and continues to grow, with treated universities publishing around 25 additional computer-intensive papers per year by year ten – about 10% of their mean annual output of 228 searchable publications.[31] The timing aligns closely with our hand-collected installation dates, validating the data.

Figure 7: Callaway-Sant'Anna event-study coefficients for the number of papers flagged with computer-related keywords at the university-year level. Sample: $N = 184$ treated universities, each aligned to event time 0 at its own computer-installation year. Points are aggregated cohort ATTs; shaded band is the 95% CI. The dashed vertical line marks installation.



Notes: Standard errors clustered at the university level.

Results are robust to using alternative estimators (de Chaisemartin and D'Haultfoeuille, 2020; Sun and Abraham, 2021) and to the exclusion of 'computer' as a standalone keyword.

[31]We use searchable publications in the denominator since keyword detection requires full text. Searchable papers are somewhat more likely to be in Physical or Social Sciences. See Section 4.2 and Appendix E.2.

6.2 Subject-Level Triple Differences

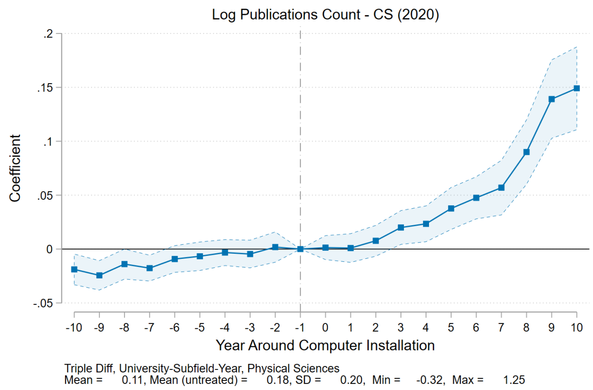
We study the effect of computers on research outcomes using the triple-difference (DDD) design described in Section 5. Our design exploits, at each event time k, the *within-university* gap between subfields that were ex ante more amenable to computation ("exposed") and less amenable ("unexposed") in treated universities, relative to the same gap in not-yet-treated universities (see Equation 3). Two implications follow. First, if computers raise outcomes broadly across a university ("lift all boats") or generate spillovers that also benefit less amenable subfields, our DDD will understate total effects. Second, any catch-up of initially less amenable subfields shrinks the exposed-unexposed gap over long horizons, biasing estimates toward zero.

Consistent with the weaker first-stage relationship between numerical intensity and computer penetration in the Life and Health Sciences (see Sections 4.4 and 5), we typically not observe DDD effects in these domains. We relegate full results for Life and Health Sciences to Appendix G.2.

Direction of science. Figure 8 reports DDD effects on *log* publication counts by domain. In the Physical Sciences, point estimates imply that, a decade after installation, computers increased publications by 15% on average, or 0.75 standard deviations, in numerically intensive fields relative to non-numerically intensive ones for treated universities. For context, the mean pre-treatment difference among not-yet-treated universities is only 0.014. In the Social Sciences, the effect is about 10% (0.77 SDs), from a pre-treatment baseline of -0.083. By contrast, the results for Health Sciences and Life Sciences are noisy and inconclusive, potentially reflecting the fact there is less variation and predictive power in the pre-computer numerical intensity proxy in those domains (see Appendix G.2).

Consistent with our results showing that computers caused numerically intensive areas to surge over non-intensive ones, we also observe that the number of authors for the former subjects increased significantly relative to the unexposed fields. On average, computers increased the difference in the number of unique coauthors between numerically-intensive vs non-intensive fields by 1.5 in the Physical Sciences, starting from a baseline of 0.09. For the Social Sciences, the effect was smaller, of 0.4 (baseline -0.37). These results are shown in Appendix G.2.

(a) Physical Sciences



(b) Social Sciences

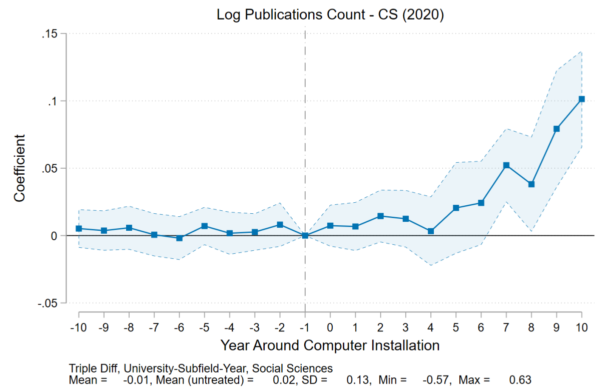


Figure 8: Triple-difference (DDD) event study for *log publication counts*, Physical and Social Sciences. Outcome Y^*_{ut} is the within-university gap in log pub counts between compute-amenable and less-amenable subfields (defined by the median 1940–1944 numerical-intensity share within each domain); estimator is Callaway-Sant'Anna on this gap, comparing treated cohorts with not-yet-treated. Sample: $N = 184$ treated universities; 252 subfields partitioned within domains. Points are aggregated cohort ATTs; shaded band is the 95% CI. The dashed vertical line marks installation.

Quality of science. We next study the effect of computers on the quality of research. To do this, we focus on the number of "top 1%" papers by citations (within their field) and at average citations per paper as measures of quality.

Figure 9 plots event study DDD coefficients using the yearly log of *top 1%* papers per university as the outcome. In the Physical Sciences, effects are large and persistent: at event time +10, on average, the exposed subfields see an increase of 2% in the number of top papers relative to the unexposed ones.[32]

Figure 10 looks at log average citations per paper. At +10, the Physical Sciences show an increase of about 0.20 on a baseline mean gap of 0.003, and the Social Sciences about 0.28 on a baseline of 0.0002. Results for Health and Life Sciences are imprecise and noisy (see Figure G.20, panels a–b). These intensive-margin gains align with the top-tail evidence in Figure 9 and strengthen the conclusion that computer adoption shifts research toward more

[32]In Figure G.21 (panels e–f), we show results hold whether we use the top 10% of papers instead of the top 1%.

computationally intensive, higher-impact subfields.

(a) Physical Sciences (b) Social Sciences

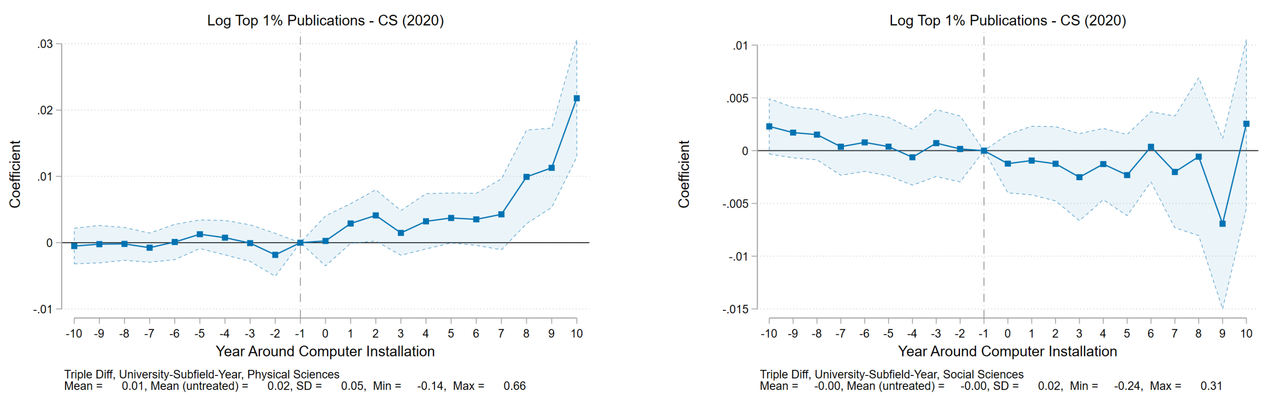


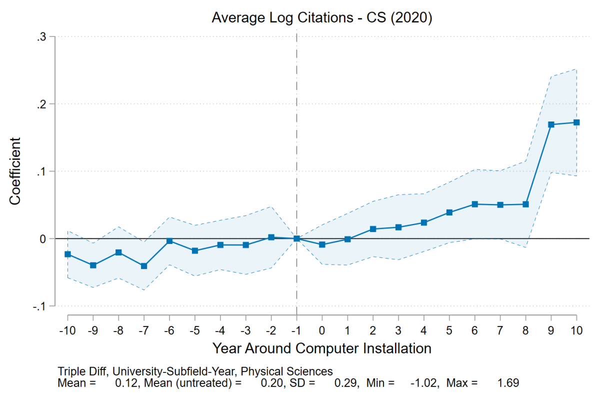
Figure 9: DDD event study for the *log count of top 1% papers* (by within-subfield citation rank), Physical and Social Sciences. Y^*_{ut} is the within-university gap between compute-amenable and less-amenable subfields; estimator is Callaway-Sant'Anna; sample is $N = 184$ treated universities. Points are aggregated cohort ATTs with 95% CIs; dashed line marks installation. Appendix Figure G.21 (panels e–f) shows the same pattern for the broader top 10% threshold.

Content of science. Finally, we focus on the actual contents of scientific research, exploring how computers changed the *breadth* and the *novelty* of the different topics being explored in academic works.

Figure 11 presents DDD estimates for the *average number of OpenAlex topics per paper*. The Physical Sciences exhibit a sustained increase of roughly 1.8 topics at +10 (baseline: 0.04), and the Social Sciences about 0.8 (baseline: -0.06), while effects in the Life and Health Sciences are muted or null (see Appendix G.2). These topic-level results mirror the concept-level findings and are consistent with weaker within-domain first-stage predictive power of early-1940s numerical intensity in Life and Health Sciences. Results are robust to using OpenAlex concepts instead.

To explore the novelty of research, following the concept-based construction in Section 4.5, for each paper we enumerate all topic (or concept) pairs and compute their historical co-occurrence frequency using the corpus from 1920 up to (but not including) the paper's publication year. A paper's novelty is the *maximum* pair-level novelty score, averaged within

(a) Physical Sciences



(b) Social Sciences

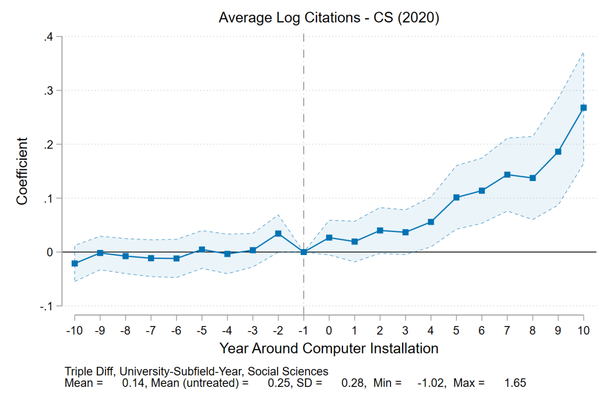
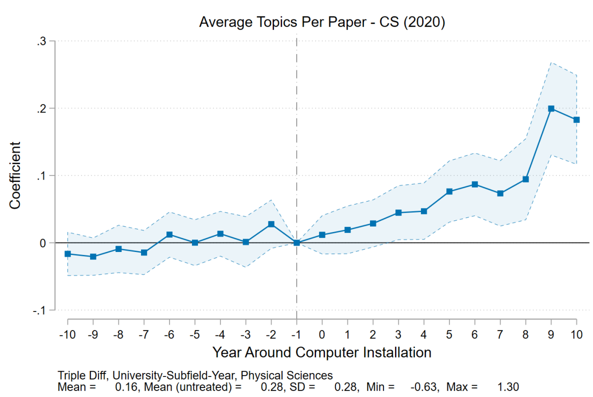


Figure 10: DDD event study for *log average citations per paper* (5-year citation window), Physical and Social Sciences. Y^*_{ut} is the within-university gap in log avg. citations between compute-amenable and less-amenable subfields; Callaway-Sant'Anna estimator; $N = 184$ treated universities. Points are aggregated cohort ATTs with 95% CIs; dashed line marks installation.

(a) Physical Sciences



(b) Social Sciences

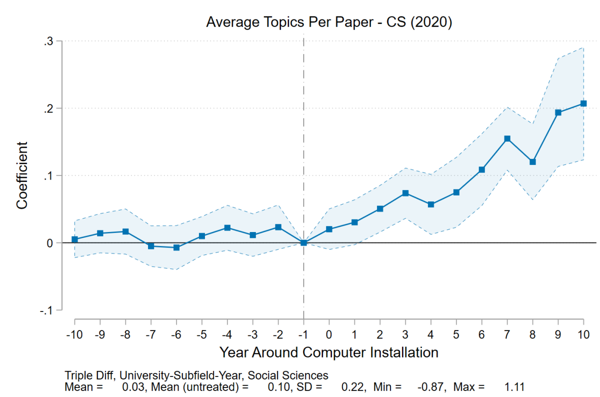


Figure 11: DDD event study for the *average number of OpenAlex topics per paper*, Physical and Social Sciences. Each paper in OpenAlex is tagged with a variable number of topics; the outcome averages this count within university-subfield-year. Y^*_{ut} is the within-university gap between compute-amenable and less-amenable subfields; Callaway-Sant'Anna estimator; $N = 184$ treated universities. Points are aggregated cohort ATTs with 95% CIs; dashed line marks installation. Read as breadth of topical coverage per paper.

university-subfield-year cells.[33] As with all DDD results, our coefficients capture the *relative* within-university shift toward compute-amenable subfields in treated universities versus not-yet-treated.

Figure 12 shows clear increases in topic-combination novelty in the Physical Sciences – about 2.0 points at event time +10 (against a baseline mean pre-treatment gap of 0.08) – and smaller but statistically meaningful gains in the Social Sciences of about 0.8 (baseline: -0.17). These patterns are consistent with computer adoption pushing research toward more unconventional topic pairings, especially in the Physical and Social Sciences. Finally, estimates for the Life Sciences and Health Sciences are close to zero (see Appendix G.2).

(a) Physical Sciences (b) Social Sciences

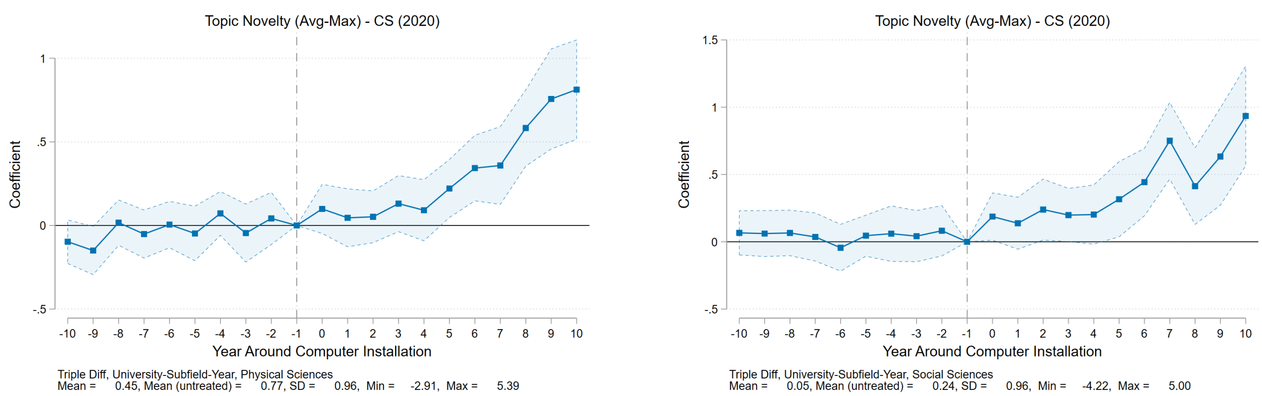


Figure 12: DDD event study for *topic-combination novelty*, Physical and Social Sciences. Each paper receives a novelty score equal to the maximum novelty across its pairs of OpenAlex topics (higher = more unusual combinations, based on their co-occurrence frequency in prior literature); scores are averaged within university-subfield-year. Y^*_{ut} is the within-university gap between compute-amenable and less-amenable subfields; Callaway-Sant'Anna estimator; $N = 184$ treated universities. Points are aggregated cohort ATTs with 95% CIs; dashed line marks installation.

All our results are robust to using de Chaisemartin and D'Haultfoeuille (2020) as an alternative estimator.

[33]Results are robust to alternative aggregations (e.g., averaging pair-level novelty within papers, taking the maximum across papers in a cell, or averaging across papers).

6.3 Narrowing Down on Specific Areas

In the previous section, we show that computers caused a shift in research production towards more numerically intensive subjects of research, particularly within the Physical and Social Sciences. This shift in output was also paralleled by similar increases in quality, topical breadth, and topic-combination novelty.

To further unpack these results, here we estimate separate event–study regressions at the university level for research outcomes in specific subjects. This approach lets us recover the absolute impact of computer adoption on each subject and determine whether our triple-differences estimates are driven by growth concentrated in numerically intensive areas, declines in less numerically intensive areas, or a combination of both.

Operationally, we estimate the university-year specification in (1) subject-by-subject, i.e., we replace Y_{ut} with $Y_{ut}^{(f)}$ (the outcome computed *only from papers in subject* f) and continue to use the Callaway and Sant'Anna (2021) cohort-aggregated DiD estimator to address staggered timing and treatment-effect heterogeneity.[34] This choice trades off some identification strength (we no longer leverage variation within u) but allows us to detect level effects that the triple-differences design, by construction, would difference out.

To summarize the subject-level estimates across many subfields, we use pooled binned-scatter plots of subfield-specific pre/post CS-DiD estimates against each subfield's pre-1944 share of papers mentioning non-generic numerical-computation keywords. In each figure, the underlying Physical and Social Science subfields are pooled into 20 equal-count bins, while the fitted line is estimated over the underlying subfields with publication-count weights.

Direction of science. Figure 13a shows, overall, that sizable post-installation increases in the logarithm of total papers are concentrated in more numerically intensive subfields. Using the fitted weighted relationship, subfields near the 90th percentile in pre-1944 numerical

[34] In Appendix F, we show how TWFE returns sometimes very different estimates to other more robust DiD estimators for even our regressions on computer-usage. A Goodman-Bacon decomposition (panel (c) of Figure F.16) shows that the TWFE estimator, in our scenario, inserts a lot of negative weights in the estimation, flipping the sign of the ATE. The results when using Sun and Abraham (2021) or de Chaisemartin and D'Haultfoeuille (2020) are in line with the ones presented here, further confirming the problems of TWFE for our setting.

intensity see publication effects of about 31%, versus about 17% for subfields near the 10th percentile.

Quality of science. Figure 13b reports effects on log citations per paper.[35] Similar to our results for output, we see a larger response for the most numerically intensive subfields, with computer adoption raising citations per paper by about 51% for subfields near the 90th percentile in numerical intensity, versus about 24% for those near the 10th percentile.

Taken together with publication counts, computers raised both the scale and the impact of research in the most compute-amenable areas. A pure cost-shifter would predict dilution and lower citations; a pure possibility-frontier expansion would predict quality gains without the rise in quantity. The joint increase is consistent with both margins moving at once: digital computing lowered the cost of existing work and expanded the set of feasible research paths – simulation, algorithmic or numerical work, and measurement-intensive analysis – so that researchers could act on valuable ideas that had been understood in principle but remained hard to execute.[36] Less numerically intensive subfields, on the other hand, show much more muted effects for the initial years of computer adoption.

Content and novelty of science. Figure 13c reports subfield event study estimates for the average number of OpenAlex topics per paper at the university-year level. Again the magnitude of the effects is strongly correlated with the numerical intensity of the subfields. Using the fitted weighted relationship, computers increase topics per paper by about 0.37 for subfields near the 90th percentile in numerical intensity, versus about 0.17 for those near the 10th percentile. To put this in perspective, the mean of the variable is 0.43.[37] Thus the fitted 10th–90th percentile comparison corresponds to roughly 0.20 extra topics per paper in the most numerically intensive subfields. Figure 13d shows the same positive gradient for

[35] We do not use count of top 1% of citations like we did for the triple differences because top 1% counts at the subfield level are very sparse and small, leaving little variation to generate point estimates.

[36] It is possible that the quantity effect in specific subject areas is entirely driven by researchers migrating toward "greenfield" areas where the research possibility frontier expanded. Nevertheless, if the quantity shift were merely due to reallocation, we would likely observe *negative spillovers* on other subjects, which we do not observe (see Figure 13a).

