What Do Graduate Student Unions Do?

John Fallon Jesse Meredith

Working Draft, November 2026

Abstract

About 36 percent of U.S. research doctorates and 37 percent of university publications now come from universities whose graduate students bargain under a union contract, up from about 3 percent of doctorates in 1990. Using stacked event studies on the date each first contract is ratified, we find that stipends rise about 6 percent beyond local cost-of-living growth. Doctorates awarded change little for three years. Six or more years after ratification, when the first cohorts admitted under a contract reach graduation, they are about 11 percent lower. The decline is deepest in education and the humanities and smaller, though not absent, in science and engineering. Priced from the contracts themselves, the implied elasticity of doctorates with respect to their cost is 1.08, about 2.6 times the median for workers in firms. Treated universities also employ fewer ladder faculty and publish less in total. They award fewer bachelor's degrees as well, though that decline began before the campaign. We rule out large changes in output per faculty member and in the placement and research of the remaining graduates. Across the treated universities, the decline amounts to about 2.5 percent of all U.S. research doctorates each year, so how universities absorb a contract now moves the national supply of researchers.

Keywords: graduate student unions; collective bargaining; doctoral education; higher education finance; research universities; stipends.
JEL codes: J51, I23, J24, I26, J45.

1 Introduction

How do graduate student unions change what universities do? Universities with a graduate union contract now award about 36 percent of U.S. research doctorates (Figure 1) and produce 37 percent of university publications. That reach matters beyond the campus: regions with more universities grow faster (Valero and Van Reenen, 2019), university research spills over into the local economy (Hausman, 2022; Kantor and Whalley, 2014), and PhDs are overrepresented among inventors (Toivanen and Väänänen, 2016). We know what contracts do to pay (Liu, 2005; Schenk, 2010) and how students experience them (Rogers et al., 2013), but not what they do to the rest of the university: whom it trains, whom it employs, what it publishes and whom it teaches. We compare each research university whose graduate workers have ratified a first contract, before and after, with research universities whose graduate workers never unionized. Both groups are R1 universities, those with the highest research activity in the 2015 Carnegie classification. First contracts raise stipends about 6 percent beyond local living costs. Six or more years later, these universities award about 11 percent fewer doctorates, employ fewer ladder faculty and publish less. A contract is also associated with 12 percent fewer bachelor's degrees. The cuts are deepest in education and the humanities. We can rule out large changes in output per faculty member, and in where the remaining graduates work and what they publish.

A university chooses how many graduate workers to hire, as a firm does, but it is an unusual employer. Economists have long found the demand side of the market for new scientists hard to model, because so little is known about how universities behave (Stephan, 1996). A university is typically a nonprofit that sells its product below cost, covers the difference from appropriations, gifts and endowment earnings, and competes on prestige, which draws those gifts (Winston, 1999). After grants, students at R1 universities paid about 50 percent of what their education cost in 2024 (IPEDS Finance; Desrochers and Hurlburt, 2016). Graduate students sit inside that structure twice: undergraduate revenue helps pay for their training, and as teaching and research assistants they are inputs the university pays through aid.

When graduate workers organize, such an employer meets both faces of unionism (Freeman and Medoff, 1979, 1984). As a monopoly, a union raises the wage, which induces employers "to hire fewer workers." As collective voice, it lets workers change the conditions of work rather than leave, with effects that depend on how management responds. Graduate workers have unusual reason to use voice, because the research is their own training. The contracts we collected set the price of an appointment, never the number, so the number stays the university's choice. If the university uses that margin, the effect reaches beyond

the workers who organized, to the doctorates it trains and the research assistants it funds on grants.

The price moves first: contracts raise pay, as unions do elsewhere. Faculty unions at Canadian universities raise salaries 6.1 percent within six years of certification (Baker et al., 2024). For graduate students, Liu (2005) finds teaching-assistant stipends 4.7 percent higher where a union bargains, and 6.6 percent higher where one is organizing without recognition. Schenk (2010) finds a similar premium confined to covered titles. Both compare institutions at one time. We follow each university through its first contract, on a union roster we expanded to the campus level and re-dated against the Hunter College directory of bargaining agents in higher education (Herbert et al., 2024). Stipends rise 6.3 percent net of metropolitan cost-of-living growth, and 9.0 percent by the fourth contract year. That rise is larger than in close private-sector elections and in bargaining laws for public school teachers, where union effects on pay are near zero (DiNardo and Lee, 2004; Frandsen, 2016, 2021; Lovenheim, 2009; Paglayan, 2019).