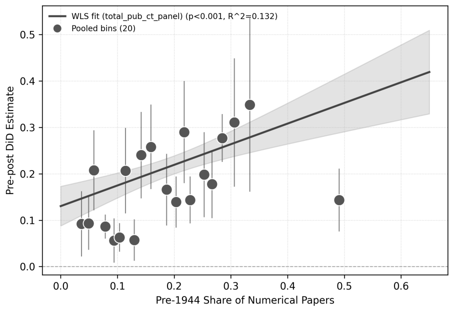
[37] The low value reflects the fact that some subfield-university-year bins have no publications. When this happens, we set the average topics per publications to 0.

topic-combination novelty. Using the fitted weighted relationship, computers raise the novelty score by about 1.94 for subfields near the 90th percentile in numerical intensity, versus about 0.82 for those near the 10th percentile, a difference of roughly 1.12 points.

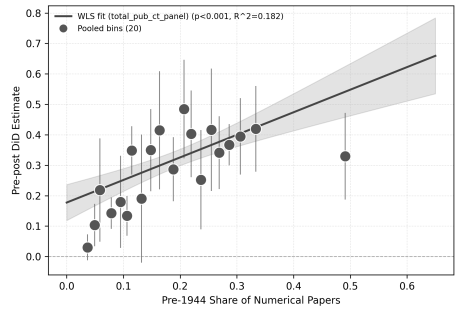
Across outcomes, the positive gradient in effects at the subject level strengthens our causal interpretation. Pre-computer numerical intensity predicts both greater computer adoption (as we show in Section 4.4) and larger post-adoption effects on downstream outcomes. This pattern suggests that computers themselves, rather than confounding factors, drove the observed transformations in scientific output.[38]

Finally, the positive gradient is robust to using de Chaisemartin and D'Haultfoeuille (2020) as an alternative estimator, to weighting or not weighting the fitted relationship by subject size, and to excluding very small subjects based on either pre-1944 keyword support or total panel output. The same qualitative pattern also appears when we aggregate from subfields to fields.

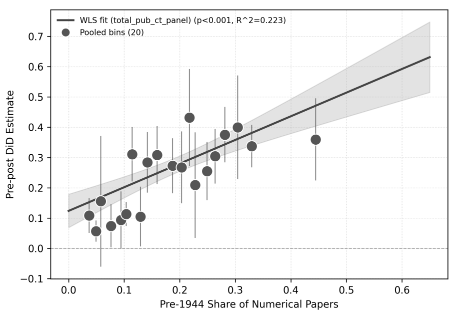
[38] If we were capturing other contemporaneous shocks correlated with computer adoption like funding windfalls or hiring sprees, we would not expect effects to follow the contours of latent demand for computation across subject areas.



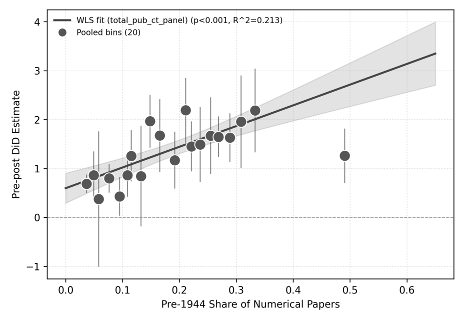
(a) Quantity of Papers (log)



(b) Citations per Paper (log)



(c) Topics per Paper



(d) Topic-Combination Novelty

Figure 13: Effect of computers on research outcomes by subfield pre-digital numerical intensity. Each dot is one of 20 equal-count bins of Physical- and Social-Science subfields (pooled; 137 subfields in panels a–b, 135 in c, 136 in d). y-axis: subject-specific pre/post Callaway-Sant'Anna ATT from a subfield-by-subfield university-year regression; outcomes are log publications (a), log citations per paper (b), average OpenAlex topics per paper (c), and average max topic-combination novelty (d). x-axis: subfield's 1940–1944 share of papers mentioning non-generic numerical-computation keywords. Red line is the weighted least-squares fit; bin means and fit both weight subfields by total 1940–1980 panel output. Vertical lines are 95% CIs for the binned means.

6.4 Research Methodologies

6.4.1 Classifying research methodologies.

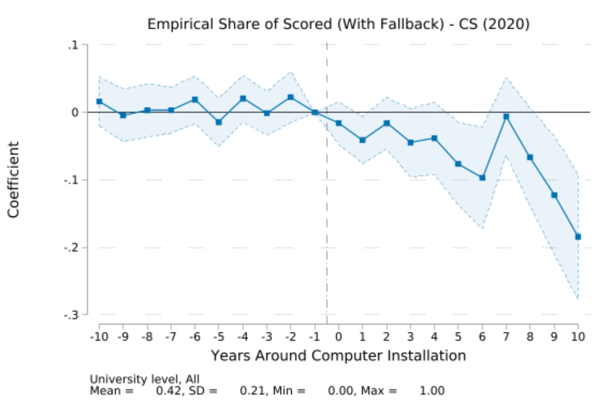
To study whether computers changed how universities conducted research, we classify papers by *methodology* rather than topic. Each paper receives a probability distribution over five categories – *empirical*, *theory*, *methods*, *simulation*, and *other* – and university-year outcomes aggregate this predicted mass rather than an argmax label. In the classified 1951–1969 slice, the average predicted mass is 48.3% empirical, 18.3% theory, 11.1% methods, 1.1% simulation, and 21.3% other. We build pseudo-ground truth from Gemini 3.0 Flash full-text reads on

177,243 papers, then train a gradient-boosting model to score papers when local full text is unavailable.[39] On a 10,000-paper full-text holdout, the model achieves a weighted F_1 of 0.825. Appendix B.1 reports details on the procedure.

6.4.2 Computers shift the composition of research methods.

Figure 14a shows that the empirical *share* of scored papers falls within treated universities after installation, even though raw empirical counts continue to rise – what changes is composition, not level. Figure 14b makes the reallocation explicit: theory, methods, and simulation gain share as empirical loses share. Appendix B.2 reports event studies for theory and methods, and paper-level regressions showing that computer-related papers are less empirical and more methods- and simulation-oriented, even after absorbing author, year, university, and topic fixed effects.

(a) Empirical share, event study



(b) Collapsed pre/post by category

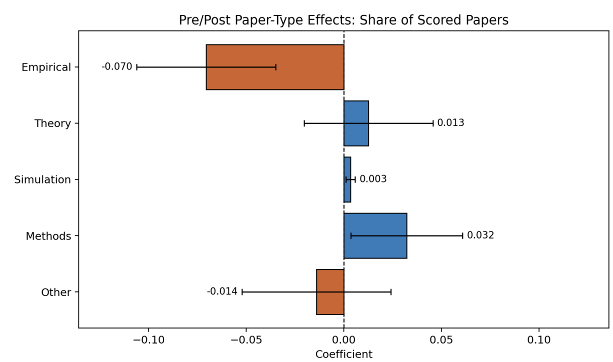


Figure 14: Effect of computer adoption on the methodology composition of university output, 1951–1969 classified panel slice (268,887 classified papers across 184 treated universities; five-bucket methodology taxonomy: empirical, theory, methods, simulation, other). Panel (a): Callaway-Sant'Anna event study for the *empirical share* of scored papers at the university-year level; companion event studies for theory and methods are in Appendix B.2. Panel (b): collapsed pre/post ATT estimates, one per methodology category (separate share regressions, so coefficients do not mechanically add to zero). 95% CIs; standard errors clustered by university. Read as a composition shift within the classified subset, not an exact adding-up decomposition.

[39]Our production classifier is a LightGBM model, a gradient-boosted ensemble of decision trees optimized for high-dimensional prediction (Ke et al., 2017). We report weighted F_1, the harmonic mean of precision and recall averaged across classes using class frequencies as weights.

6.4.3 What is rising after installation?

We validate the shift with two descriptive pre-post exercises: title-level TF-IDF and transparent full-text keyword families. Neither is a treatment-effect estimator. The content aligns with the composition shift: methods titles move toward *computer*, *algorithm*, *programming*, and *digital*; theory toward *model*, *systems*, *control*, and plasma-related language; simulation toward *computer simulation*, *electronic structure*, and *dynamics*. The keyword families in Figure 15 sharpen the same pattern – methods toward computer-systems and algorithmic or numerical content, theory toward control, optimization, and matrix methods, and simulation toward direct simulation language.

Empirical work looks different. Full-text reads (Appendix B.4) show computers enter empirical research through measurement, monitoring, signal analysis, and numerical processing of observed systems – not the regression-heavy style that dominates later decades. This helps explain why the empirical share falls even as counts keep rising: in this period, computers complement methods, theory, and simulation more than they do classically empirical work.

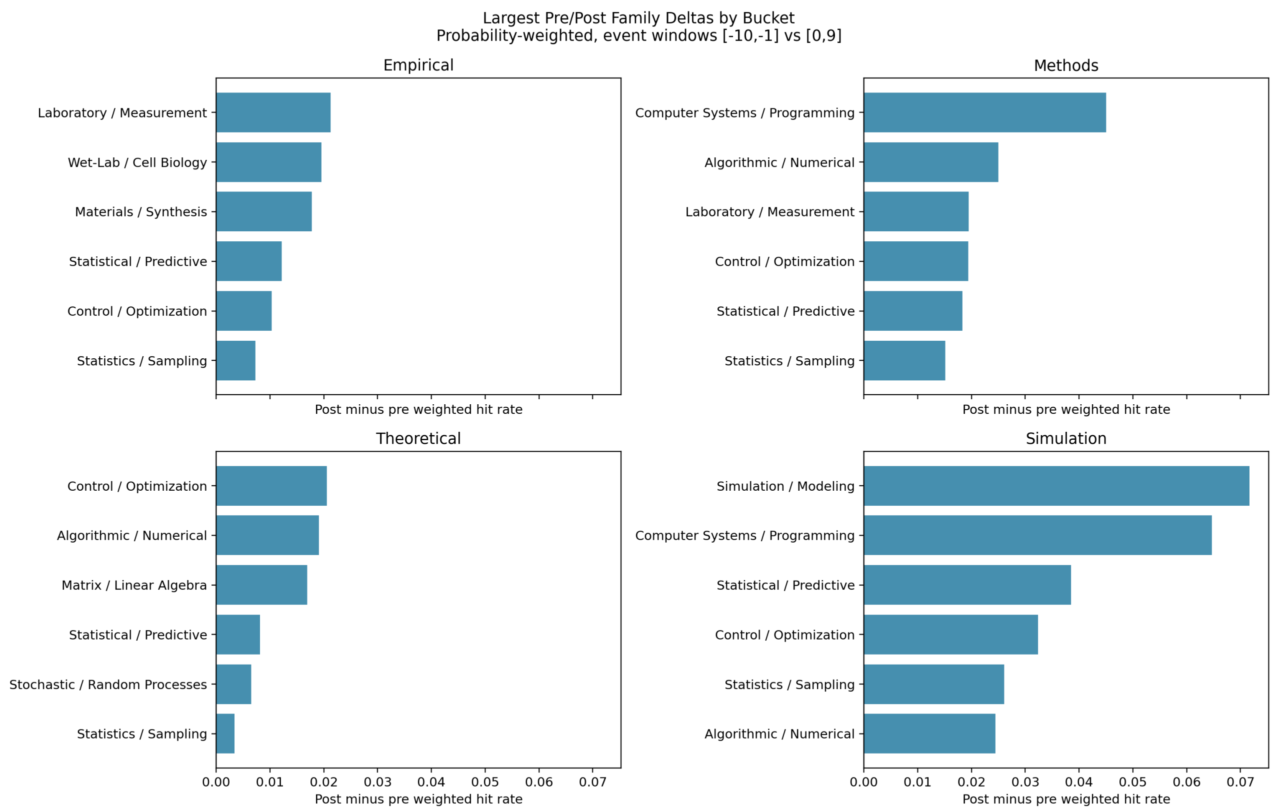


Figure 15: Pre/post differences in probability-weighted keyword-family hit rates within each methodology bucket (1951–1969 classified slice, 268,887 papers, 10-year window around installation). Rows are methodology buckets (empirical, theory, methods, simulation); columns are transparent keyword families (e.g., computer-systems, algorithmic/numerical, control/optimization; full definitions in Appendix Table 10; title-based TF-IDF companion in Appendix B.3). Cell color = post-minus-pre differential in the family's paper-weighted hit rate. Families are interpretive probes, not an exhaustive vocabulary partition.

7 Conclusion

We study the first diffusion of digital computers through American university research, from 1950 to 1970. Combining a new database of 2,200 university installations with full-text publication data, we exploit the staggered rollout of university computing centers and pre-digital differences across subjects in their reliance on numerical computation to identify the causal effects of computer access on science.

Computer papers carry a sizable citation premium, and the paper-level descriptives help explain why. Two patterns stand out. Computer papers are cited disproportionately as canonical sources of concepts and as formal methods, and less often as sources of established empirical findings. And their reference lists sit in a distinctive "hotspot" position: younger work at the center of the age distribution paired with a long tail of older, more dormant ideas (Mukherjee et al., 2017), a combination strongly associated with high-impact research. The underlying combinations are also more novel in the atypicality sense of Uzzi et al. (2013). The common thread is that computers let researchers act on ideas that had been scientifically understood but computationally infeasible to execute.

The triple-difference estimates confirm this reading. Within treated universities, numerically intensive subfields expand in both publication volume and citations per paper, with no quantity–quality trade-off – the pattern we would expect if computers relaxed a feasibility constraint rather than scaled up existing work. Subject-specific DiD estimates show that the size of the effect grows smoothly with pre-digital numerical intensity: subfields at the 90th percentile gain roughly twice as much in output (31% vs. 17%) and per-paper citations (51% vs. 24%) as those at the 10th percentile. The composition of research shifts in a way that echoes the descriptive evidence: simulation, methods, and computational theory rise, while the empirical share falls even as raw empirical counts continue to grow. Computational theory gains appear most clearly in fields like plasma physics, where governing equations had long been written down but not solvable. Where computers entered empirical work in this period, they did so through measurement and numerical processing of observed systems, not the regression-heavy style that would later dominate – and consistent with computer papers of this era functioning as canonical and methodological sources rather than as empirical findings.

The common thread is that computers unlocked research paths that were understood in principle but bottlenecked by computation. This single reading explains the joint rise in quantity and quality (the new paths carried genuine value rather than cheaper versions of existing work), the gradient with respect to pre-digital numerical intensity (the constraint had been binding where latent demand was highest), and the function of computer papers as methods and conceptual sources rather than as empirical findings or literature precedents (they encoded newly practical ways of doing something, not new things to know).

The gains also materialize quickly, which is telling. If ideas had been the binding constraint, new research paths could take years to formulate and more years still to mature. This is the long gestation lag that standard accounts of general-purpose technology diffusion emphasize. Instead, effects appear within years of installation and concentrated precisely where computational feasibility had been limiting, exactly the pattern we would expect if researchers had long known what to do and were waiting only on the means to do it. Where a new research tool meets latent demand – problems scientifically mature but computationally blocked – the scientific returns can be quick, concentrated, and visible in the record.

References

Agrawal, A. and Goldfarb, A. (2008). Restructuring research: Communication costs and the democratization of university innovation. *American Economic Review*, 98(4):1578–1590.

Ahmadpoor, M. and Jones, B. F. (2017). The dual frontier: Patented inventions and prior scientific advance. *Science*, 357(6351):583–587.

Azoulay, P. and Greenblatt, W. H. (2025). Does peer review penalize scientific risk taking? evidence from nih grant renewals. NBER Working Paper 33495, National Bureau of Economic Research.

Borjas, G. J. and Doran, K. B. (2012). The collapse of the soviet union and the productivity of american mathematicians. *The Quarterly Journal of Economics*, 127(3):1143–1203.

Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, 91(6):3253–3285.

Boudou, J. and Mckeon, J. (2024). Innovation under resource constraints: Supercomputing in scientific research.

Bresnahan, T. F. and Trajtenberg, M. (1995). General purpose technologies 'engines of growth'? *Journal of Econometrics*, 65(1):83–108.

Bryan, K. A. and Williams, H. L. (2021). Innovation: market failures and public policies. In *Handbook of industrial organization*, volume 5, pages 281–388. Elsevier.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2):200–230.

Campbell-Kelly, M., Aspray, W. F., Yost, J. R., Tinn, H., and Díaz, G. C. (1980). *Computer: A History of the Information Machine*. 4 edition.

Carnegie Commission on Higher Education (1973). *A classification of institutions of higher education: A technical report*. The Commission.

Ceruzzi, P. E. (2003). *A History of Modern Computing*. MIT University Press.

Charney, J. G., Fjörtoft, R., and Neumann, J. v. (1950). Numerical integration of the barotropic vorticity equation. *Tellus*, 2(4):237–254.

Comin, D. and Hobijn, B. (2004). Cross-country technology adoption: making the theories face the facts. *Journal of monetary Economics*, 51(1):39–83.

Comin, D. and Hobijn, B. (2010). An exploration of technology diffusion. *American economic review*, 100(5):2031–2059.

David, P. A. (1990). The dynamo and the computer: An historical perspective on the modern productivity paradox. *American Economic Review*, 80:355–361.

de Chaisemartin, C. and D'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110:2964–2996.

Gaetani, R. and Bergolis, M. L. (2015). The economic effects of scientific shocks. *Unpublished Manuscript*.

Gao, J. and Wang, D. (2023). Quantifying the benefit of artificial intelligence for scientific research. *arXiv preprint arXiv:2304.10578*.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of econometrics*, 225(2):254–277.

Grier, D. A. (2013). *When computers were human*. Princeton University Press.

Hill, R., Yin, Y., Stein, C., Wang, D., and Jones, B. F. (2021). Adaptability and the pivot penalty in science. *SSRN Electronic Journal*.

Jaffe, A. B. (1989). Real effects of academic research. *The American economic review*, pages 957–970.

Kantor, S. and Whalley, A. (2019). Research proximity and productivity: Long-term evidence from agriculture. *Journal of Political Economy*, 127:819–854.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, volume 30.

Ke, Q., Ferrara, E., Radicchi, F., and Flammini, A. (2015). Defining and identifying sleeping beauties in science. *Proceedings of the National Academy of Sciences*, 112(24):7426–7431.

Kendrew, J. C. (1963). Myoglobin and the structure of proteins: Crystallographic analysis and data-processing techniques reveal the molecular architecture. *Science*, 139(3561):1259–1266.

Krauss, A. (2026). *The Engine of Scientific Discovery: How New Methods and Tools Spark Major Breakthroughs*. Oxford University Press, New York.

Lin, Z., Yin, Y., Liu, L., and Wang, D. (2023). Sciscinet: A large-scale open data lake for the science of science research. *Scientific Data*, 10:315.

Liu, S. (2015). Spillovers from universities: Evidence from the land-grant program. *Journal of Urban Economics*, 87:25–41.

Mansfield, E. (1991). Academic research and industrial innovation. *Research policy*, 20(1):1–12.

Mokyr, J. (2002). *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton University Press, Princeton, NJ.

Moser, P. and Nicholas, T. (2004). Was electricity a general purpose technology? evidence from historical patent citations. *American Economic Review*, 94(2):388–394.

Mukherjee, S., Romero, D. M., Jones, B., and Uzzi, B. (2017). The nearly universal link between the age of past knowledge and tomorrow's breakthroughs in science and technology: The hotspot. *Science Advances*, 3:e1601315.

Myers, K. (2020). The elasticity of science. *American Economic Journal: Applied Economics*, 12(4):103–134.

Nagaraj, A. and Tranchero, M. (2023). How does data access shape science? evidence from the impact of u.s. census's research data centers on economics research.

National Research Council (1966). Digital computer needs in universities and colleges. Technical Report 1233, National Academy of Sciences – National Research Council, Washington, D.C.

National Science Foundation (1958). Eighth annual report for the fiscal year ended june 30, 1958. Government Printing Office Report. Eighth Annual Report of the National Science Foundation.

National Science Foundation (1959). Eighth annual report for the fiscal year ended june 30, 1959. Government Printing Office Report. Eighth Annual Report of the National Science Foundation.

Nordhaus, W. D. (2007). Two centuries of productivity growth in computing. *Journal of Economic History*, 67:128–159.

Olden, A. and Møen, J. (2022). The triple difference estimator. *The Econometrics Journal*, 25(3):531–553.

Park, M., Leahey, E., and Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. *Nature*, 613:138–144.

Picker, R. C. (2018). The invention of the computer. Conference abstract / draft chapter. Discusses wartime R&D, patent choices, and the ENIAC litigation.

Priem, J., Piwowar, H., and Orr, R. (2022). Openalex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv preprint arXiv:2205.01833*.

Pritchard, H. O. and Sumner, F. H. (1954). The application of electronic digital computers to molecular orbital problems. i. the calculation of bond lengths in aromatic hydrocarbons. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 226(1167):128–140.