The quantity follows. Firms buy less labor after a union wins: after a narrow NLRB election win, establishment employment falls 5.3 percent, payroll 7.7 percent, and survival 4 points within three years (Frandsen, 2021). Across a wide range of vote shares, employment is about 13 percent lower five years after a win, most of it through closure (Wang and Young, 2023). Nursing homes cut aide hours per resident by about 15 percent (Sojourner et al., 2015). For graduate unions, Schenk (2010) names the admissions response, and Julius (2023), writing around 2000 from the employer's side, predicts it. Neither tests it. Universities, we find, buy less graduate labor too: doctorates awarded fall by 11.5 percent, and research assistantships in science and engineering by about 14 percent. The decline is small for three years (-5.2 percent at $\tau + 3$) and reaches its full size by $\tau + 5$ to $\tau + 6$, when the first cohorts admitted under a contract reach graduation. A regional trend has no reason to wait for that clock, so for doctorates the timing is our sharpest identification evidence. The decline is confined to the parts of the university graduate workers touch. Professional schools, which employ few graduate workers, do not shrink: law, dentistry and pharmacy degrees hold steady (Section 4.2). To compare the doctorate response with other employers', we price the cost increase from the contracts themselves. We collected 48 contracts, the first where we could find it (26) and otherwise a successor, and large language model agents coded them provision by provision. A contract raises the annual cost of a student by about 11.2 percent. The implied elasticity of doctorates with respect to that cost is 1.08 (95 percent interval 0.5 to 1.7), about 2.6 times the median labor-demand elasticity for workers in firms, 0.42 (Lichter et al., 2015).

Unionized firms do less research than other firms, and unionized universities produce

less research in total. Three years after a union narrowly wins an election, firms patent about 8.7 percent less, and publicly traded ones cut R&D spending relative to their assets (Bradley et al., 2017). Cross-sectional comparisons put union firms' R&D spending about 15 percent lower (Hirsch, 2004). To follow research into every field, we matched 2.07 million U.S. doctoral dissertations from ProQuest to OpenAlex author records. Compared with R1 and R2 universities, treated universities publish less in total after a contract, and they employ fewer ladder faculty. Faculty unions, by contrast, leave faculty headcount unchanged (Baker et al., 2024). We can rule out large changes in output per faculty member and in the placement and research records of the graduates still produced.

The cut is deepest outside science and engineering, though stipends rise by similar amounts across fields. Education falls 24.8 percent and the humanities 17.4, against 10.5 in engineering, and the physical sciences do not fall (+0.7). Outside science and engineering the contraction is 1.1 to 2.1 times as deep as inside it. The gap rests partly on one large university, without which the ratio is 1.7 (Section 5). Who pays for a student is measured less precisely, so for those measures we add R2 universities, the next Carnegie tier. On that wider sample, fields contract roughly in the order of the share of their students who pay their own way (rank correlation -0.67, permutation $p = .042$). Graduate tuition also rises after a contract (Appendix E.3), and the rise falls on students who pay their own way. In the model of Appendix E, calibrated, it can account for the gap between fields, though the data do not establish that it does. A university trains in many fields, and the same cost shock reaches them unevenly.