Richardson, L. F. (1922). *Weather Prediction by Numerical Process*. Cambridge University Press, Cambridge.

Rosenberg, N. (1992). Scientific instrumentation and university research. *Research policy*, 21(4):381–390.

Rosenberg, N. and Trajtenberg, M. (2001). A general purpose technology at work: the corliss steam engine in the late 19th century us.

Schmallenbach, L., Bärnighausen, T. W., and Lerchenmueller, M. J. (2024). The global geography of artificial intelligence in life science research. *Nature Communications*, 15:7527.

Shvadron, D., Zhang, H., Fleming, L., and Gross, D. P. (2025). Funding the u.s. scientific training ecosystem: New data, methods, and evidence. NBER Working Paper 33944, National Bureau of Economic Research.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of econometrics*, 225(2):175–199.

Swade, D. (2001). *Difference engine: Charles Babbage and the quest to build the First Computer*. Viking Penguin.

Tamkin, A., McCain, M., Handa, K., Durmus, E., Lovitt, L., Rathi, A., Huang, S., Mountfield, A., Hong, J., Ritchie, S., Stern, M., Clarke, B., Goldberg, L., Sumers, T. R., Mueller, J., McEachen, W., Mitchell, W., Carter, S., Clark, J., Kaplan, J., and Ganguli, D. (2024). Clio: Privacy-preserving insights into real-world ai use.

Truffa, F. and Wong, A. (2024). Undergraduate gender diversity and the direction of scientific research.

Uzzi, B., Mukherjee, S., Stringer, M., and Jones, B. (2013). Atypical combinations and scientific impact. *Science*, 342:468–472.

War Department, Bureau of Public Relations (1946). Eniac announcement. Press release announcing the ENIAC, the first all-electronic general-purpose computer. Released for publication on February 16, 1946.

Wernsdorf, K., Nagler, M., and Watzinger, M. (2022). Ict, collaboration, and innovation: Evidence from bitnet. *Journal of Public Economics*, 211:104678.

A Data Construction

The work to compile the dataset involved two main steps:

1. **Processing surveys**: this step involved finding, obtaining physical copies, scanning, tabulating, cleaning the surveys. Finally, we harmonized university and computer model names, and matched universities to a global identifier. In the end of this step, we get 1,180 universities.[40]

2. **Processing universities**: after the base data was processed, each university was consolidated and verified manually. For each university from step 1, there is a mean of 6.7 snapshots across surveys. The consolidation process involves collapsing several snapshots of the same computer at a university into a single entity. We fill in remaining gaps in installation years and locations with university sources, either online or archival. In this step, we consolidate and prepare for analysis our final 184 universities.

A.1 Computer Dataset Variables

The computer installations dataset contains the following variables:

1. **University**: Research Organization Registry (ROR)[41] and university name;

2. **Computer Model**:

   (a) Department/unit of university where computer was installed;

   (b) Year and month of installation;

   (c) Year and month of decommission;

   (d) Average hours of usage per month;

[40]This includes all doctoral-granting US universities. We also have data for several universities outside the US that were tabulated but not processed further.

[41]The Research Organization Registry (ROR) is a global, community-led registry that provides open persistent identifiers for research organizations. These identifiers facilitate the disambiguation of institutional affiliations and the accurate linking of the computer installations dataset to the publication data. For more information, check https://ror.org/about/

(e) Flag of whether computer is analog;

(f) Flag of whether computer was built in-house;

(g) List of survey-years that report the computer.

3. **Miscellaneous**: we also include additional information that is not systematically reported:

(a) Flag of whether computer was commissioned by the military;

(b) Flag of whether the computer was a donation.

Other relevant but unstructured information is provided in comments.

A.2 List of Sources

Figure A.1 shows the timeline coverage of our survey sources:

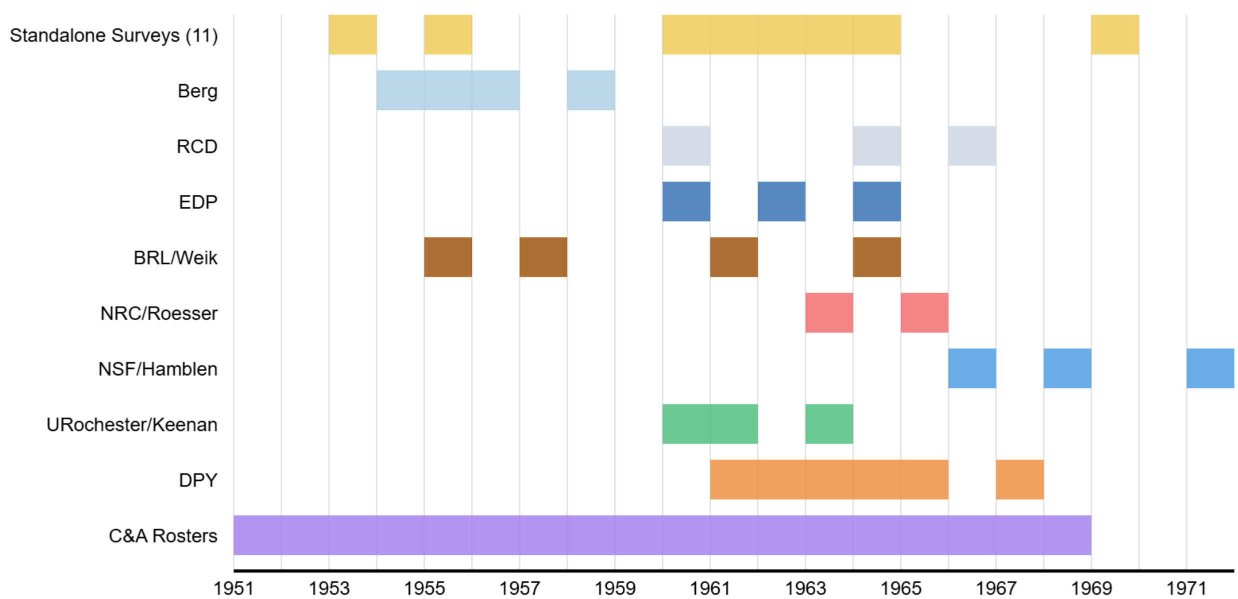


Figure A.1: Yearly coverage of surveys in the computer-installations database, 1950–1971. Each row is a distinct survey source; shaded cells indicate the year(s) in which that source published a computer census or inventory used to extract installations. Database draws on 82 survey-year pairs from 24 distinct sources. See Appendix A.2 for the full list of sources and acronyms.

Coverage of surveys varies substantially, and so does coverage within universities. The median survey in sample reports computers from 72 universities, with the 25th percentile surveying just 15 universities. Computers & Automation, for example, covered just a handful of universities per year before 1961, when it started a separate survey dedicated to universities. It also typically just reported computers within computer centers; computers installed in specific departments were not reported. Finally, all sources except for the Southern Education Board/NSF surveys conducted by John Hamblen do not consistently report installation dates for computers.

To supplement the survey information, we contacted 74 university archives and collected around 10,000 pages of documentation about the workings of university computer centers and installations.

As reported in Section 3.1, we collect data from 24 survey sources comprising 82 survey-year pairs. Our main sources are:

1. Computers and Automation – **Rosters of Organizations in the Computer Field** (1951-1968). Computers and Automation was a general interest magazine about computing technology. Rosters were published in July for each year. The magazine started a separate roster of school, college and university computer centers in 1961. Before that, only a handful of universities were mentioned in the general roster of organizations in the computer machinery field. Source acronym: *c&a*

2. Inventory of Computers in U.S. Higher Education - **Computers in higher education**: report by the John W. Hamblen of the Southern Regional Education Board commissioned by the NSF (1966; 1968; 1971). This is the only source that universally reports computer installation dates. Source acronym: *hamblen*

3. Thomas A. Keenan Surveys – **University of Rochester Annual Survey of University Computing Centers** (1960; 1961; 1963). Keenan conducted six surveys of university computer centers with detailed data for each, starting in 1957. We were able to recover the last three surveys thanks to Melissa Mead at the University of Rochester archives, but the first three seem to have been lost to time. Source acronym: *keenan*

4. National Research Council's Roesser Report - **Digital Needs in Universities and Colleges** (1966; covers 1963 and 1965). This is a survey within the context of a thorough NRC report about computer usage in higher education. Source acronym: *nrc*

The full list, including minor and supplemental sources, of sources is available upon request.

A.3 List of Universities

The 184 universities in our sample are listed in Table 4 and Table 5.

Table 4: Universities In Sample

Universities	
Abilene Christian College	Ohio University
American University	Oklahoma State University
Amherst College	Oregon State University
Arizona State University	Pennsylvania State University
Auburn University	Pomona College
CUNY, Baruch College	Princeton University
Baylor University	Providence College
Boston College	Purdue University
Boston University	Queensborough Community College
Brandeis University	Rensselaer Polytechnic Institute
Brigham Young University	Rice University
Brown University	Rose Polytechnic Institute
Bryn Mawr College	Rutgers University
California Institute Of Technology	Saint Louis University
California State University, Los Angeles	San Diego State University
Carnegie Institute Of Technology	Smith College
Carnegie Mellon University	South Dakota State University
Case Institute Of Technology	Southern Illinois University
Case Western Reserve University	Southern Methodist University
CUNY, City College	Stanford University
Clark University	SUNY At Buffalo
Clemson University	Stephen F. Austin State College
College Of William And Mary	Stevens Institute Of Technology
Colorado School Of Mines	Swarthmore College
Colorado State University	Syracuse University
Columbia University	Temple Junior College
Cornell University	Texas A&M University
Dartmouth College	Texas College Of Arts And Industries
Duke University	Texas Tech
Emory University	The King's College - Pennsylvania
Fairleigh Dickinson University	The Ohio State University
Florida State University	Tufts University
Foothill College	Tulane University
Fordham University	University Of Akron
Franklin Institute	University Of Alabama
George Washington University	University Of Alaska
Georgetown University	University Of Arizona
Georgia Institute Of Technology	University Of Arkansas
Georgia State University	UC, Berkeley
Harvard University	UC, Davis
Harvey Mudd College	UC, Irvine
Haverford College	UC, Los Angeles
Howard University	UC, Riverside
Illinois Institute Of Technology	UC, San Diego
Indiana Institute Of Technology	UC, San Francisco
Indiana University, Bloomington	UC, Santa Barbara
Iowa State University	UC, Santa Cruz
Jackson State College	University Of Chicago
Johns Hopkins University	University Of Cincinnati
Kansas State University	University Of Colorado Boulder
Kent State University	University Of Connecticut
Lehigh University	University Of Delaware
Long Island University	University Of Denver

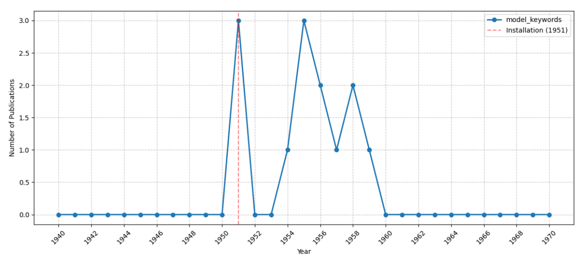
(Continued on next page)

Table 5: Universities In Sample (continued)

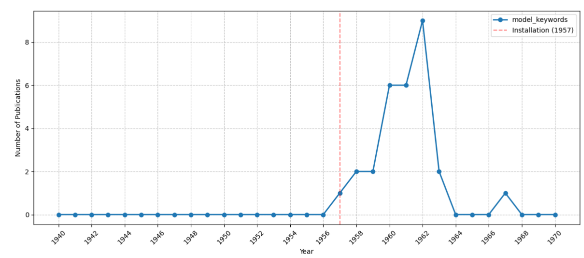
Universities	
Louisiana State University	University Of Florida
Lowell Technological Institute	University Of Georgia
Marquette University	University Of Hawaii
Massachusetts Institute Of Technology	University Of Houston
Michigan State University	University Of Idaho
Mississippi State University	University of Ill., Urbana-Champaign
Missouri Univ Of Science And Tech	University Of Iowa
Montana State University	University Of Kansas
New Mexico Inst Of Mining And Tech	University Of Kentucky
New Mexico State University	University Of Louisville
New School For Social Research	University Of Maine
NY St Col Of Agriculture At Cornell U	University Of Maryland
New York University	University Of Massachusetts, Amherst
North Carolina State University	University Of Miami
North Dakota State University	University Of Michigan, Ann Arbor
Northeastern University	University Of Minnesota
Northern Illinois University	University Of Mississippi
Northwestern University	University Of Missouri
University Of Nebraska	University Of Pennsylvania
University Of Nebraska, Omaha	University Of Pittsburgh
University Of Nevada	University Of Puerto Rico, Mayagüez
University Of New Hampshire	University Of Puerto Rico, Río Piedras
University Of New Mexico	University Of Puget Sound
UNC, Chapel Hill	University Of Rhode Island
University Of North Dakota	University Of Rochester
University Of Notre Dame	University Of South Carolina
University Of Oklahoma	University Of South Florida
University Of Oregon	University Of Southern California
Vanderbilt University	University Of Southwestern Louisiana
Vassar College	University Of Tennessee
Virginia Polytechnic Institute	University Of Texas, Austin
Washington And Lee University	University Of Utah
Washington State University	University Of Vermont
Washington University Of Saint Louis	University Of Virginia
Wayne State University	University Of Washington
Wesleyan University	University Of Wisconsin, Madison
West Virginia University	University Of Wisconsin-Milwaukee
Western Michigan University	University Of Wyoming
Western Reserve University	Utah State University
Wichita State University	Yale University
Williams College	

A.4 Computer-Intensive Research

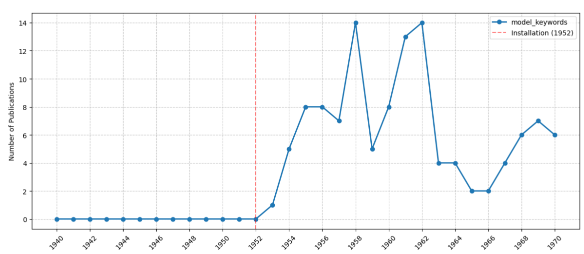
To validate treatment timing, we compare the first appearances of specific computer models in university-affiliated papers with our recorded installation dates. Figure A.2 plots mentions of the first computers installed at MIT (Whirlwind), Northwestern (IBM 650), UIUC (ILLIAC I), and Michigan (MIDAC). First-year mentions track installation dates closely. The absolute counts are small and should be read as a lower bound: explicit model-name references are common for early, distinctive machines but largely absent for later, commoditized ones like the IBM 1620.



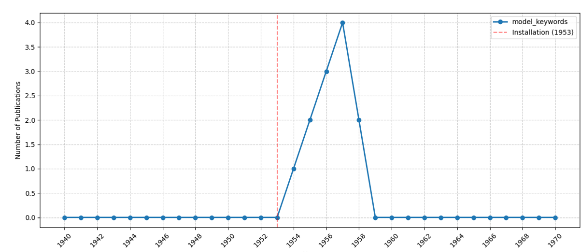
(a) Massachusetts Institute of Technology



(b) Northwestern University



(c) University of Illinois, Urbana-Champaign



(d) University of Michigan, Ann Arbor

Figure A.2: Raw yearly counts of paper mentions for each computer model (by name) at four example treated universities. Series are extracted from a dictionary of brand- and model-specific keywords applied to the full text n-grams of works affiliated with each university in OpenAlex, 1940–1980.

A.5 Computer-Related Keywords

The full list of keywords used to flag papers that use computers is:

- (computer), electronic computer, digital computer, automatic computer, high-speed computer, mainframe computer, high-speed computing device, electronic brain, data processing equipment

- computer program, computer algorithm, programming language, fortran, cobol

A paper is thus flagged as using or mentioning computers if any of those keywords is present in the OpenAlex searchable n-grams from paper full text, which is inherited from the Internet Archive's General Index. As noted above, because the word "computer" is possibly confounding for human computers in the early days,[42] we run the analysis with and without it and confirm results are robust to the inclusion/exclusion of "computer" as a keyword.

For the LLM analysis, as described in Section 3.2.1, we first screen papers for computer usage or mention using a set of high recall keywords in a hand-annotated sample of 200 papers. For screening purposes, we add to the set of keywords above the following keywords to improve recall:

- calculator, computing, computational, punch card, punched card, Monte Carlo, simulation, algol, algorithm, program, programming
- electronic machine, analog computer, electromechanical computer, automatic equipment, data processing, edp, adp, ibm, univac, burroughs

To flag numerically intensive tasks across research areas, as described in Section 4.4, we screen papers for several keywords:

- **Manual/hand calculation:** manual calculation, computed by hand, hand computation, longhand calculation, checked by hand
- **Proto-calculators & punched cards:** punched card, Hollerith, keypunch, desk calculator, mechanical calculator, adding machine, comptometer, Friden, Marchant, Brunsviga, Burroughs, tabulating department, machine accounting
- **Linear algebra and systems:** matrix inversion, matrix multiplication, Gaussian elimination, Gauss-Jordan, normal equations, eigenvalue, eigenvector
- **Numerical methods and equations:** Runge-Kutta, Newton-Raphson, finite difference, difference equation, differential equation, simplex method

[42]Normally, human computers were referred to as "computors" and not "computers."

- **Analog instruments:** differential analyzer, network analyzer, harmonic analyzer, slide rule, nomogram, integrator, analog simulator, model board, patch board, A-C network analyzer
- **Statistical analysis:** least squares, maximum likelihood, log-likelihood, regression analysis, ANOVA, principal component analysis, probit, logit, sample size, data analysis
- **New computational uses:** numerical simulation, stochastic simulation, numerical experiment, random number table, pseudo-random, Monte Carlo

We also include obvious morphological and hyphenation variants (e.g., plurals).

B Research Methodologies

B.1 Classifier Construction and Taxonomy

We classify research methodologies using titles, metadata, embeddings, and full-text-derived signals. The pseudo ground truth comes from Gemini full-text reads, after which we train LightGBM to score papers even when local full text is unavailable. The production classifier is evaluated both on the rebuilt validation split and on a separate 10,000-paper full-text hold-out. Table 6 summarizes these benchmark scores and the corresponding analysis universe.

Table 6: Methodology Classifier Benchmarks and Sample Accounting

Benchmark	Accuracy	Weighted F1	Macro F1
Main LightGBM (validation)	0.752	0.749	0.657
10k full-text benchmark	0.824	0.825	0.760
Fallback no-lexicon model	0.814	0.814	0.751

Universe	N
Unique papers in 1951–1969 panel slice	470,395
Classified unique papers	268,887
Keyword-matched computer papers in panel slice	25,665
Keyword-matched computer papers among classified	17,667
Classified papers with local full text	100,378
Already in Gemini first-pass set	56,650

Notes: The main production model is LightGBM. The benchmark table keeps only the validation score we use in the paper, the 10k full-text benchmark, and the no-lexicon fallback. Universe counts refer to the 1951–1969 panel slice used for the methodology diagnostics.

The raw classifier predicts seven classes, which we collapse into the five buckets used in the main text. Computational papers are rare in this period and are folded into the simulation bucket; the *other* category is mostly reviews, bibliographies, and similar reference material rather than a generic model-failure residual.

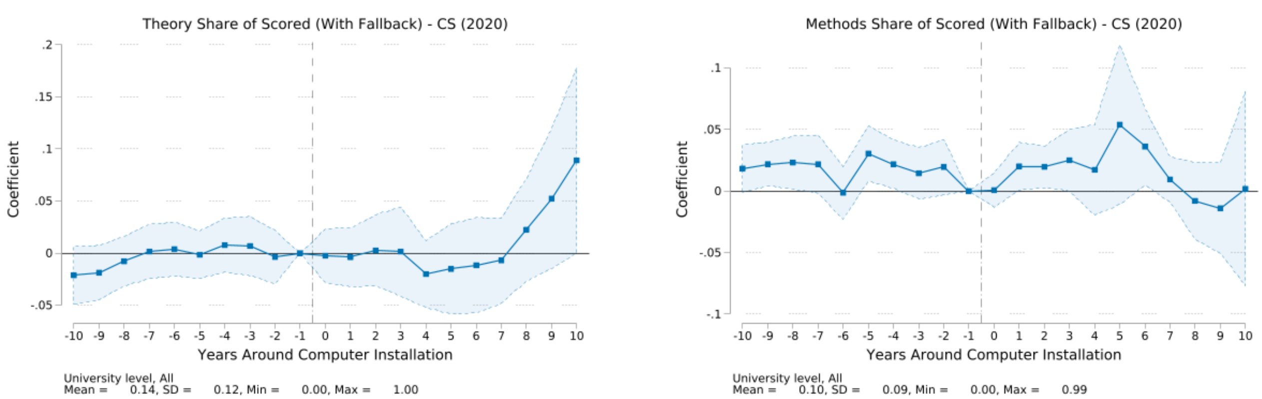
Table 7: Raw Seven-Class Taxonomy and Five-Bucket Mapping

Raw class	Mass share	Maps to
Quantitative empirical	31.1%	Empirical
Qualitative empirical	17.2%	Empirical
Formal theoretical	12.9%	Theory
Discursive theoretical	5.4%	Theory
Methods	11.1%	Methods
Computational	1.1%	Simulation bucket
Other	21.3%	Other

Notes: Shares are average predicted mass across the 268,887 classified papers in the 1951–1969 panel slice. The paper's main methodology shares use the collapsed five-bucket taxonomy and aggregate probability mass rather than argmax assignments.