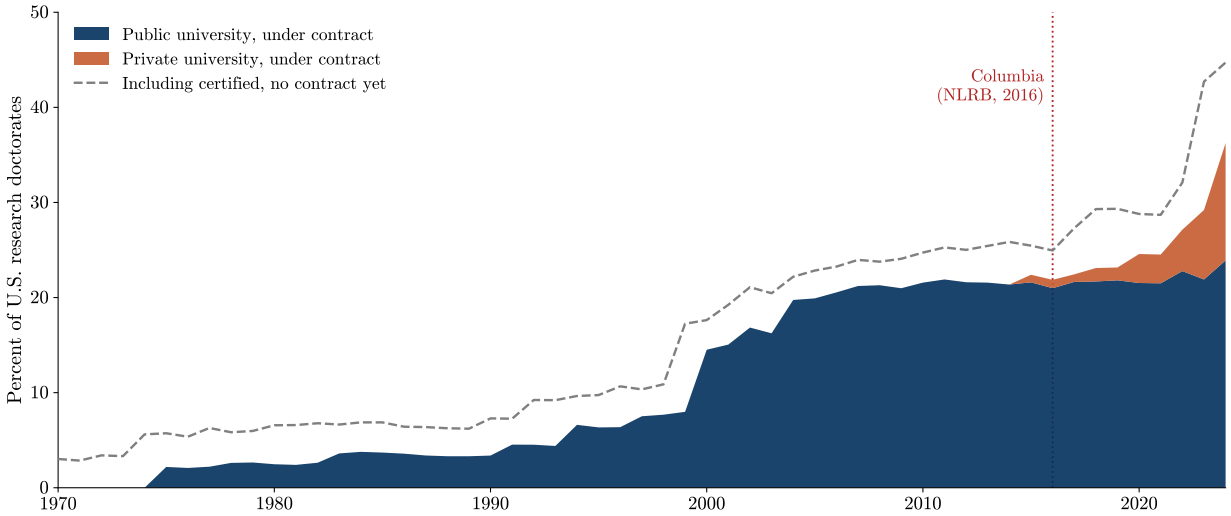
2 Setting

Graduate unions bargained for decades under state public-sector law before private universities could organize at all. The first contracts came at public universities in the upper Midwest and the Northeast in the 1970s, and thirty of the sixty-six dated first contracts were signed before 2014. Yet in 1990 universities with a graduate contract awarded only about 3 percent of U.S. research doctorates, against 36 percent in 2024 (Figure 1).

For private universities, the National Labor Relations Board held in *New York University* (2000) that graduate assistants were employees, reversed itself in *Brown University* (2004) and restored the *NYU* position in *Columbia University* (2016). Certifications at private universities followed within a year, and organizing accelerated from 2021 (Herbert et al., 2024). Representation has more than doubled since 2012, when 64,424 graduate student employees were represented. First contracts therefore came at public universities from the 1970s through the 2000s and at private universities after 2016 (Figures 2 and 3). Appendix B

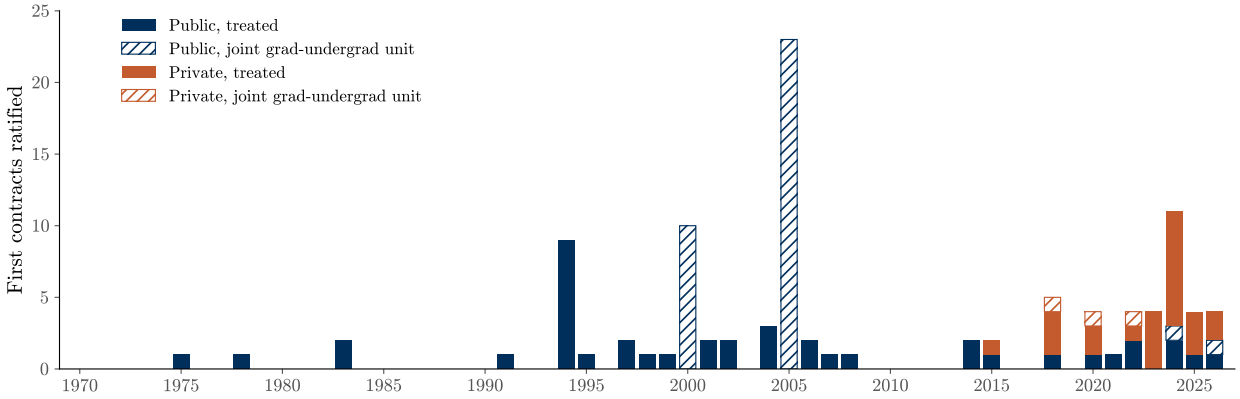
documents the event clocks and how the roster was built.

Figure 1: Share of U.S. Research Doctorates Awarded Under a Graduate Contract



Notes: Share of all research doctorates in the Survey of Earned Doctorates awarded at universities whose graduate workers bargain under a ratified contract, by year and IPEDS sector (HD2024 control). A university counts from its first-contract year; campuses of joint graduate-undergraduate units count from their first contract before the unit became joint. Contracts with no known year are left out. The dashed line adds universities with a certified union and no contract yet. The dotted line marks the NLRB's *Columbia University* decision (2016).

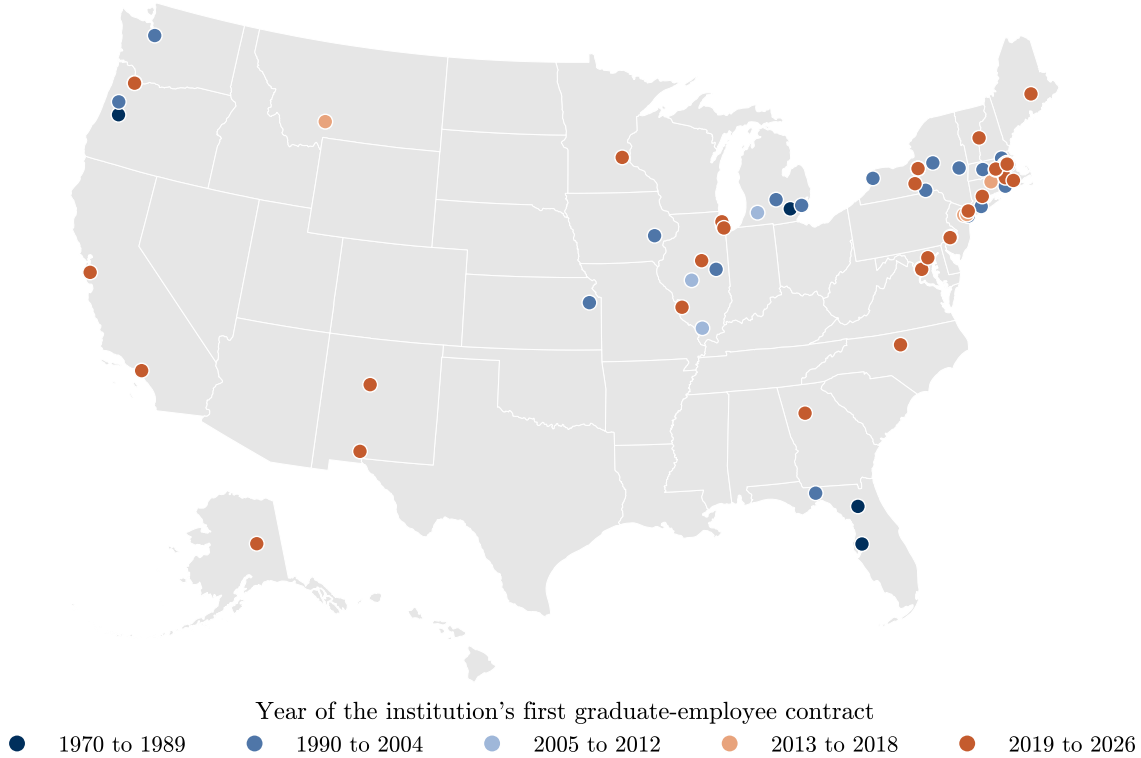
Figure 2: First Graduate-Employee Contracts by Year and Sector



Notes: First contracts ratified per year, by IPEDS sector (HD2024 control). Solid bars are the 66 institutions with a dated first contract that form the treated set. Hatched bars are campuses of joint graduate-undergraduate bargaining units, dated at their first contract, which are organized but held out of the treated set and every control pool because their contracts price undergraduate labor along with graduate labor (Appendix B). A system-wide contract counts once per campus.

A contract sets the price of an appointment and not the number of them. Of the 48

Figure 3: Institutions with a First Graduate-Employee Contract



Notes: Each point is one of the 66 treated institutions at its IPEDS coordinates, colored by the year of its first contract, on Census cartographic state boundaries (2023, 1:20m), Albers equal-area projection with Alaska and Hawaii inset. Joint graduate-undergraduate units are not plotted.

agreements we coded, 45 set a minimum pay rate, and where one states the university's share of the health premium, the share averages 95.2 percent. None sets a number of appointments or an admissions target. The university chooses how many students to fund at the contract price, as in the model of Appendix E, so the quantities in Section 4 are its response rather than a term of the bargain. Table A17 in Appendix I reports the full coding.