B.2 Additional Paper-Type Results

Figure B.3 reports the companion event studies for theory and methods. Both series rise after installation, complementing the empirical-share decline in the main text. Table 8 then moves to the paper level and shows that computer-related papers are less empirical, more methods- and simulation-oriented, and relatively more theoretical than empirical after controlling for author, year, university, and topic fixed effects.



(a) Theory share of scored papers

(b) Methods share of scored papers

Figure B.3: Companion event studies to Figure 14: Callaway-Sant'Anna estimates for the theory share (panel a) and methods share (panel b) of scored papers at the university-year level. Shares are measured within the classified subset, so they capture reallocation rather than changes in scoring coverage. 95% CIs; standard errors clustered by university.

Table 8: Paper-Level Methodology Regressions

	Empirical	Theory	Simulation	Methods	log(Emp./Theory)
Computer-Keyword Flag	-0.043***	-0.018***	0.028***	0.036***	-0.034***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
# Authors	0.028***	-0.015***	0.000	0.000	0.049***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
NSF grants (paper)	0.029***	-0.005*	0.001	-0.006*	0.038***
	(0.005)	(0.002)	(0.001)	(0.003)	(0.007)
Observations	2,128,533	2,128,533	2,128,533	2,128,533	2,128,533
R^2	0.798	0.816	0.572	0.667	0.841
Mean of Dep Var	0.466	0.146	0.009	0.124	0.375
Author/Year/Univ FE	Yes	Yes	Yes	Yes	Yes
Subject FE	Topic	Topic	Topic	Topic	Topic

Notes: Paper-author observations, weighted by the inverse number of authors. Treatment indicates computer-related keywords in the paper full text. Controls: number of authors and NSF grants citing the work. Fixed effects: author, publication year, university, and primary topic. Standard errors are clustered at the paper level (OpenAlex Work ID). Negative values in $\log(\text{Emp./Theory})$ indicate relatively more theoretical than empirical mass.

B.3 Title-Based Validation

To validate that the methodology shifts correspond to plausible computer-related content, we compare title language before and after adoption within each bucket. We compute document-level TF-IDF for each title and then contrast treated-post with treated-pre papers within the 1951–1969 classified slice. Unlike the family dictionaries below, this exercise spans the *full universe of title vocabulary*. It is a descriptive validation device rather than another DiD design. Figure B.4 shows the strongest post-treatment title terms, while Table 9 highlights the most representative terms by bucket.

Table 9: Representative Post-Treatment Title Terms by Bucket

Bucket	Representative post > pre title terms
Empirical	spectral analysis, measurements, detection, prediction, correlation
Methods	computer, algorithm, programming, digital, automated
Theory	model, systems, control, optimal, plasma
Simulation	simulation, computer simulation, electronic structure, dynamics, wavefunctions

Notes: These are representative terms drawn from the post > pre title contrasts. The full title exercise ranges over the entire title vocabulary, not a hand-picked lexicon.

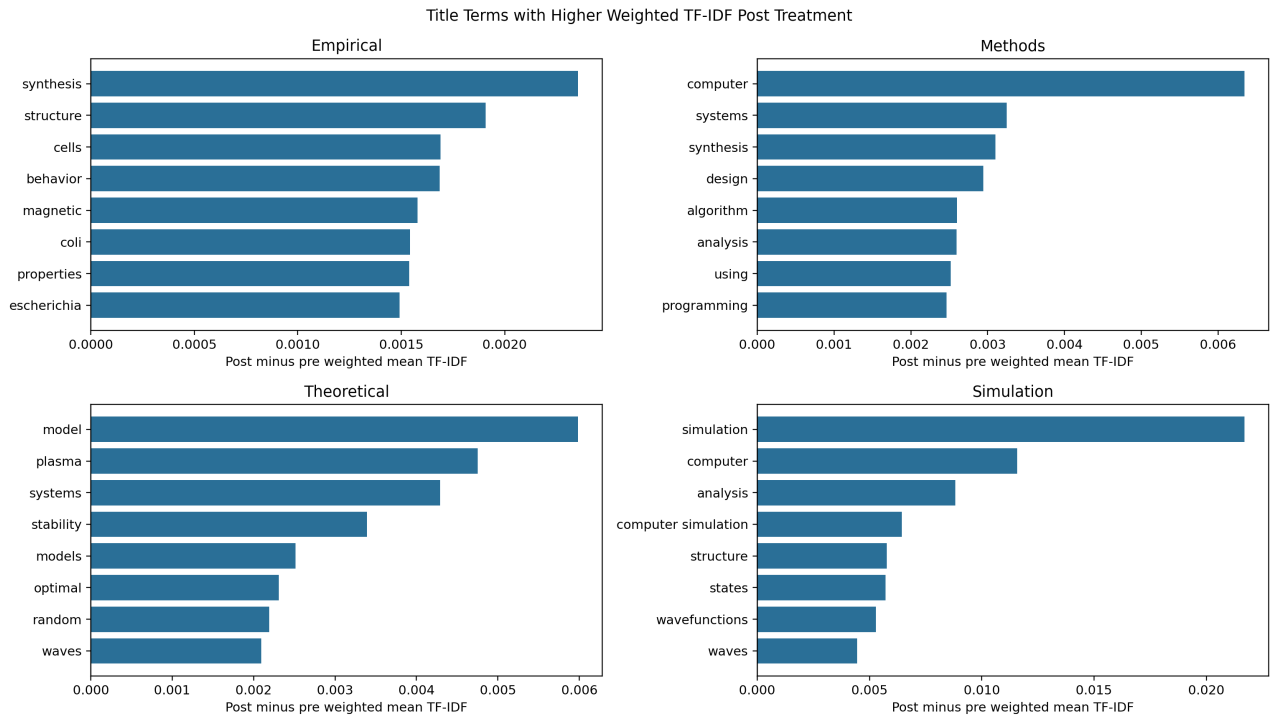


Figure B.4: Title terms that rise most, post minus pre, by methodology bucket (1951–1969 classified slice, 268,887 papers; 10-year pre/post window around each university's computer-installation year). Within each bucket row, bars rank title tokens by the difference between the post-treatment and pre-treatment probability-weighted mean TF-IDF (i.e., the paper's methodology-bucket probability is used as the weight). Bar color indicates the bucket. Computed over the full title vocabulary, not a hand-picked lexicon; representative examples drawn from these rankings appear in Table 9.

Notes: The empirical row is noisier than the others, but dominant post-treatment terms still point toward measurement, detection, and prediction. On the theory/simulation side, the appearance of terms such as *plasma*, *model*, *systems*, and *computer simulation* matches historically computer-intensive areas.

B.4 Full-Text Family Validation and Manual Examples

The full-text family exercise is narrower and more interpretable. We build transparent keyword families from manual reading of full-text packets and then apply them to the refreshed local full-text base. After integrating the consolidated `fulltext_master` repository, the methodology diagnostics now have usable local snippets for 111,842 paper-university rows in the 1951–1969 classified slice. These families are *not* intended to partition the full vocabulary. Instead, they are selective probes used to interpret the kinds of content that rise after adoption.

Table 10 reports the main family definitions. Table 11 gives representative manual examples. The empirical examples reinforce a key point from the main text: empirical computer use in this period mostly reflects measurement, monitoring, signal analysis, diagnosis, and numerical processing of observed systems, rather than the later regression-heavy style familiar from modern social science.

Table 10: Transparent Family Dictionaries Used in the Validation Exercise

Family	Definition and example language
Computer systems / programming	Explicit computer, compiler, programming-language, assembler, time-sharing, or computer-center language. Example terms: computer, compiler, ALGOL, program, time-sharing.
Algorithmic / numerical	Algorithms, numerical procedures, matrix computation, finite-difference, and finite-element schemes. Example terms: algorithm, numerical, matrix, finite element.
Simulation / modeling	Simulation, simulation models, numerical forecasting, or model-based experimentation. Example terms: simulation, model, modeling, dynamics.
Control / optimization	Feedback, control, queueing, encoding, and related systems or optimization problems. Example terms: control, optimal, feedback, queueing, encoding.
Laboratory / measurement	Experimental apparatus, assay, protocol, or measurement language that often accompanies computer-assisted recording and monitoring. Example terms: apparatus, assay, procedure, perfusion, measurement.
Statistical / predictive	Quantitative processing of observed data. Example terms: regression, statistical, variance, correlation, prediction.
Biomedical signal / detection	Empirical signal-processing and diagnostic language. Example terms: electrocardiogram, heart sounds, body-surface potential, detection, spectral analysis.
Statistics / sampling	Statistical inference, regression, variance analysis, or survey/sampling design. Example terms: statistical, regression, sampling, inference.
Differential equations	PDE/ODE and boundary-value or continuum-model problems. Example terms: differential equation, boundary value, PDE, ODE.
Materials / synthesis	Chemistry and materials language used as a placebo-style comparison family rather than a claim about the main computer mechanism. Example terms: chemical synthesis, enzymatic synthesis, polymer, alloy.

Notes: Families are transparent dictionaries used for descriptive interpretation only. They were built from manual reading of full-text papers in the methods, theory, simulation, and empirical-computer subsets, then applied to the refreshed local full-text base.

Table 11: Illustrative Manual Examples behind the Methodology Labels

Empirical computer use

- *Study of High Frequency Components in Electrocardiogram by Power Spectrum Analysis* (W2003343840): measured biomedical signals processed with spectral analysis.
- *The Detection of Heart Disease in Children* (W2974997846): tape-recorded heart sounds used for computer-assisted screening and classification.
- *The Prediction of Flow Patterns, Liquid Holdup and Pressure Losses Occurring During Continuous Two-Phase Flow In Horizontal Pipelines* (W1975029220): numerical prediction built on observed engineering measurements.

Methods

- *The structure of yet another ALGOL compiler* (W2050572856): compiler architecture and implementation.
- *Programming Technique: An improved hash code for scatter storage* (W2003248512): programming-method paper on data storage and retrieval.
- *The simplex method of linear programming using LU decomposition* (W2071877138): numerical optimization method built for computation.

Computer-adjacent theory

- *Feedback Queueing Models for Time-Shared Systems* (W2074065133): queueing/control theory aimed at time-sharing computer systems.
- *Levels of computer systems* (W1983161084): formal systems-theory discussion explicitly about computer architecture.
- *Source encoding in the presence of random disturbance* (W1584278176): information/encoding theory that sits on the computation-communications interface.

Simulation

- *The simulation of time sharing systems* (W2113442017): explicit simulation of computer-system performance.
- *Experimental and Numerical Simulation of Two-Phase Flow with Interphase Mass Transfer in One and Two Dimensions* (W2149866374): explicit numerical simulation paper centered on finite-difference solution of a physical system.

Notes: These examples come from manual full-text packet reads and are used only to anchor interpretation. The empirical examples show how computers entered measured and observational work through diagnosis, spectral analysis, monitoring, and numerical processing.

B.5 Additional Keyword-Family Robustness

As a robustness check, we also estimate simple Callaway-Sant'Anna pre/post DiD specifications for the family outcomes at the university-year level. We do *not* use these family DiD estimates as the main interpretation device, because the families are selective rather than exhaustive. Still, Figure B.5 shows that the same broad content themes appear when the family outcomes are estimated directly as simple post-treatment contrasts.

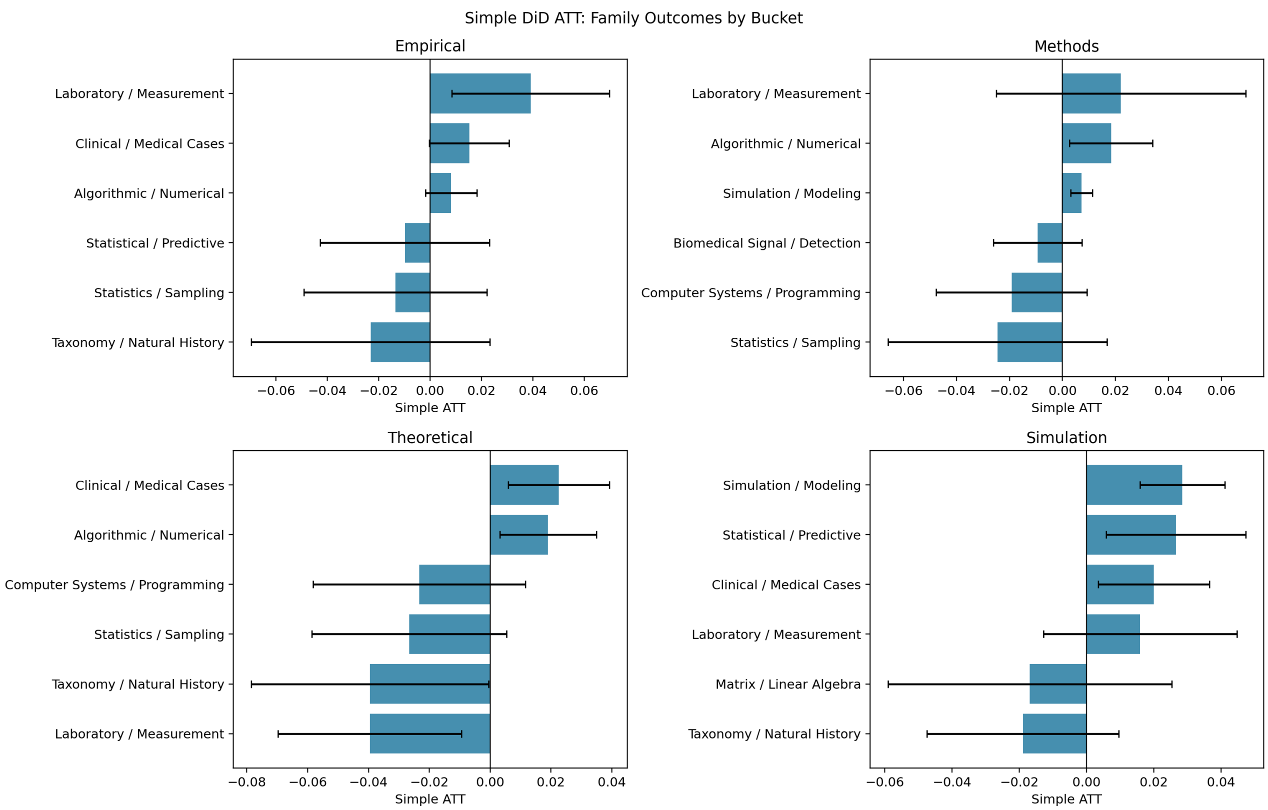


Figure B.5: Simple Callaway-Sant'Anna pre/post ATT estimates for each full-text keyword-family outcome, estimated at the university-year level across the 184 treated universities. The outcome for each family is the count (or share) of papers containing at least one family keyword in their full text. Markers are aggregated cohort ATTs; horizontal bars are 95% CIs. These are robustness objects and should not be read as the primary evidence, since the families are selective interpretive probes rather than an exhaustive vocabulary partition; see Appendix Table 10 for family definitions.

C Taxonomy of Computer Uses on Research

To understand in more detail how researchers were using computers, we rely on LLMs to classify and cluster computer usage by type.

To begin, we use a two-step procedure to screen which papers were using computers, described in Section 3.2.1. In summary, we first screen papers for a list of keywords generating high recall of computer usage, and then have an LLM (OpenAI's o3-mini) classify the papers into the four broad categories of computer mention or usage or no usage at all.

To benchmark the coarse computer-use classifier, we compare model predictions against a human-labeled validation sample of 184 papers. The headline benchmark in the paper collapses all computer-use and computer-mention categories into a single positive class against *No Computer Use or Mention*. On the benchmark papers with both human labels and model predictions, this binary comparison yields an F1 score of 0.967, with precision 0.956 and recall 0.978. Evaluating the full five-category taxonomy yields a weighted F1 of 0.866 and a macro F1 of 0.762.[43] The remaining errors are concentrated mainly in distinguishing papers that merely mention computers from papers that use computers as tools or treat computers as objects of study.

The four categories of computer usage we are the ones we describe in Section 4.3: (i) *Computer as a Tool*; (ii) *Computer as an Object of Study*; (iii) *Hardware or Software Papers*; and (iv) *Other Mentions*. Although these categories are not strictly mutually exclusive, the LLM is instructed to select the most salient role. As shown in Figure C.6, among papers classified as computer use or mention, the majority (62.7%) employ computers primarily as tools for scientific work, such as solving equations or running simulations. While informative, this four-way taxonomy is intentionally coarse and serves only as a first pass.

Before we move to the next step, we further restrict our sample adding a third step to classify article type within papers that use computers from another round of LLM classifications. In this step, we narrow our sample only to research articles – discarding for example

[43]Weighted F1 averages class-specific F1 scores using class frequencies as weights, whereas macro F1 gives each class equal weight regardless of how common it is.

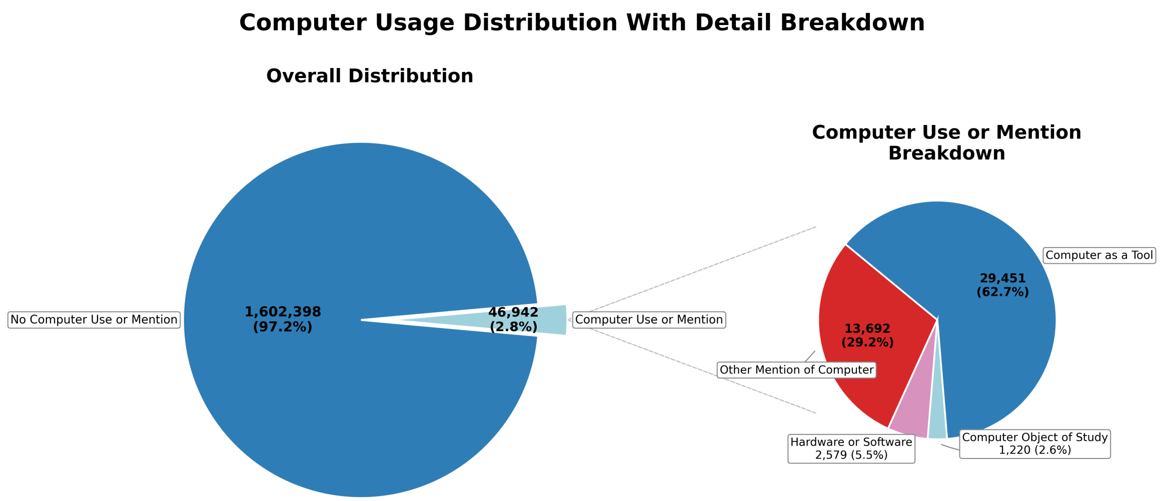


Figure C.6: Computer Usage Intensity in the Local Full-Text Pooled-Paper Sample. $N = 1{,}649{,}340$. We use o4-mini to classify papers from local full text after a keyword first pass; the figure excludes papers classified as "Unclear" or "Text Truncated or Mangled."

conference papers, book reviews, news articles, and other paratext not properly classified as such by OpenAlex. Intersecting these clustered classifications with the distinct pooled-paper universe used in the manuscript yields a final sample of $N = 30{,}959$ research papers.

We then follow the procedure proposed by Tamkin et al. (2024). Specifically, we first use Gemini 2.5 Flash to generate a short description of how each paper is using computers, based on its full text. We then embed these descriptions using a dense embedding model (`gemini-embedding-001`, with 1536 dimensions). Next, we run K-Means clustering on the embeddings, setting $K = 61$.[44] For each cluster, we feed the 20 papers closest to the cluster centroid plus 80 random paragraph descriptions of papers within the cluster, along with 50 descriptions of the closest neighbors outside the cluster, to Gemini 2.5 Pro in order to generate a title and description of each cluster. Finally, we manually assign each of the 60 lower-level clusters to 11 high-level clusters (see below).

Figure C.7 pools computer usage between 1950–1970 within this clustered research-article sample. In aggregate, three categories – Numerical Computation and Mathematical Modeling (22.2%), System Simulation & Dynamic Modeling (21.3%), and Statistical Analysis & Data Processing (19.3%) – account for more than 60% of uses or mentions of computers.

[44]We choose the number of clusters using a Kneedle algorithm.

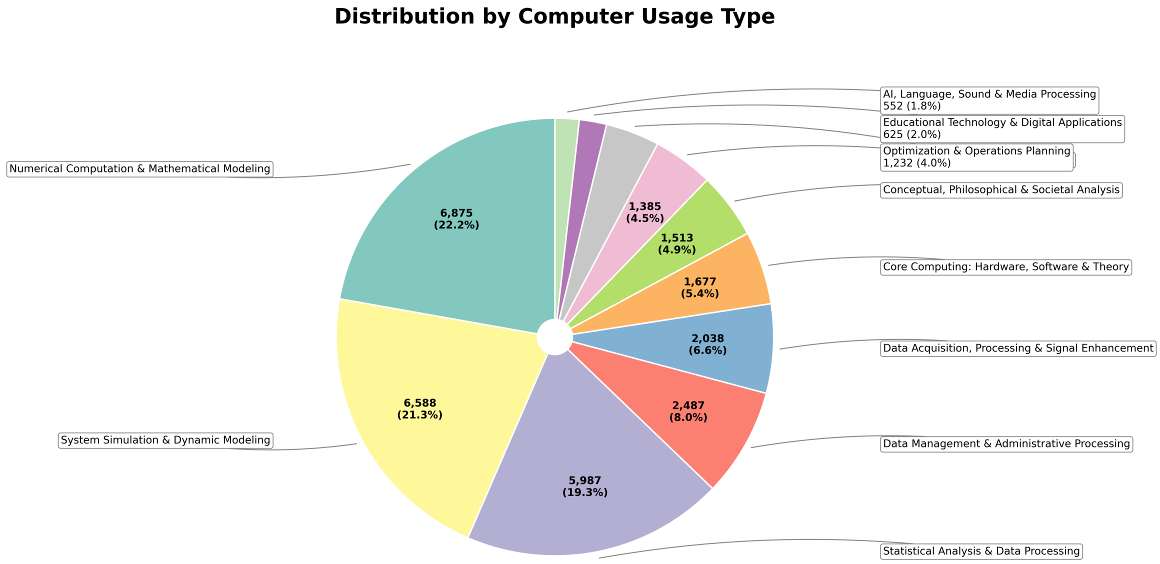


Figure C.7: Pooled distribution of computer-usage types across the clustered research-article sample, 1950–1970 ($N = 30{,}959$). Usage categories are generated by a two-step procedure: we extract free-text LLM descriptions of how computers are used in each paper, embed the descriptions, and cluster the embeddings (see Section 4.3 and Appendix C). Each slice is the share of the sample assigned to a given cluster; category labels are post-hoc human summaries of the clusters.

In Figure C.8, we plot the evolution we showed in Figure 3 broken down by domain. The graph shows numerical and modeling uses dominate uses in the Physical Sciences, whereas on other domains Statistical Analysis & Data Processing dominates and increase clearly over time.

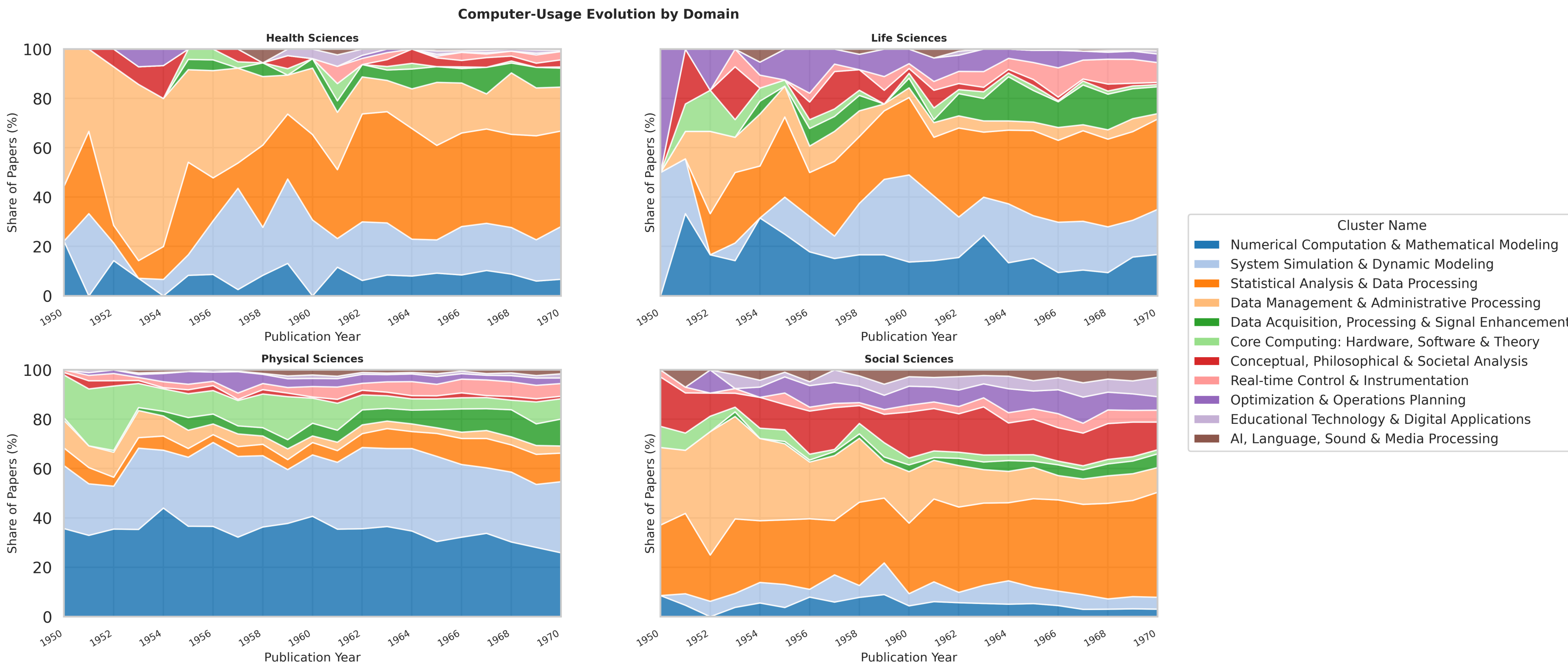


Figure C.8: Evolution of computer-usage category shares over time by OpenAlex domain (Physical, Life, Health, Social Sciences), 1950–1970, clustered research-article sample ($N = 30{,}959$). Categories are defined as in Figure C.7. Each panel shows yearly shares within the domain-specific subset, normalized to 100%. Physical Sciences exhibit the earliest and broadest transition toward computational categories; Life and Health Sciences adopt later and remain more empirical.

1. Numerical Computation & Mathematical Modeling

- 0: Solve quantum mechanical equations
- 1: Solve equations for stress and stability
- 8: Refine crystal structures
- 16: Numerically evaluate mathematical functions
- 19: Implement algorithms for matrix algebra
- 22: Develop numerical methods for differential equations
- 26: Solve theoretical equations for nuclear structure
- 41: Numerical analysis for statistical methods
- 49: Verify conjectures in number theory
- 60: Analyze biochemical kinetics and macromolecular properties

2. Statistical Analysis & Data Processing

- 5: Calculate radiation dose
- 12: Statistical analysis for medical studies
- 13: Process and analyze earth-surface data
- 15: Statistical analysis for agricultural experiments
- 17: Analyze research data with statistical software
- 25: Automate scoring and analysis of educational data
- 33: Econometric modeling
- 34: Analyze biomedical data
- 39: Factor analysis on psychological data
- 42: Statistical analysis for archaeology
- 43: Statistical analysis for ecology
- 44: Statistical analysis of survey data
- 45: Statistical analysis for psychology

3. System Simulation & Dynamic Modeling

- 14: Model evolutionary dynamics
- 27: Analog and hybrid simulation
- 29: Model atmospheric dynamics
- 31: Model chemical kinetics and thermodynamics
- 32: Numerical modeling of geophysical systems
- 36: Numerically model physical phenomena
- 40: Model flow for water resources
- 46: Simulate complex systems
- 48: Numerical modeling of plasma
- 53: Numerically model fluid systems
- 54: Simulate physiological systems

4. Optimization & Operations Planning

- 2: Implement optimization algorithms
- 28: Model financial instruments
- 51: Analysis and planning in agriculture

5. Data Acquisition, Processing & Signal Enhancement

- 3: Model and visualize geographic systems
- 6: Process data from analytical instruments
- 9: Process astronomical data and model stellar systems
- 18: Perform signal averaging on physiological recordings
- 38: Analyze particle-detector data with Monte Carlo simulation

6. Data Management & Administrative Processing

- 7: Automate business data processing
- 11: Punch-card medical record tabulation
- 21: Automate library operations
- 52: Process administrative data with punch-cards

7. Real-time Control & Instrumentation

- 10: Automate real-time control for experiments
- 24: Simulate and control space systems
- 59: Integrated systems for simulation, control, and analysis

8. Core Computing: Hardware, Software & Theory

- 4: Foundational programming and system software
- 50: Theoretical foundations of computation and logic
- 56: Design and analyze early computer hardware
- 58: Algorithms for computational statistics and data management

9. AI, Language, Sound & Media Processing

- 20: Computational linguistics and machine translation
- 23: Analyze and generate musical sound

10. Conceptual, Philosophical & Societal Analysis

- 30: Computers referenced incidentally or metaphorically
- 35: Conceptualizing intelligence and cybernetic systems
- 47: Computerization of gov't and soc policy

11. Educational Technology & Digital Applications

- 37: Computer-assisted instruction
- 57: Applying digital technologies in education

D Citation Functions of Computer Papers

This appendix provides additional detail for Section 4.6. The workflow adapts the clustering approach in Appendix C, but applies it to citation contexts rather than to descriptions of how computers are used within papers.

D.1 Pipeline and Sample Construction

We begin from the full 1940–1980 OpenAlex paper universe used elsewhere in the project. The computer-paper group is the set of computer-keyword flagged papers, and the control pool is made up of papers that do not pass that keyword screen. For each computer paper, we search for non-computer controls in the same OpenAlex domain and field and within a ± 1 publication-year window. If that eligible pool is larger than 200 papers, we evaluate a random subset of 200 candidates and choose the nearest neighbor in abstract-embedding space, using cosine distance. Matching is therefore exact on domain and field, tight on year, one-to-one, and with replacement on the control side. We then keep citing papers for which open-content full text can be recovered from OpenAlex assets, recover inline citation contexts using GROBID-style tags or string and pattern matches, and ask an LLM to write short descriptions of why the cited paper is being used in the citing paper. We embed those descriptions, cluster them, and aggregate the resulting clusters into five high-level citation-function buckets. The final analysis sample contains 98,051 citing-cited pairs with usable context descriptions. Of those, 20,910 fall in the strict clustered sample used below: 16,114 links to computer papers and 4,796 to matched controls. These links map into a cited-paper panel of 16,120 papers, of which 12,213 are computer papers and 3,907 are matched controls.

D.2 Regression Specification and Bucket Scores

The bucket-premium scores are estimated at the cited-paper level. Let n_{pk} be the number of classified incoming citations to paper p that fall in citation-function bucket k, let $N_p = \sum_{k=0}^{4} n_{pk}$, and define $s_{pk} = n_{pk}/N_p$. Omitting bucket 0 (*Technical evidence and protocols*) in estimation and re-centering the fitted coefficients so that the five reported bucket scores average to zero, we estimate

$$FWCI_p = \alpha + X_p'\gamma + \lambda_{\text{year}(p)} + \delta_{\text{domain}(p)} + \sum_{k=1}^{4} \theta_k s_{pk} + u_p,$$

where X_p includes number of authors and NSF grants. Figure D.9 reports these re-centered coefficients with and without a control for $\log(1+\text{classified citation count})$. Canonical sources remain the highest-premium bucket and established findings the lowest, even after partialling out citation volume.

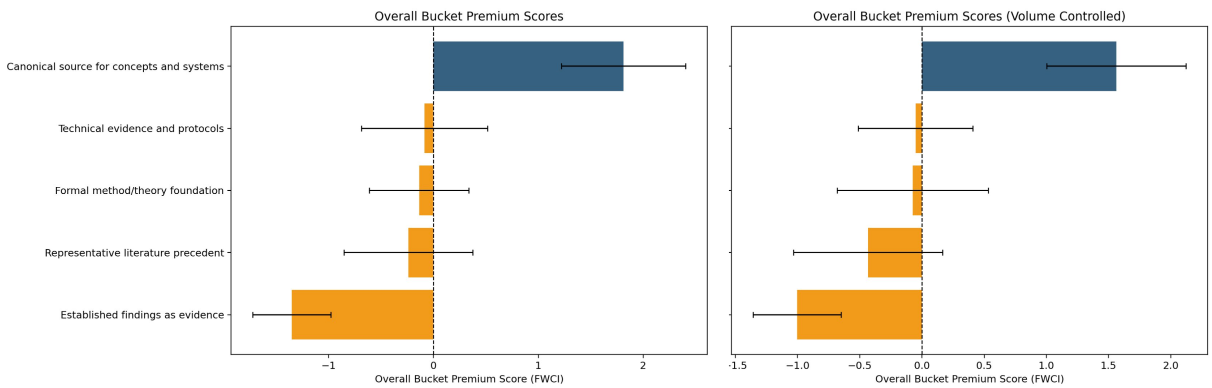


Figure D.9: Citation-function bucket premium scores. Cited-paper regression of FWCI on citation-function shares, controlling for number of authors, NSF grants, cited-paper year fixed effects, and domain fixed effects. Sample: 15,938 cited papers with at least one classified citation. The left panel omits citation volume, so coefficients summarize each bucket's overall association with citation impact; the right panel adds $\log(1 + \text{classified citation count})$ to partial out citation volume. Scores are re-centered so all five buckets are directly comparable; whiskers show 95% CIs.

D.3 Decomposition of the Citation Premium

The descriptive decomposition in Table 12 starts from the same cited-paper panel and adds two objects built from the bucket shares $\{s_{pk}\}_{k=0}^{4}$: composition and excess breadth. We measure the raw breadth of paper p's citation uses by the effective number of citation-function buckets,

$$B_p = \exp\left(-\sum_{k=0}^{4} s_{pk} \log s_{pk}\right),$$

which equals one when all classified citations fall in a single bucket and rises as those citations are spread more evenly across buckets. Because B_p mechanically rises with citation volume, we first residualize it using

$$B_p = a + \sum_b \rho_b \mathbf{1}\{\text{classified citation count bin} = b\} + X_p'\gamma + \lambda_{\text{year}(p)} + \delta_{\text{domain}(p)} + \nu_p,$$

where the citation-count bins are 1, 2, 3, 4, and 5+ classified citations. Excess breadth is the residual $\widehat{\nu}_p = B_p - \widehat{B}_p$. We then estimate the cited-paper regressions

$$FWCI_p = \alpha + \beta T_p + X_p'\gamma + \lambda_{\text{year}(p)} + \delta_{\text{domain}(p)} + \varepsilon_p,$$

$$FWCI_p = \alpha + \beta^{EB} T_p + \eta\widehat{\nu}_p + X_p'\gamma + \lambda_{\text{year}(p)} + \delta_{\text{domain}(p)} + \varepsilon_p,$$

$$FWCI_p = \alpha + \beta^{C} T_p + \sum_{k=1}^{4} \theta_k s_{pk} + X_p'\gamma + \lambda_{\text{year}(p)} + \delta_{\text{domain}(p)} + \varepsilon_p,$$

and

$$FWCI_p = \alpha + \beta^{EBC} T_p + \eta\widehat{\nu}_p + \sum_{k=1}^{4} \theta_k s_{pk} + X_p'\gamma + \lambda_{\text{year}(p)} + \delta_{\text{domain}(p)} + \varepsilon_p,$$

where T_p is an indicator for computer papers and bucket 0 is omitted because the shares sum to one. Comparing the baseline and augmented treated coefficients implies that breadth and composition together explain about 10% of the computer-paper FWCI premium. A Shapley decomposition of that explained portion assigns roughly 90% to composition and 10% to excess breadth, reinforcing that the mix of citation reasons matters more than breadth alone.

D.4 Illustrative Citation Functions

Table 13 summarizes the substantive distinction between the five citation-function buckets. These examples are condensed from the full cluster example brief generated from centroid-nearest papers, nearby outside examples, and matched computer-versus-control pairs.

Table 12: Citation Premium Decomposition

Spec	Treat. coef.	Atten.	p-value
Baseline	1.193	–	0.000
+ Excess breadth	1.179	1.2%	0.000
+ Composition	1.086	9.0%	0.000
+ Breadth + composition	1.077	9.8%	0.000
+ Volume (robustness)	0.773	35.2%	0.000
+ Volume + breadth + composition	0.657	45.0%	0.001

Notes: Unit of observation is the cited paper. Baseline specification regresses FWCI on a computer-paper indicator, number of authors, NSF grants, cited-paper year fixed effects, and domain fixed effects. Let s_{pk} denote the share of paper p's classified incoming citations in citation-function bucket k. The raw breadth object is the effective number of buckets, $B_p = \exp(-\sum_k s_{pk} \log s_{pk})$. Excess breadth is the residual from regressing B_p on flexible classified-citation-count bins $(1, 2, 3, 4, 5+)$, the baseline controls, and the same fixed effects. Composition refers to the bucket shares $\{s_{pk}\}_{k=0}^{4}$, with *Technical evidence and protocols* omitted as the reference bucket in estimation because the shares sum to one. The attenuation from 1.193 to 1.077 implies that citation breadth and composition account for about 10% of the computer-paper FWCI premium. This is a descriptive accounting exercise rather than a causal mediation result.

Table 13: Illustrative Citation-Function Buckets

Bucket	What the citation is doing	Illustrative example
Technical evidence and protocols	Supplies a protocol, benchmark, calibration, or measurement input.	"As previously described" citations to earlier activation or purification procedures, or to benchmark values used to calibrate the current analysis.
Established findings as evidence	Provides factual support for a substantive claim.	Waterman and Horch (1966) cited as evidence on polarization sensitivity, or Daley et al. (1971) cited for measured radar backscatter results.
Representative literature precedent	Shows that a phenomenon, debate, or method already exists in the literature.	Siegel (1974) or Posner, Nissen, and Klein (1976) cited as one example in a broader list of prior work.
Formal method or theory foundation	Gives the estimator, derivation, or analytical framework the paper actually uses.	Eisenthal and Cornish-Bowden (1974) or Wilkinson (1961) cited for the enzyme-kinetics procedure used in estimation.
Canonical source for concepts and systems	Establishes the recognized origin of a concept, language, system, or benchmark.	Hoare on monitors or the TAXIS data model cited for provenance rather than as direct empirical evidence.

Notes: Examples are shortened descriptions of representative citation contexts from the classified citation-link sample. They are intended to clarify the meaning of each bucket rather than to estimate any additional coefficients.

E Additional Descriptives

E.1 Computer Database

In this subsection, we present additional descriptive facts on the computer installations dataset.

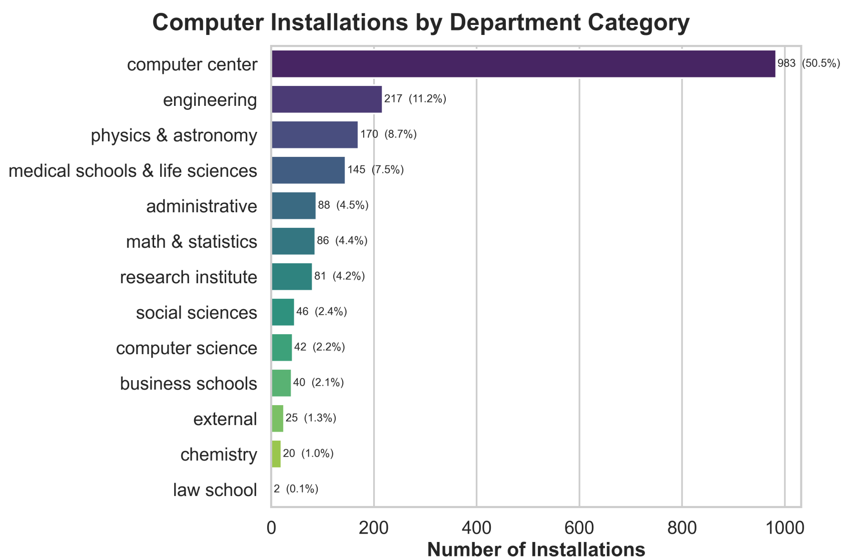
IBM dominated the academic computing market with 58% of all installations, followed by Digital Equipment Corporation (DEC) at 9%. Before 1955, most universities with computers (12 out of 16, or 75%) built their own machines, usually based on the Princeton IAS design. As commercial options became more available and affordable, in-house development declined. In total, 27 universities built 44 computers internally. Finally, analog computers represent 8% of total installations.[45]

We are able to determine the location of computer installation within university for around 90% of installations in our sample. To parse raw installation strings into consistent categories, we use an LLM to create a mapping between unique raw strings and location categories and then manually validate the mapping. In Figure E.10a, we show the distribution of installation by location type.