We date treatment to ratification of the first contract, the first date at which the stipend floor and any funding guarantee bind. An organizing campaign comes first. Certification follows, by an election or voluntary recognition, and obliges the university to bargain but sets no new terms of pay. Ratification comes last, at most institutions quickly. The median lag is 1 year, and 61 of the 66 institutions where we observe both dates ratify within two years. The longest wait is 4 years, at the New Jersey Institute of Technology, and 6 certified units have no contract yet. On the same treated institutions, the stipend effect is +5.5 percent dated from ratification and +3.1 dated from certification. The certification clock spreads the ratification-year jump across the lag. Appendix B.1 reports the stipend estimates on both clocks.

3 Data and Design

3.1 Data

Stipends come from 17,958 crowd-sourced reports on phdstipends.com, validated against University of California payroll (Appendix C). Doctorates come from the NSF Survey of Earned Doctorates (SED), the estimate of record, with IPEDS Completions as a corroborating census. Research assistantships come from the NSF Survey of Graduate Students and Postdoctorates (GSS), which covers only science, engineering and health. University research output and graduates' placements and research come from the ProQuest–OpenAlex match, faculty headcount from the IPEDS Faculty Salaries and Human Resources surveys, and undergraduate outcomes from IPEDS and the College Scorecard. Union events come from a hand-collected roster validated against the Hunter directory (Herbert et al., 2024). Every institution is treated at its first ratified contract, except the campuses of joint graduate-undergraduate bargaining units. Their contracts also raise the price of undergraduate student workers, the nearest substitute for graduate workers, at the same time, so they are neither treated nor in any control pool (Appendix B.2). Treated and comparison universities alike are research universities, R1 in the 2015 Carnegie classification, so a unionized research university is compared with research universities whose graduate workers never organized. Graduate tuition, self-supported students and every input to the

model of Appendix E are less precisely measured, and for them we add R2 universities to both groups. On R1 universities alone they are similar but less precise, and we report both. Appendix B states the rule, the alternatives and every source in detail, and Table A2 traces the roster to each estimation sample.

Even among research universities, treated institutions are larger than the controls over the years before treatment (Table 1). They award 331 research doctorates a year against 208, and employ 1,060 ladder faculty against 888. About the same share of their doctorates is outside science and engineering (0.29 against 0.28). Institution fixed effects absorb level differences of this kind. What the design requires is parallel trends, tested below.

Table 1: Treated and Control Institutions Before Treatment

	Treated	Control	Normalized	N	
	mean (s.d.)	mean (s.d.)	difference	Treated	Control
A. Institution (SED)					
Research doctorates awarded per year	331 (184)	208 (113)	0.81	34	43
Share of doctorates outside S&E	0.29 (0.12)	0.28 (0.11)	0.09	37	43
Private	0.41 (0.50)	0.26 (0.44)	0.33	34	43
B. Graduate students (NSF GSS, S&E and health)					
Full-time graduate students	2,206 (1,400)	1,628 (778)	0.51	25	43
Research assistants	667 (599)	527 (357)	0.28	25	43
Teaching assistants	442 (261)	384 (261)	0.22	25	43
C. Stipends (phdstipends.com, 2012–2015)					
Median stipend, dollars	28,001 (3,570)	22,969 (4,142)	1.30	18	32
D. Faculty and research output (IPEDS, OpenAlex)					
Ladder faculty	1,060 (421)	888 (336)	0.45	30	41
Articles published per year	4,173 (3,848)	2,184 (1,116)	0.70	30	41
Citations to a year's output, thousands	268.9 (276.0)	116.2 (75.2)	0.75	30	41
E. Graduating cohorts (ProQuest matched to OpenAlex)					
Dissertations per cohort	258 (182)	182 (117)	0.50	22	43
Share with any publication after the degree	0.69 (0.06)	0.74 (0.06)	-0.88	22	43
Academic share of named first placements	0.73 (0.03)	0.73 (0.04)	-0.03	22	43
F. Undergraduates (IPEDS)					
First-year retention, percent	94.0 (6.6)	87.3 (6.9)	0.99	8	43
Six-year graduation rate	0.65 (0.15)	0.65 (0.15)	-0.07	7	42

Notes: Means with standard deviations in parentheses, across institutions. Each variable is read from the stacked panel on which its outcome is estimated, with that panel's control pool (never-organized R1 universities in the 2015 Carnegie classification), over the fully untreated pre-period of each stack, $\tau \in [-10, -5]$ on the ratification clock. An institution's values are averaged over its baseline years first, so each institution counts once; shares are weighted within the institution by the graduates or cohort behind them. Stipends: the main stipend sample (contracts ratified 2014–2025 against never-organized R1 universities), calendar years 2012–2015, treated institutions before ratification, report-weighted. Normalized difference: the treated minus control mean over the square root of the average of the two variances. N: institutions.