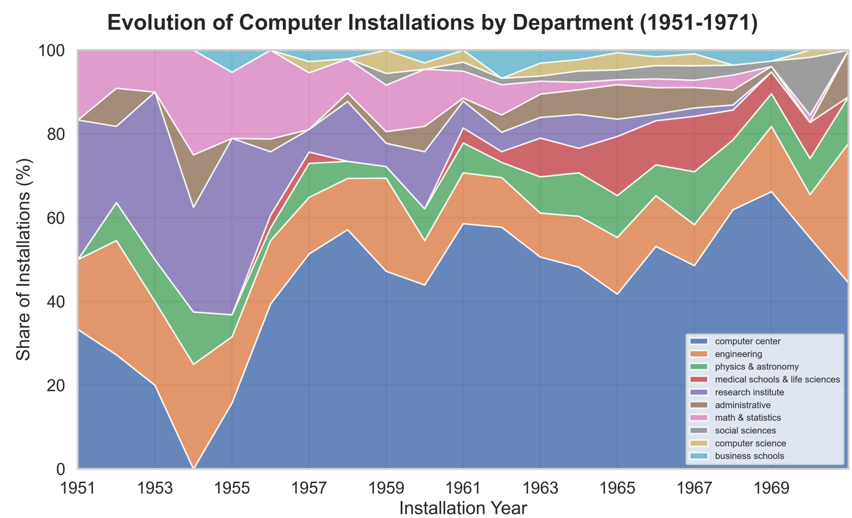
In Figure E.10b, we display the location categories of computer installations over time, using the same data from Figure E.10a.

Figures show that computer installations were heavily concentrated at shared university computer centers (50.5%), followed by engineering departments (11.2%) and Physics & Astronomy departments (8.7%). Over time, we see installations shift from Math & Stats departments and research institutes like Princeton's IAS and University of Michigan's Willow Run laboratories, towards computer centers.

[45]Not all surveys report analog computer installations, so this number should be interpreted as a lower bound.



(a) Number of installations by department category



(b) Evolution of installations by department category over time

Figure E.10: Department/unit categories of computer installations in the 1950–1971 US university sample. Panel (a): bars show the number of installations assigned to each category; shares computed over installations with an identifiable department (omitting the $\approx 10\%$ of records lacking location information). Panel (b): yearly installations by category over 1950–1970, normalized to 100% within each year so colors sum to the year's total. Categories include computing centers, engineering departments, physical-science departments, business/administration, and other academic units.

E.2 Computer Usage Across Disciplines and Diffusion

This appendix presents robustness checks and further results for Section 4.2.

To analyze the speed of computer use diffusion, we plot the intensity of computer usage in 1950, 1965 and 1980 across domains in Figure E.11. The results show that Life and Health Sciences begin to catch up in terms of computer usage later in our period, though not enough to overcome the initial lag.

In Figure E.12, we show that qualitative results of Figure 2 do not change if we use exclude "computer" as a standalone keyword or whether we use LLMs to classify computer use or mention from full text of papers instead.

In Figure E.13, we plot the distributions by field within each domain, using keyword matches including "computer".[46] In relative terms, Engineering, Physics & Astronomy, and Computer Science, are the more computer intensive field within the Physical Sciences. Within the Social Sciences, Decision Sciences and Psychology lead the pack.

Zooming in selected subjects (Figure E.14), we see some particular subfields are very computer-intensive: Numerical Analysis and Statistics & Probability within Mathematics; Aerospace Engineering within Engineering; Nuclear & High Energy Physics within Physics & Astronomy; Management & Operations Research within Decision Sciences.

E.3 Paper Characteristics

This subsection shows the same regression specifications of Section 4.5 on other outcomes. It also reports a robustness check that additionally partials out semantic embeddings from paper titles and abstracts together with SPECTER embeddings that capture citation-neighborhood position.

Table 14 revisits the citation outcomes from Table 1 under this richer control set. The estimated citation premia remain positive and statistically significant across all reported outcomes, including top-tail citation status and citations accrued within five years of publication.

[46] For conciseness, we do not plot results excluding computers or using LLM classifications, but results are close and qualitatively unchanged.

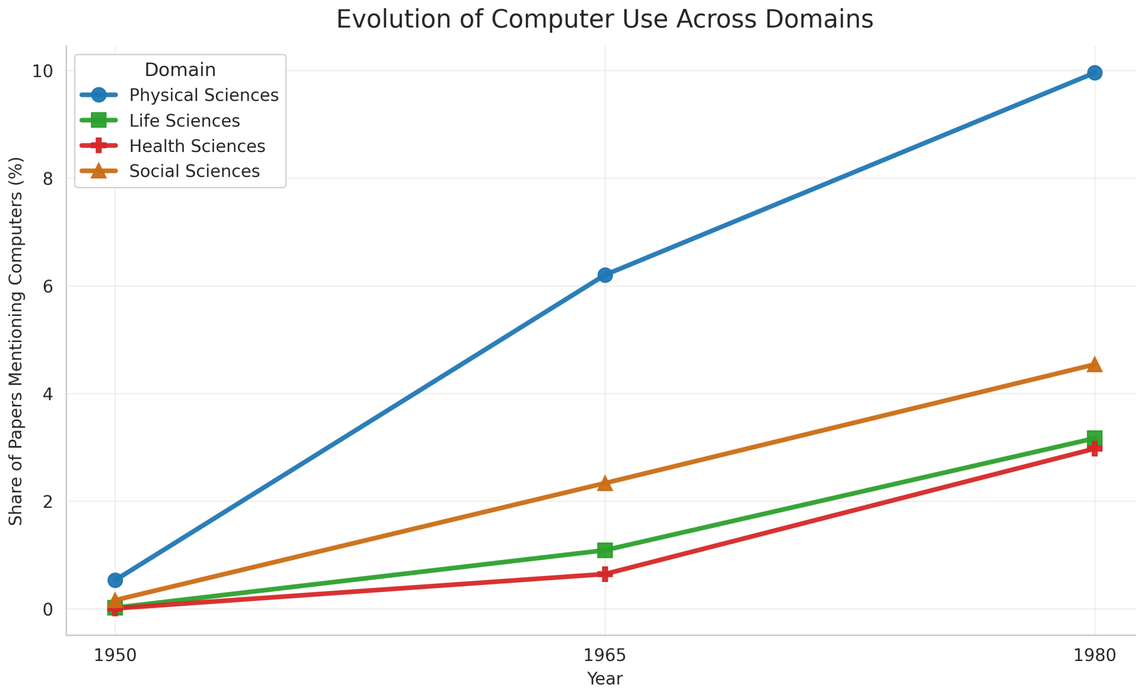
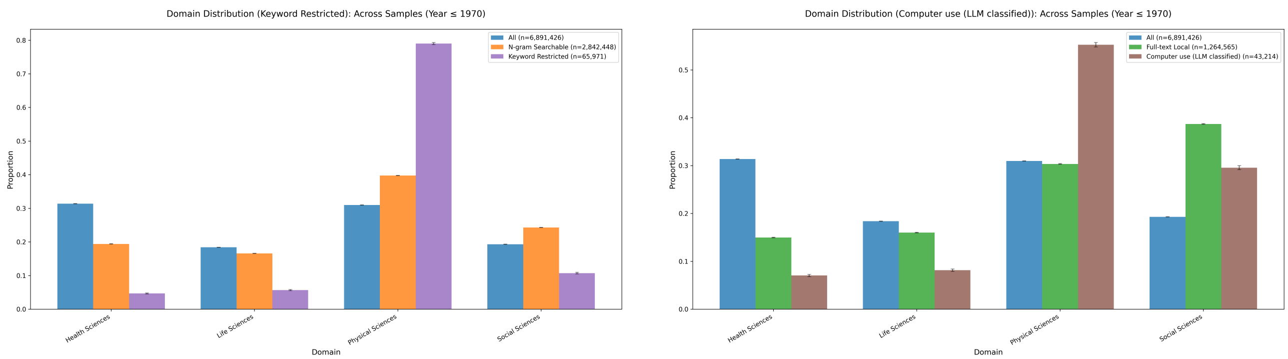


Figure E.11: Slope graph of the share of papers mentioning computer-related keywords across OpenAlex domains, comparing 1940–1944 to 1960–1964. Each line tracks one domain's share across the two windows; steeper slope indicates a larger diffusion of computer use. Plot includes "computer" itself as a keyword (unlike the restricted-keyword variant in Figure E.12, which excludes it). Computed over the searchable full-text sample (≈40% of OpenAlex papers in the period).

Table 14: Citation Outcomes with Embedding Controls

	Log cites	Top 10%	Top 1%	FWCI	C5	Cite pct.
Computer-Keyword Flag	0.196***	0.034***	0.005***	0.337***	0.530***	0.031***
	(0.005)	(0.001)	(0.001)	(0.019)	(0.042)	(0.001)
Observations	1,882,763	1,882,654	1,882,654	1,882,004	1,882,763	1,882,654
R^2	0.002	0.000	0.000	0.000	0.000	0.001
Mean of Dep Var	-0.016	0.046	0.001	0.171	0.644	0.003

Notes: Columns report second-stage OLS coefficients from a double-machine-learning robustness specification. In a first stage, both the computer-keyword indicator and each outcome are residualized with respect to author, publication year, university, and primary-topic fixed effects; controls for the number of authors and NSF grants awarded to the paper; dense title and abstract embeddings; and SPECTER paper embeddings. Residuals are estimated with 3-fold cross-fitting and SGD, after which the residualized outcome is regressed on the residualized treatment. Paper-author observations are weighted by the inverse number of authors, and second-stage standard errors are clustered at the paper level (OpenAlex Work ID). Sample restricted to the 1947–1975 searchable-full-text analysis sample used in the paper-level regressions. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.



(a) Keywords restricted

(b) LLM classified

Figure E.12: Robustness variants of the domain-level distribution of computer papers shown in Figure 2. Panel (a): restricted keyword set (excludes the word "computer" itself). Panel (b): LLM-classified computer papers from the local full-text sample. Each panel compares three slices: (i) all searchable papers, (ii) papers flagged as computer papers by the given method, (iii) baseline share of each domain in searchable OpenAlex. Bars of the same color (slice) add to 1 across domains within each panel.

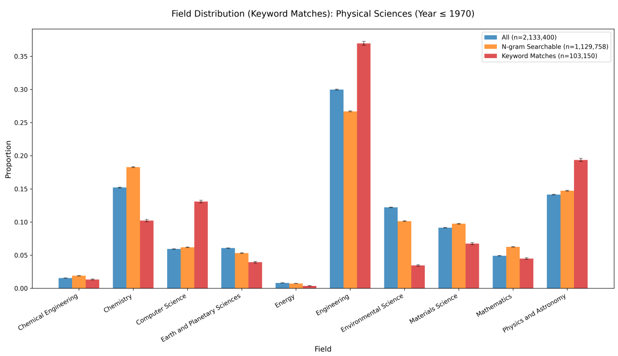
Table 15 shows that computer papers exhibit greater conceptual breadth: they use roughly 2.5% more OpenAlex concepts and have a lower concept Herfindahl Index (HHI), indicating their focus is spread more evenly across concepts rather than concentrated in just a few.[47]

Results for topics are similar but smaller: computer papers use about 1% more topics, have lower topic and concept HHI, and display higher topic- and concept-pair novelty. These breadth effects remain modest, ranging from about 1% to 4% of the mean depending on the outcome.

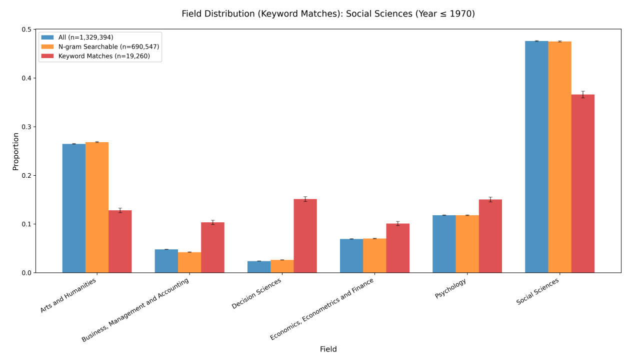
Table 16 shows regressions on outcomes used in the Science of Science literature, and obtained from merging OpenAlex data with SciSciNet (Lin et al., 2023).

The first two outcomes are an atypicality measure introduced by Uzzi et al. (2013), capturing how rarely the paper's cited journal pairs have been co-cited before – that is, the more negative the z-score, the more unconventional the reference combination. The results show that computer-intensive papers' distribution of journal-pairs is more atypical than others at both the left-tail and median cutoffs.

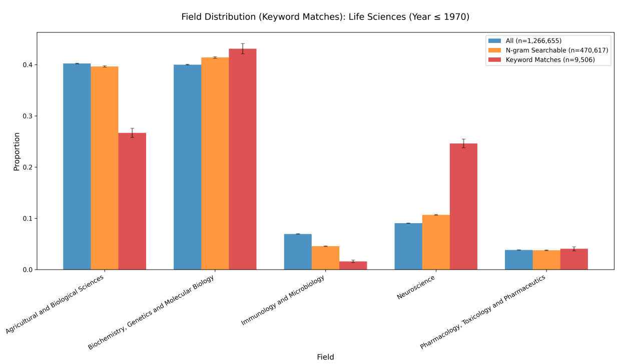
[47]OpenAlex *concepts* are hierarchical subject tags – about 65 000 in total – automatically assigned to each work by a machine-learning classifier. The classifier examines the work's title, abstract, and venue; each concept carries a `score` measuring relevance, and concepts with score ≥ 0.3 (plus all of their ancestors) are linked to the work. See https://docs.openalex.org/api-entities/concepts for full details.



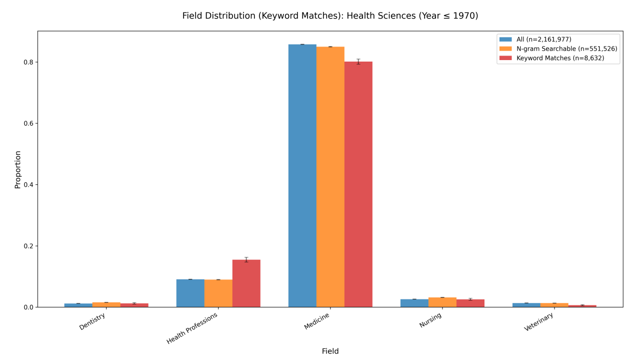
(a) Physical Sciences



(b) Social Sciences



(c) Life Sciences

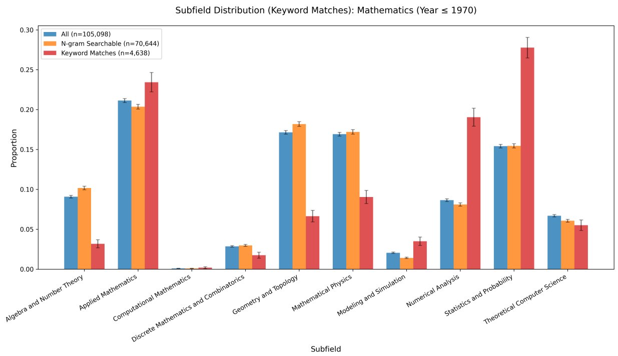


(d) Health Sciences

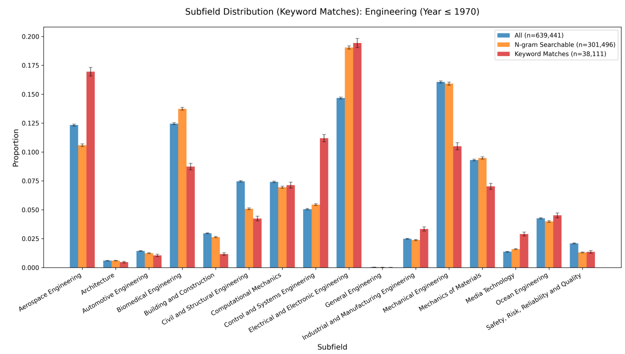
Figure E.13: Distribution of computer papers within each OpenAlex major field (Physical, Social, Life, Health Sciences). Within each domain panel, bars compare the field-level composition across three slices: (a) all papers, (b) searchable full-text papers, (c) papers flagged for computer keywords. Within-panel bars of the same color add to 1, so these are within-domain compositional shares and not cross-domain comparisons. Plots include "computer" as a keyword; restricted-keyword variants appear in Figure E.12.

The third outcome is the disruption score introduced by Park et al. (2023), which captures how often later papers cite the focal work without also citing its references; positive values mean the paper is disruptive in the sense of become a new focal reference for subsequent work, whereas negative values indicate it mainly builds on prior work. The regressions show a clear null, suggesting that though computer papers are more cited, broader, and using more novel reference combinations, they do not seem to be more disruptive.

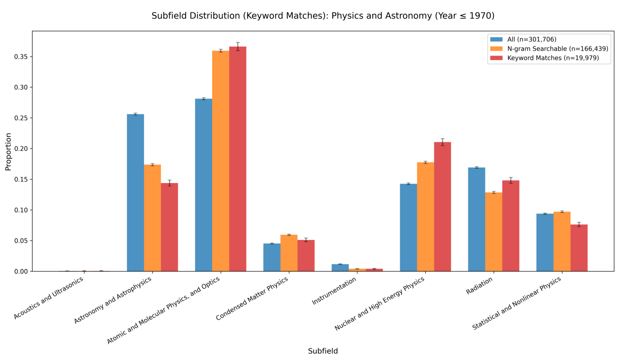
The last two outcomes, the Sleeping Beauty score and the Awakening score, were introduced by Ke et al. (2015). The Sleeping Beauty score B is calculated by integrating, year by year from publication to the citation peak, the positive gap between the paper's observed yearly citations c_t and a reference line that joins the point (t_0, c_0) (publication year and its



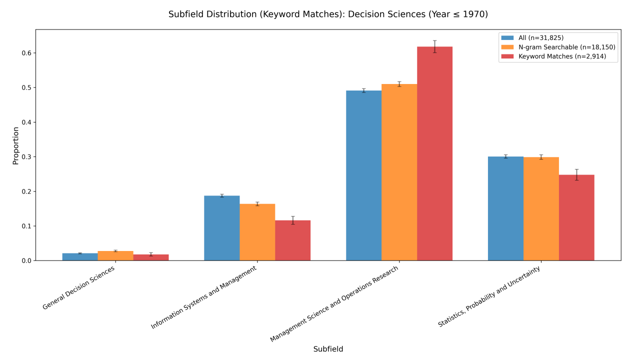
(a) Mathematics



(b) Engineering



(c) Physics & Astronomy



(d) Decision Sciences

Figure E.14: Distribution of computer papers across subfields of four selected fields: Mathematics, Engineering, Physics & Astronomy, and Decision Sciences. Within each field panel, bars compare subfield-level composition across three slices: (a) all papers, (b) searchable full-text papers, (c) papers flagged for computer keywords. Within-panel bars of the same color add to 1, so values are within-field compositional shares. Plots include "computer" as a keyword. Fields selected for visibly sharp pre-digital numerical-intensity gradients; full set of fields appears in the underlying panel data.

citations) to $(t_{\max}, c_{\max})$ (peak-year citations); each annual gap is divided by c_t before summation. The result is the area of under-performance relative to a linear growth baseline, so larger B indicates a longer or deeper dormant period before recognition. The Awakening score A (reported as t_a in the original paper) is the publication age at which that gap reaches its maximum; it identifies the exact year when the citation trajectory turns upward most sharply, signaling the start of the paper's rapid ascent. These results suggest that computer-intensive papers are more likely to be sleeping beauties and awaken slightly later. This result echoes a higher citation count after 10 years than after 5 years.

Table 17 shows regressions on patent citations, and other outcomes. Its main result is that papers mentioning computers are not more likely to be cited by patents.

Table 15: Effect of Computer-Keyword Flag on Topic and Concept Outcomes

	Topics			Concepts		
	# Tpks	Tpk HHI	Pair Avg Nov	# Cncpts	Cncpt HHI	Pair Avg Nov
Computer-Keyword Flag	0.022***	-0.004***	0.064***	0.086***	-0.007***	0.203***
	(0.002)	(0.001)	(0.014)	(0.008)	(0.000)	(0.016)
Number of Authors	0.023***	-0.005***	0.097***	0.072***	-0.002***	0.060***
	(0.001)	(0.000)	(0.004)	(0.002)	(0.000)	(0.004)
NSF Grants (paper)	0.108***	-0.028***	0.429***	0.565***	-0.016***	0.448***
	(0.015)	(0.004)	(0.081)	(0.062)	(0.002)	(0.087)
Observations	3,903,691	3,903,691	3,903,691	3,903,691	3,903,691	3,903,691
R^2	0.593	0.519	0.506	0.510	0.507	0.462
Mean of Dep Var	2.095	0.491	7.392	3.485	0.197	9.168
Author/Year/Univ FE	Yes	Yes	Yes	Yes	Yes	Yes
Subject FE	Topic	Topic	Topic	Topic	Topic	Topic

Notes: Paper-author observations, weighted by the inverse number of authors. Treatment is an indicator for the presence of computer-related keywords in the paper's full text. Controls are the number of authors and the number of NSF grants awarded to the paper. Fixed effects are author, publication year, university, and primary topic. Standard errors are clustered at the paper level (OpenAlex Work ID). Sample restricted to papers with at least one in-sample university affiliation and searchable full text in OpenAlex. HHI variables are computed at the paper-level topic/concept score distribution. For novelty scores, we assign a frequency score to every concept/topic pair in the paper computed over all pairs up to the paper publication year, then compute cross-entropy over those scores for each paper.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 16: Science of Science Outcomes

	Atp-Z (10%)	Atp-Z (Med)	Disrupt	SB	Awak
Computer-Keyword Flag	-10.175***	-2.895***	0.000	1.432***	0.308***
	(0.783)	(1.082)	(0.000)	(0.531)	(0.047)
# Authors	-1.134***	-1.176***	0.000***	0.119	0.065***
	(0.141)	(0.193)	(0.000)	(0.120)	(0.010)
NSF grants (paper)	-2.925	-2.278	-0.006***	-1.146	-0.140
	(2.480)	(3.385)	(0.001)	(1.282)	(0.339)
Observations	2,434,402	2,428,263	3,903,691	3,353,525	3,353,525
R^2	0.479	0.503	0.377	0.282	0.404
Mean of Dep Var	48.063	105.177	0.012	19.686	6.856
Author/Year/Univ FE	Yes	Yes	Yes	Yes	Yes
Subject FE	Topic	Topic	Topic	Topic	Topic

Notes: Paper-author observations, weighted by the inverse number of authors. Treatment is an indicator for the presence of computer-related keywords in the paper's full text. Controls are the number of authors and the number of NSF grants awarded to the paper. Fixed effects are author, publication year, university, and primary topic. Standard errors are clustered at the paper level (OpenAlex Work ID). Sample restricted to papers with at least one in-sample university affiliation and searchable full text in OpenAlex. First two columns refer to atypicality scores (Uzzi et al., 2013) at the 10th percentile and median cut-off. Disruption scores are measured as in Park et al. (2023). Sleeping Beauty (SB) and Awakening (Awak) follow the methodology from Ke et al. (2015). Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 17: Other Outcomes

	# Patent Cites	# Countries	# Institutions	# Refs
Computer-Keyword Flag	0.244	0.003***	0.013***	1.967***
	(0.197)	(0.001)	(0.002)	(0.042)
# Authors	0.013	0.039***	0.140***	0.080***
	(0.012)	(0.001)	(0.002)	(0.010)
NSF grants (paper)	-0.084	-0.006	0.009	4.132***
	(0.154)	(0.004)	(0.013)	(0.297)
Observations	3,903,691	3,903,691	3,903,691	3,903,691
R^2	0.289	0.490	0.537	0.454
Mean of Dep Var	0.259	0.605	0.643	5.492
Author/Year/Univ FE	Yes	Yes	Yes	Yes
Subject FE	Topic	Topic	Topic	Topic

Notes: Paper-author observations, weighted by the inverse number of authors. Treatment is an indicator for the presence of computer-related keywords in the paper's full text. Controls are the number of authors and the number of NSF grants awarded to the paper. Fixed effects are author, publication year, university, and primary topic. Standard errors are clustered at the paper level (OpenAlex Work ID). Sample restricted to papers with at least one in-sample university affiliation and searchable full text in OpenAlex. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In Table 18, we relate the initial task-based classification from Section 4.3 to downstream outcomes in citations and concept-based novelty. The refreshed backfilled specification continues to show that the positive citation premium is concentrated among papers classified as using computers *as tools*. These tool papers also use more concepts and exhibit higher concept-level novelty. By contrast, papers categorized as *Objects of Study*, *Hardware or Software*, or *Other Mentions* do not display comparable positive premia relative to baseline articles, and hardware/software and other-mention papers often carry negative coefficients. This suggests that the recognized impact of computers in the scientific literature was driven primarily by their instrumental role in enabling new research, rather than by work on computing systems themselves.

E.4 Author Characteristics

This subsection shows other results that further suggest authors using computers have greater impact and are positively selected among their counterparts.

Table 18: Computer-Usage LLM Classification

	Log Cites	Top 1%	# Cncpts	Cncpt Max Nov	Atp-Z (10%)
Computer as a Tool	0.226***	0.010***	0.140***	0.340***	-5.321**
	(0.014)	(0.002)	(0.024)	(0.076)	(2.122)
Computer as an Object of Study	-0.032	-0.004	-0.132	-0.782*	-7.187
	(0.080)	(0.008)	(0.122)	(0.466)	(9.215)
Computer (Hardware/Software context)	-0.133**	-0.010*	-0.057	-0.107	17.834
	(0.055)	(0.006)	(0.087)	(0.313)	(14.974)
Mention of Computer (other)	-0.071***	-0.000	-0.221***	-0.497***	-2.970
	(0.025)	(0.003)	(0.040)	(0.148)	(6.913)
Number of Authors	0.092***	0.003***	0.059***	0.140***	-0.947***
	(0.004)	(0.000)	(0.004)	(0.013)	(0.260)
NSF Grants (paper)	0.040	0.043	0.206	-0.141	-12.219
	(0.184)	(0.042)	(0.357)	(1.015)	(8.918)
Observations	1,144,696	1,144,645	1,144,696	1,144,696	608,277
R^2	0.714	0.406	0.614	0.577	0.583
Mean of Dep Var	1.534	0.020	2.853	10.264	37.643
Author/Year/Univ FE	Yes	Yes	Yes	Yes	Yes
Subject FE	Topic	Topic	Topic	Topic	Topic

Notes: Paper-author observations, weighted by the inverse number of authors. Categories are generated by passing paper full text to an LLM. In this backfilled specification, papers with persisted first-pass labels keep those labels. Among remaining locally available full-text papers, only those without a computer-keyword hit are assigned to the baseline category, *No Computer Use or Mention*; keyword-hit papers lacking a persisted first-pass label are reclassified from local full text using GPT-5.4. Regressions include author, publication-year, university, and primary-topic fixed effects and control for the number of authors and the number of NSF grants awarded to the paper. Standard errors are clustered at the paper level (OpenAlex Work ID). Sample restricted to locally available searchable full-text papers. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In Table 19, we show that results from Table 3 also hold at the intensive margin. Authors that publish more computer-intensive papers are more productive, have more citations, affiliations, and employ more topics. Though we should expect some mechanical association between publication counts and computer papers even in the absence of selection or causal effects, outcomes for other variables control for the number of publications, in addition to affiliation, cohort, and primary topic fixed effect.[48]

Table 19: Author-Level Regressions: Intensive Margin

	(1) Log Works	(2) Log Cites	(3) H-Index	(4) Topics	(5) Affiliations
Computer Paper Count	0.188***	0.181***	1.166***	0.553***	0.258***
	(0.00790)	(0.0256)	(0.250)	(0.0639)	(0.0294)
Number of Works		0.00553**	0.0582**	0.0128**	0.00631**
		(0.00185)	(0.0189)	(0.00435)	(0.00208)
Observations	316,970	316,970	316,970	316,970	316,970
R^2	0.280	0.417	0.524	0.269	0.317
Mean of Dep Var	2.732	5.158	12.16	16.77	3.525
Affiliation/Cohort/Topic FE	Yes	Yes	Yes	Yes	Yes

SE clustered at affiliation level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Same sample and specification as Table 3, but the key regressor is *Computer Paper Count* (the author's total number of papers flagged for computer-related keywords) rather than a binary adopter indicator. Outcomes are measured over the author's entire career and are defined as in Table 3. Columns (2)–(5) control for the author's total number of works. All columns include affiliation, cohort, and topic fixed effects. Standard errors clustered at the affiliation level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

In Table 20, we examine computer adopters' characteristics at the time their universities first acquired computers. Results indicate significant selection effects: even before computer adoption, these researchers already demonstrate higher productivity and influence compared to their peers. However, the magnitude of these differences is smaller than in the full sample including the post period. Prior to exposure to computers, adopters publish approximately 22% more papers and receive about 60% more citations than non-adopters.

Table 21 restricts the sample to computer adopters and compares early and late adopters. We use adoption lag – the number of years elapsed between the first computer publication and the acquisition of a computer – as a proxy for speed of adoption. Results suggest no

[48]For affiliation FEs, we use the author's modal affiliation as measured by publication counts. For subject fixed effects, we choose again the modal domain/field/subfield/topic as measured by publication counts.

	(1) Log Works	(2) Log Cites	(3) H-Index	(4) Top 1%	(5) Top 10%
Computer Adopter	0.227***	0.592***	1.652***	0.576**	0.0967***
	(0.0182)	(0.0364)	(0.161)	(0.177)	(0.0217)
Number of Works		0.000731*	0.00350	0.00347	0.000412
		(0.000369)	(0.00184)	(0.00204)	(0.000242)
Observations	122,159	134,521	134,521	122,159	122,159
R^2	0.344	0.251	0.246	0.169	0.0964
Mean of Dep Var	1.530	3.460	4.438	2.581	0.267
Affiliation/Cohort/Topic FE	Yes	Yes	Yes	Yes	Yes

SE clustered at affiliation level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 20: Author-Level Regressions: Outcomes Pre-Computer Adoption

Notes: Outcomes are measured strictly *before* the author's earliest exposure to a computer, where exposure year is the lowest installation year among the author's affiliations. Regressor: *Computer Adopter* is an indicator for authors who subsequently publish at least one computer-flagged paper. Outcomes are log works, log citations, H-index, and log counts of the author's top 1% and top 10% cited papers, all measured over the pre-adoption window. Columns (2)–(5) control for number of works. Affiliation, cohort, and topic fixed effects are defined as in Table 3. Standard errors clustered at the affiliation level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

significant difference between early and late adopters. Note, however, that our sample is ends in 1970, so we are comparing authors within the set of relatively early adopters.

Finally, Figure E.15 plots for each year and by group of authors the average number of years an author has been publishing academic articles for, which we use as a proxy for experience.[49] Consistently, computer adopters have more years of experience than the non-adopting counterparts.

[49]That is, for every year in the x-axis, we get the set of authors publishing that year, and get the average years elapsed since first publication from both the set of computer adopters and non-adopters.

Table 21: Author-Level Regressions: Early vs. Late Adopters

	(1) Log Works	(2) Log Cites	(3) H-Index	(4) # Topics	(5) # Affiliations
Adoption Lag (Freq)	0.00817	0.0109	-0.0391	0.0713***	-0.0405*
	(0.00440)	(0.00628)	(0.0368)	(0.0179)	(0.0177)
Number of Works		0.00515***	0.0782***	0.00283***	0.0118***
		(0.000364)	(0.00427)	(0.000380)	(0.000983)
Observations	6,141	6,141	6,141	6,141	6,141
R^2	0.373	0.601	0.759	0.341	0.524
Mean of Dep Var	4.358	7.271	23.65	24.01	6.636
Affiliation/Cohort/Topic FE	Yes	Yes	Yes	Yes	Yes

SE clustered at affiliation level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Sample restricted to computer adopters with publications both before and after the computer installation year at their affiliations. Regressor: *Adoption Lag (Freq)* is the number of years between the author's first computer-flagged publication and the installation year at their affiliations, where the installation year is a publication-count-weighted mean across the author's affiliations; larger values indicate slower adoption relative to local availability. Outcomes are measured over the author's entire career and defined as in Table 3. Columns (2)–(5) control for number of works. Affiliation, cohort, and topic fixed effects are defined as in Table 3. Standard errors clustered at the affiliation level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

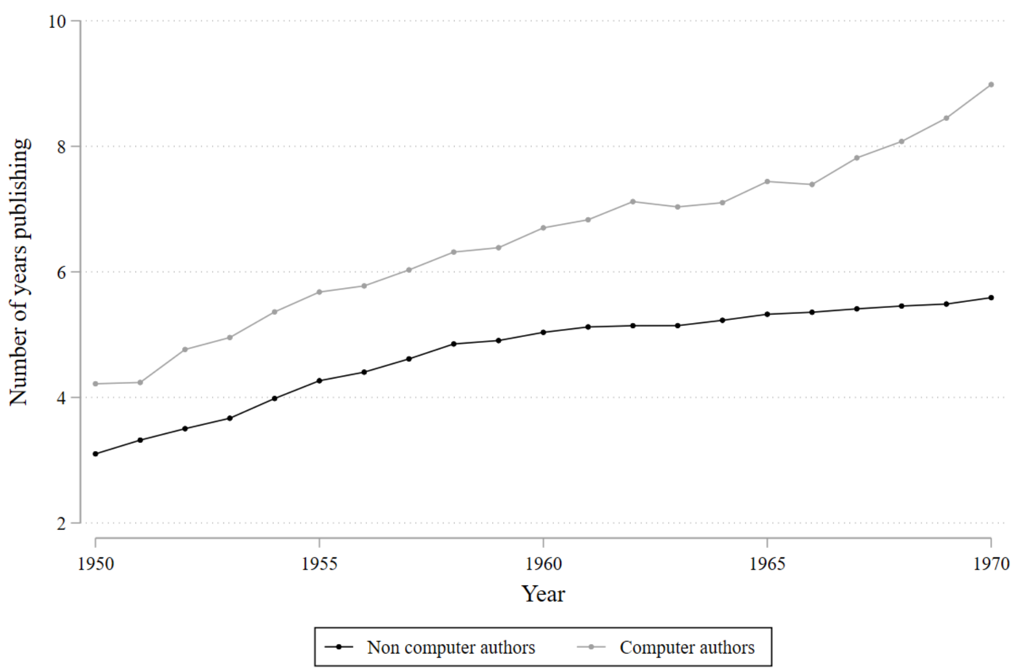


Figure E.15: Average author experience by publication year, computer adopters vs. non-adopters. For each year t, we take the set of authors with a publication in t and plot the mean number of years since their first OpenAlex-observed publication, separately for authors with at least one computer-flagged paper (adopters) and the remainder (non-adopters). Across the full 1950–1970 window, adopters are consistently more experienced than non-adopters. See Tables 19–21 for the corresponding author-level regressions.

E.5 Reference Age and Sleeping Beauties

This subsection reports a robustness check for Section 4.7 that applies the same embedding controls used in Table 14: semantic embeddings from titles and abstracts together with SPECTER embeddings that capture citation-neighborhood position. The overall patterns remain similar to the baseline specifications: computer papers still cite younger references at the minimum and median, older references in the right tail, and references with higher Sleeping Beauty scores among the oldest cited works.

Table 22: Reference Age and Dormancy Outcomes with Embedding Controls

Panel A: Reference age				
	Mean	Min	Median	Max
Computer-Keyword Flag	-0.048*	-0.121***	-0.086***	0.851***
	(0.023)	(0.015)	(0.022)	(0.068)
Observations	1,267,437	1,267,437	1,267,437	1,267,437
R^2	0.000	0.000	0.000	0.000
Mean of Dep Var	-0.405	0.062	0.220	-0.255
Panel B: Reference dormancy (Sleeping Beauty scores)				
	Mean	Median Ref.	Oldest	Age90 Mean
Computer-Keyword Flag	19494.863***	16492.462***	48145.559***	38288.376***
	(3049.162)	(3818.570)	(9517.431)	(8363.198)
Observations	1,266,675	1,262,071	1,263,557	1,264,247
R^2	0.000	0.000	0.000	0.000
Mean of Dep Var	10413.798	-61481.215	52900.751	36815.254

Notes: Columns report second-stage OLS coefficients from the same double-machine-learning robustness specification used in Table 14. In a first stage, both the computer-keyword indicator and each outcome are residualized with respect to author, publication year, university, and primary-topic fixed effects; controls for the number of authors and NSF grants awarded to the paper; dense title and abstract embeddings; and SPECTER paper embeddings. Residuals are estimated with 3-fold cross-fitting and SGD, after which the residualized outcome is regressed on the residualized treatment. Paper-author observations are weighted by the inverse number of authors, and second-stage standard errors are clustered at the paper level (OpenAlex Work ID). Sample restricted to the 1947–1975 searchable-full-text analysis sample used in the paper-level regressions. Panel A reports reference-age outcomes. Panel B reports cited-reference Sleeping Beauty scores, including the score of the median-age cited reference. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

F Issues With Two-Way Fixed-Effects

A recent literature has highlighted potential biases in traditional two-way fixed effects (TWFE) difference-in-differences estimators when treatment effects are staggered and heterogeneous across groups or over time (Goodman-Bacon, 2021; Sun and Abraham, 2021; Callaway and Sant'Anna, 2021; de Chaisemartin and D'Haultfoeuille, 2020).

Our setting presents a textbook failure case of TWFE estimators, due to a combination of staggered adoption and steeply increasing heterogeneous treatment effects over time.

Panels (a) and (b) of Figure F.16 display side-by-side the TWFE and Callaway and Sant'Anna (2021) estimates from the event study specification looking at the computer-related keywords as outcomes, using the exact same sample and treatment definition. We use keyword matches as our benchmark outcome because the effects are very clear, and positive, even by looking at the raw data.

Next, we show in Table 23 a comparison of a simple pre/post difference-in-differences design with the outcome of number of publications with computer keywords. The TWFE design completely flips the coefficients, and it is still very significant.

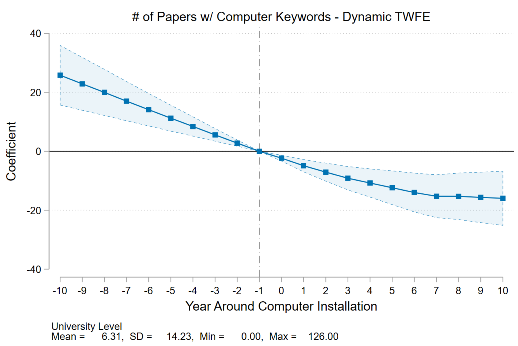
Table 23: Simple DiD estimates (pre/post)

Outcome: # *Papers w/ Computer Keywords*		**Regressor:** *Has Computer*		
Model	Estimate	SE	p-value	N
TWFE	-2.036	0.489	0.000	5 704
Callaway-Sant'Anna (2021)	5.417	0.716	0.000	-
de Chaisemartin-D'Haultfoeuille (2020)	5.417	0.663	0.000	3 312

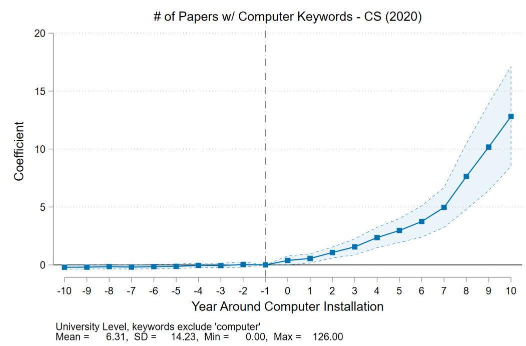
The reason TWFE fails so dramatically in our setting stems from staggered adoption with effects that grow sharply over time. Because TWFE actually compares treated cohorts with already treated ones (Goodman-Bacon, 2021), it confuses trends with effects, and subtracts effects as trends. This effect is mostly driven by "forbidden comparisons" with early-treated cohorts.