3.2 Design

Stacked event studies. We build one stacked dataset for each ratification cohort c (Cengiz et al., 2019). It contains the R1 universities whose first contract was ratified in year c and all never-organized R1 universities, over a window from $c-10$ to $c+8$. Merged universities are consolidated: every school a university later absorbed is summed into it in every year. Pooling stacks, the difference-in-differences specification is

$$y_{ict} \;=\; \alpha_{ic} \;+\; \gamma_{ct} \;+\; \beta^{\text{camp}}\, D_{ict}^{[-4,-1]} \;+\; \beta^{\text{part}}\, D_{ict}^{[0,5]} \;+\; \beta^{\text{full}}\, D_{ict}^{[6,+]} \;+\; \varepsilon_{ict}, \tag{1}$$

where α_{ic} are stack-by-institution fixed effects, γ_{ct} are stack-by-year fixed effects, and each D_{ict} indicates a treated institution of stack c in the stated event-time window. $D^{[-4,-1]}$ covers the organizing campaign, $D^{[0,5]}$ the years in which graduating cohorts were admitted before the contract, and $D^{[6,+]}$ the years in which they were admitted under it. For outcomes with no pipeline lag, such as stipends, the last two indicators pool into one post indicator. Event-study versions replace the bins with one indicator per event year. Counts are estimated by Poisson pseudo-maximum likelihood on zero-filled panels,

$$E[\, y_{ict} \,|\, \cdot\,] \;=\; \exp\!\Big(\alpha_{ic} + \gamma_{ct} + \beta^{\text{camp}} D_{ict}^{[-4,-1]} + \beta^{\text{part}} D_{ict}^{[0,5]} + \beta^{\text{full}} D_{ict}^{[6,+]}\Big), \tag{2}$$

so that an institution that stops awarding a degree stays in the risk set.

Identification and inference. The design requires that treated and never-organized institutions would have moved in parallel absent the contract. We test this on the pre-campaign years, $\tau-10$ through $\tau-6$ (Table A14). We estimate the campaign window rather than assume it away, since a campaign is when a university has most reason to respond in advance. About 86 percent of treated institutions are in Democratic-leaning states, so regional trends are the main threat to the design. Section 4.2 reports what absorbing them does. Standard errors are clustered by bargaining unit, so the campuses of a public system under one contract form one cluster. Because several outcomes rest on few treated institutions, every main estimate is re-estimated under the null-imposed wild cluster bootstrap (Cameron et al., 2008; MacKinnon and Webb, 2018), with randomization inference where the treated count is smallest or the estimator rules out the bootstrap. Appendix A.1 states the five identifying assumptions in full and Appendix H.4 the inference.