Panel (c) of Figure F.16 presents the Goodman-Bacon decomposition of the simple pre/post

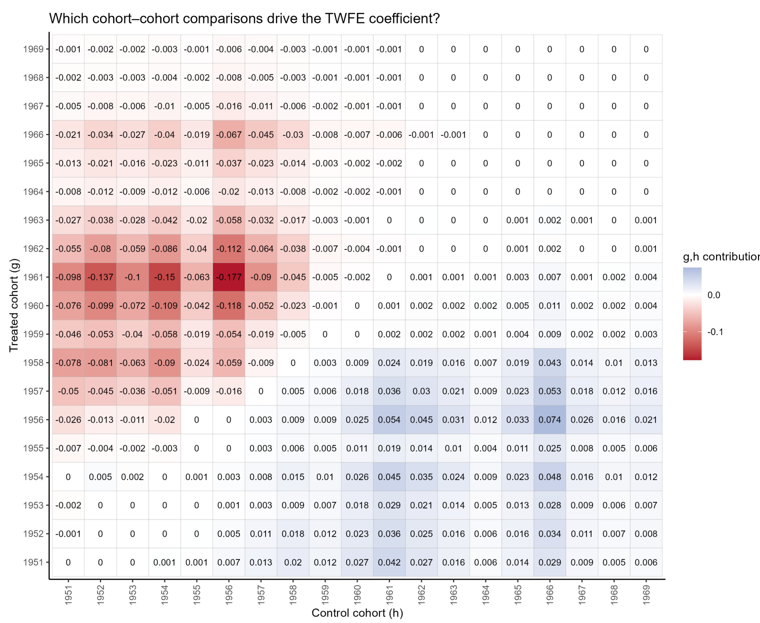
TWFE estimator, which disaggregates the overall coefficient into its component 2×2 difference-in-differences comparisons. Comparisons involving already-treated units – visible in the upper-left triangle of the matrix – yield negative weights, reflecting so-called "forbidden" comparisons that bias the estimate downward. These arise when earlier-treated cohorts are inappropriately used as controls for later-treated ones.



(a) TWFE specification



(b) Callaway-Sant'Anna specification



(c) Goodman-Bacon (2021) decomposition of the TWFE estimand

Figure F.16: TWFE bias diagnosis for the main DiD on the number of computer-related-keyword papers (excluding the word "computer"), 184 treated universities. Panels (a)–(b): event-study coefficients with 95% CIs using TWFE vs. Callaway-Sant'Anna (cohort-aggregated); standard errors clustered by university. TWFE flips the sign of the post-treatment effect. Panel (c): Goodman-Bacon (2021) decomposition of the simple pre/post TWFE estimand from Table 23 into its constituent 2×2 treated-vs.-control comparisons across cohort pairs (g, h). Rows index treated cohort g; columns index control cohort h; cell color is the signed weight of each pairwise comparison.

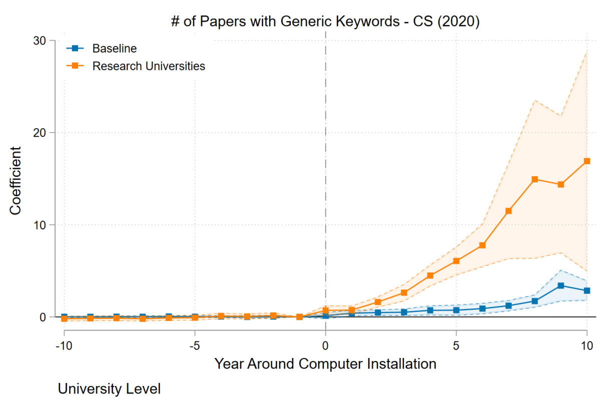
G Robustness & Additional Results

This section presents robustness checks and additional results for Section 6 not included in the main text for conciseness.

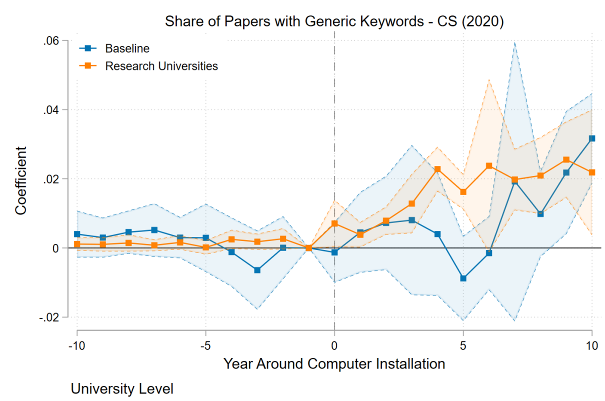
G.1 Computer Adoption

In this Appendix, we extend results in Section 6.1.

Figure G.17 summarizes heterogeneity by research category. Panel (a) shows that research-intensive universities experience larger *absolute* increases in computer-intensive publications, while panel (b) shows *proportional* increases of similar magnitude across categories. In other words, scale differences drive the level effects, but the relative response is broadly comparable.



(a) Count of computer-keyword publications



(b) Share of computer-keyword publications

Figure G.17: Heterogeneity of the main DiD estimate by university research category (Research 1 / Research 2 / other classifications), estimated via Callaway-Sant'Anna (Callaway and Sant'Anna, 2021) at the university-year level. Sample: 184 treated universities. Panel (a): outcome is the absolute count of computer-keyword papers. Panel (b): outcome is their share of total publications. Points are aggregated cohort ATTs with 95% CIs; standard errors clustered by university. Research-1 universities experience larger absolute uptake; proportional responses are broadly comparable across categories.

G.2 Triple Differences: Additional Results

In this section, we display additional relevant outcomes for the results presented in Section 6.2.

Results for Life and Health Sciences. As discussed in Section 6.2, results for the Life and Health Sciences for our outcomes are either null or noisy around zero. In the Life Sciences, we have cases where results hard to interpret due to negative placebo coefficients or pre-trends for some outcomes. We present the results for the outcomes displayed in the main results section on Figures G.18, G.19, and G.20.

(a) Life Sciences

(b) Health Sciences

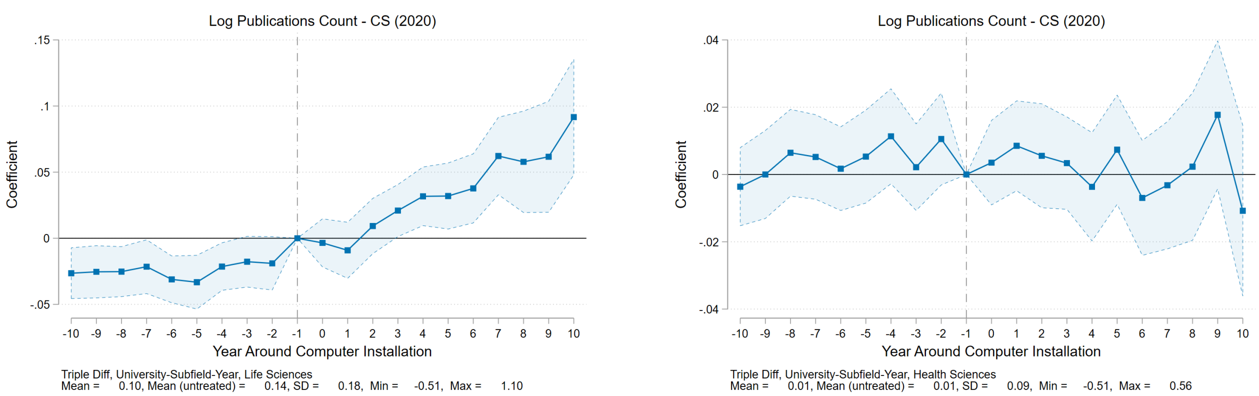
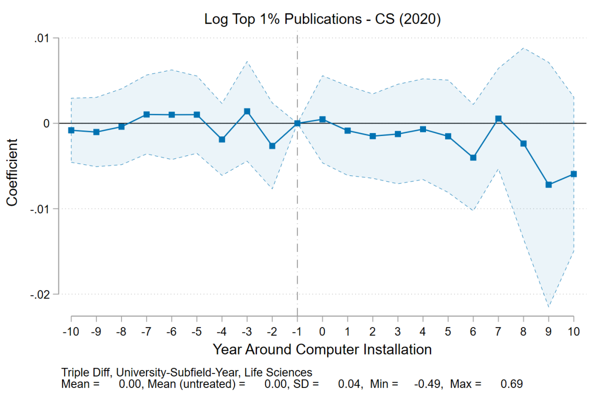


Figure G.18: DDD event study for *log publication counts*, Life and Health Sciences (companion to main-text Figure 8). Y^{*}_{ut} is the within-university gap in log pub counts between compute-amenable and less-amenable subfields (defined by the median 1940–1944 numerical-intensity share within each domain); Callaway-Sant'Anna estimator; $N = 184$ treated universities. Points are aggregated cohort ATTs with 95% CIs; dashed line marks installation. Effects in Life and Health Sciences are smaller and noisier than in Physical and Social Sciences, consistent with weaker pre-digital numerical-intensity gradient in these domains.

Figure G.21 reports two further DDD outcomes: unique authors across the four domains (panels a–d) and top 10% publication counts in Physical and Social Sciences (panels e–f).

(a) Life Sciences



(b) Health Sciences

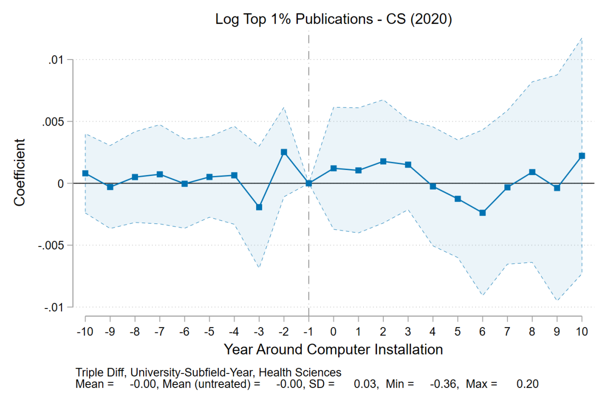
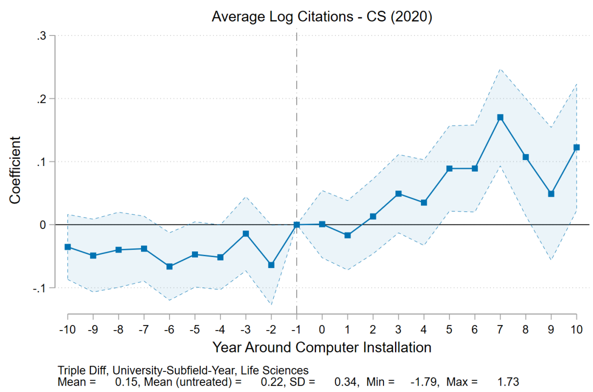
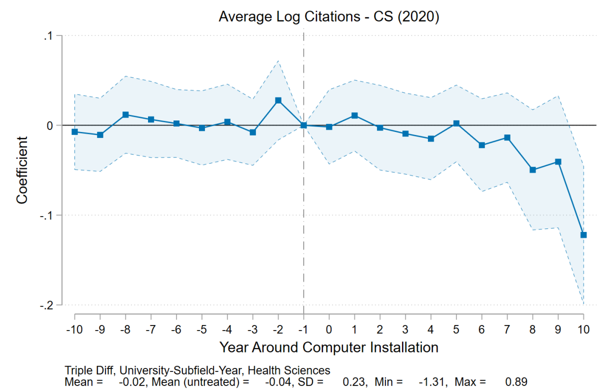


Figure G.19: DDD event study for the *log count of top 1% papers*, Life and Health Sciences (companion to main-text Figure 9). Y^*_{ut} is the within-university gap in the log number of top 1%-cited papers between compute-amenable and less-amenable subfields; Callaway-Sant'Anna estimator; $N = 184$ treated universities. Points are aggregated cohort ATTs with 95% CIs; dashed line marks installation.

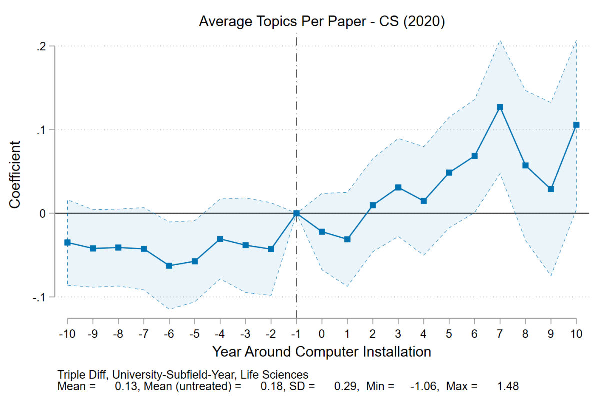
(a) Log avg. citations per paper, Life Sciences



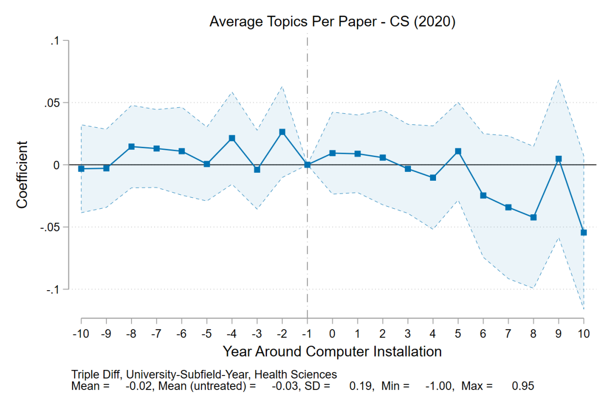
(b) Log avg. citations per paper, Health Sciences



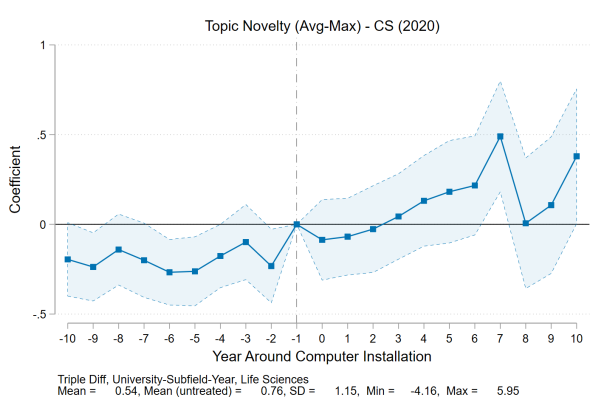
(c) Avg. topics per paper, Life Sciences



(d) Avg. topics per paper, Health Sciences



(e) Topic-combination novelty, Life Sciences



(f) Topic-combination novelty, Health Sciences

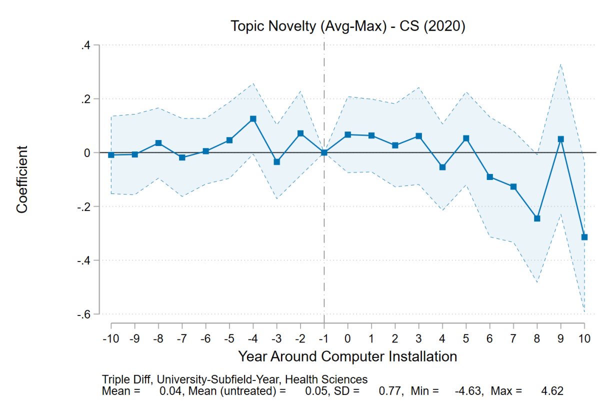
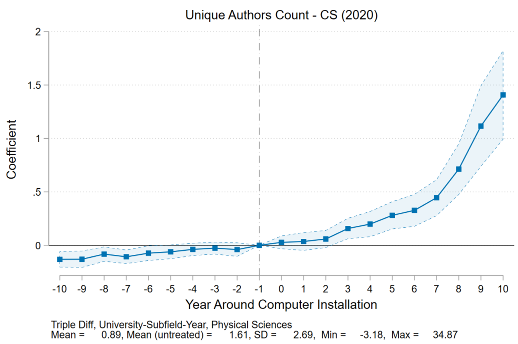
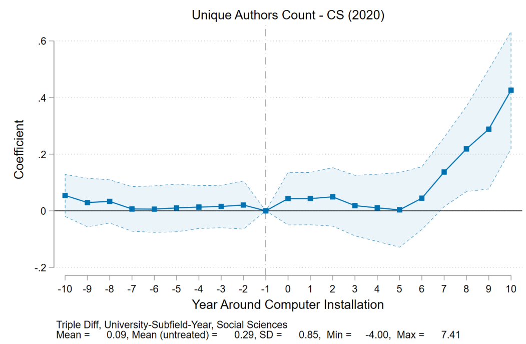


Figure G.20: DDD event studies for three quality- and breadth-related outcomes in Life and Health Sciences: log average citations per paper (panels a–b), average OpenAlex topics per paper (panels c–d), and topic-combination novelty (panels e–f, per-paper max averaged within university-subfield-year). For each outcome, Y^*_{ut} is the within-university gap between compute-amenable and less-amenable subfields; Callaway-Sant'Anna estimator; $N = 184$ treated universities. Points are aggregated cohort ATTs with 95% CIs; dashed line marks installation. Companion to main-text Figures 10, 11, and 12.

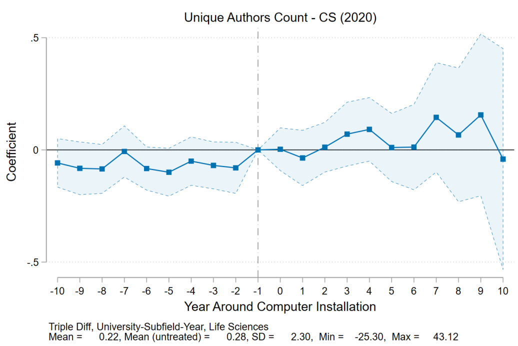
(a) Unique authors, Physical Sciences



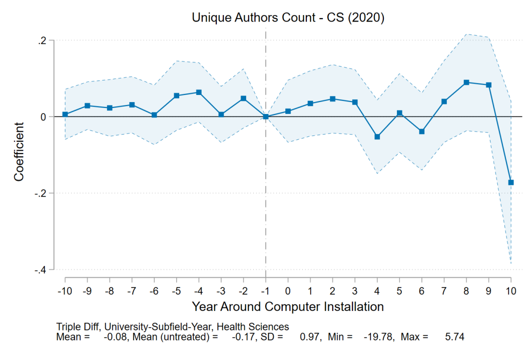
(b) Unique authors, Social Sciences



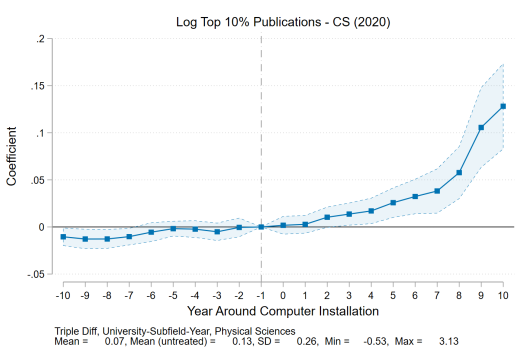
(c) Unique authors, Life Sciences



(d) Unique authors, Health Sciences



(e) Top 10% publications, Physical Sciences



(f) Top 10% publications, Social Sciences

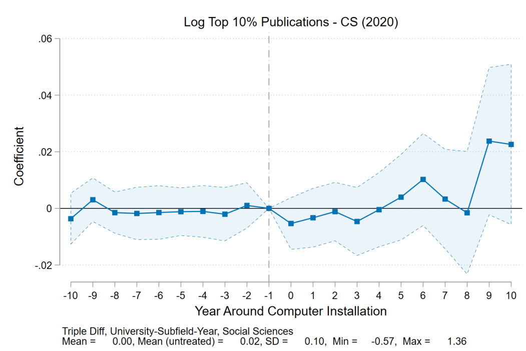


Figure G.21: Additional DDD outcomes, Callaway-Sant'Anna estimator, $N = 184$ treated universities. Panels (a)–(d): number of unique authors per university-year across the four domains; the exposed-unexposed gap grows by about 1.5 authors in the Physical Sciences and 0.4 in the Social Sciences after installation, with Life and Health Sciences imprecise and centered near zero, corroborating Figure 8. Panels (e)–(f): log count of top 10% papers in Physical and Social Sciences, showing that the top 1% results in Figure 9 also hold at the broader top 10% threshold. For each outcome, Y^*_{ut} is the within-university gap between compute-amenable and less-amenable subfields; points are aggregated cohort ATTs with 95% CIs; dashed line marks installation.