4 What a First Contract Does to the University

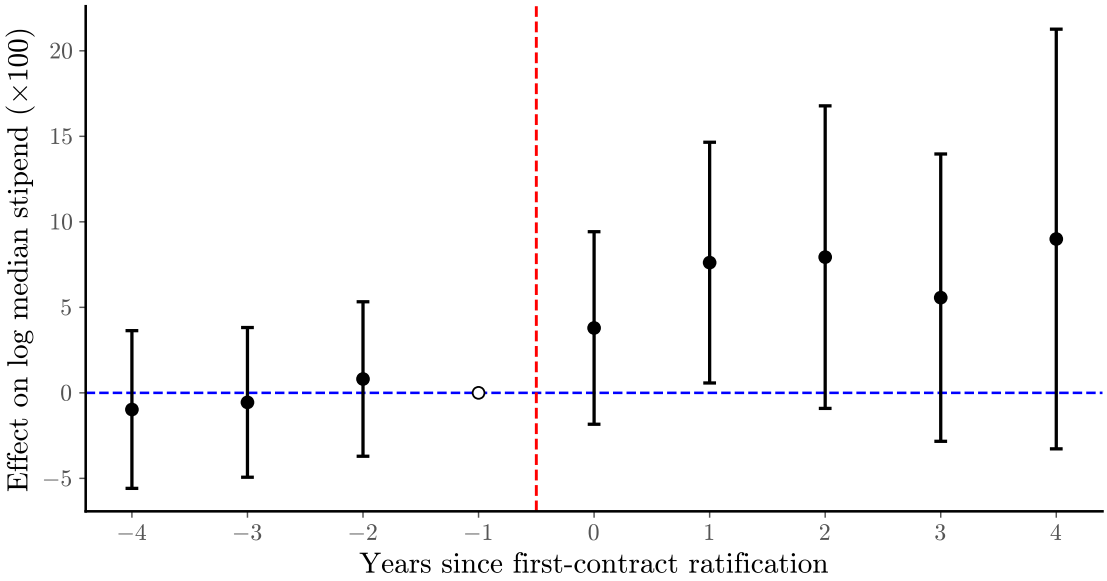
A first contract raises what a university pays for graduate labor, and the university becomes smaller in the parts graduate workers touch. Stipends rise about 6 percent. Six or more years after ratification, treated universities award about 11 percent fewer doctorates and enroll fewer first-year students. Compared with R1 and R2 universities, they employ about 6 percent fewer ladder faculty and receive about 11 percent fewer citations to a year's output, though output per ladder faculty member does not detectably change (Section 4.7). In science and engineering, research assistantships fall about 14 percent from three years after ratification. Professional schools, which employ few graduate workers, do not shrink (Section 4.2). Four or more years after ratification, a contract is also associated with about 12 percent fewer bachelor's degrees (Section 4.9).

4.1 Stipends Rise

Stipends rise at ratification and not before it (Figure 4). We measure the stipend as the median that students report on phdstipends.com, by university, broad field and year. Reports thin in the two years around ratification, but their field mix holds and the estimate does not depend on those years (Appendix C.2). Before ratification, stipends at treated universities move with those at control universities. They begin to rise in the ratification year and are 7.6 percent higher in the first full contract year, as bargained raises take effect. On the 19 universities that ratified from 2014, the ones the stipend data observe before and after, stipends rise 6.3 percent net of metropolitan cost-of-living growth, the same as without that adjustment. Pooling all 37 contracted universities with stipend reports, including those ratified before the data begin, gives 5.5 percent (Table A4), and the roster codings give 2.9 to 6.1 percent.

Our estimate is larger than the near-zero pay effects in close private-sector elections (DiNardo and Lee, 2004; Frandsen, 2021) and under bargaining laws for public school teachers (Frandsen, 2016; Lovenheim, 2009; Paglayan, 2019). Those designs date treatment to a certification or a statute, not to a contract. The estimate falls within the cross-sectional graduate premia of about 5 to 9 percent (Liu, 2005; Schenk, 2010). If control universities answered nearby organizing with raises, the threat effect Liu (2005) documents, our estimate would be biased toward zero.

Figure 4: Stipends Around First-Contract Ratification



Notes: Event-study coefficients on log median stipends (phdstipends.com) on the first-contract ratification clock, times 100, with 95 percent intervals; endpoints binned at ±4, reference $\tau = -1$. The event study restricts treatment to the 19 institutions with contracts ratified 2014–2025, against never-organized research universities; the pooled difference-in-differences of Table A4 uses 37 treated institutions. Institution and field-bucket×year fixed effects, weighted by cell report counts, standard errors clustered by bargaining unit.

4.2 Universities Award Fewer Doctorates

Doctorates do not detectably change for three years after ratification. From the third year on, the decline grows, to 11.5 percent at $\tau + 6$, when the first students admitted under a contract graduate, and 13.9 percent at $\tau + 8$ (Figure 5). We observe research doctorates at 34 treated universities in the Survey of Earned Doctorates, 21 of them for six or more years after ratification. Before the organizing campaign, doctorates at treated universities move with those at control universities. Each year stays within 2.5 percent of the year before ratification, and together the years are indistinguishable from zero. The two groups part only as each university's first post-contract students near graduation, a timing a regional trend would be unlikely to share.

Averaged over the years six or more after ratification, doctorates at treated universities are 11.5 percent lower. IPEDS, the other census of doctorates, gives about the same answer: 11.9 percent in the fields where the two censuses count the same degrees, and 9.3 percent in all fields. IPEDS also counts practice degrees such as the EdD and the DNP. By 2023 it recorded 3.6 education doctorates for every one in SED, so we take the analysis on SED as the estimate of record.

The timing does not by itself rule out a steady drift that began before ratification. In

the years before the campaign, doctorates at treated universities trend down slightly relative to controls, by 0.42 log points a year, a slope not distinguishable from zero. Extending that slope through the years after ratification leaves a decline of 7.6 log points six to eight years out, with an interval from −20.8 to +5.7, so the pre-campaign terms are too noisy to rule out a drift of that size. Under the sensitivity analysis of Rambachan and Roth (2023), the decline stays distinguishable from zero as long as departures from parallel trends after ratification stay below 1.25 times the largest change between years before the campaign.

The decline holds against most alternatives (Table A12, Appendix H). Averaging each university's percentage change gives 14.4 percent with every university counted equally and 9.9 percent with larger programs counted more. With all non-union doctoral universities as controls, a broader and less similar group than R1 universities, the decline is 8.8 percent. Treated universities sit mostly in Democratic-leaning states, and letting those states follow their own trend leaves the decline at 12.2 percent (Appendix H.2).

Nor is the decline the work of a budget shock that also prompted organizing. Treated universities' state appropriations, total tuition revenue, gifts and endowments do not detectably change during the campaign or after ratification. Controlling for them leaves the decline at 10.7 percent, against 10.5 percent on the same years without them (Appendix H.3). The decline also stops at the parts of the university graduate workers do not touch. Professional schools employ few graduate workers, and at treated universities they award as many degrees after a contract as before (Table A13). Pooled, dentistry, pharmacy and optometry degrees rule out a decline beyond about 0.7 percent. Public and private universities show no detectably different response (Appendix H.1).

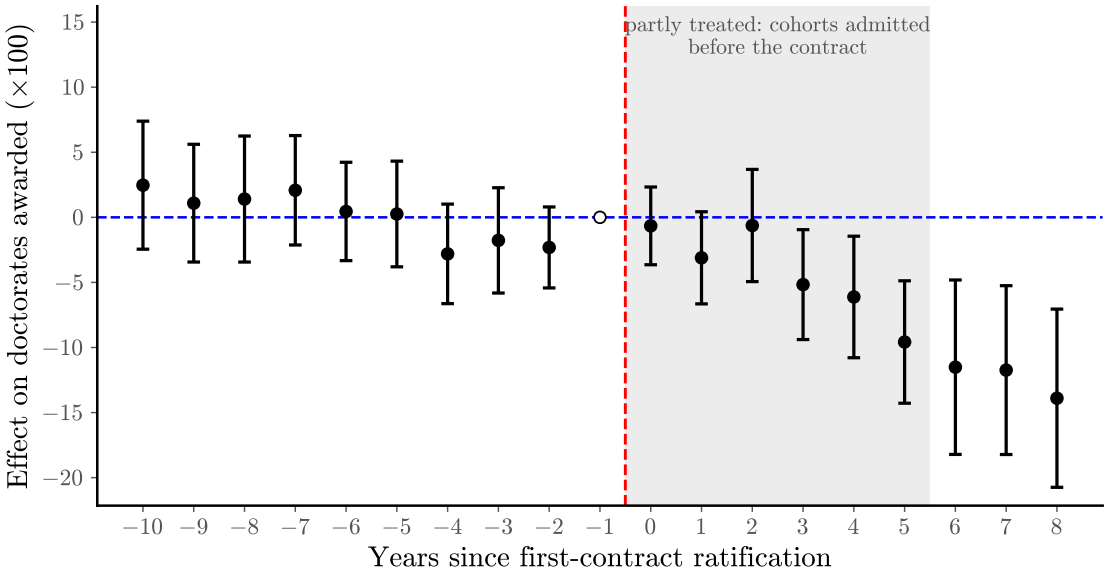
4.3 Demand for Graduate Workers Is More Elastic Than for Workers in Firms

Per unit of cost, universities cut doctorates more than firms cut workers. A contract raises the annual cost of a student by about 11.2 percent, and doctorates fall by a little more, an elasticity of 1.08, about 2.6 times the median for workers in firms. We price the cost increase from the contracts themselves and measure the elasticity as

$$\hat{\eta} = -\frac{\Delta \ln N}{\Delta \ln W}, \tag{3}$$

with $\Delta \ln N$ the fall in doctorates estimated in Section 4.2 and $\Delta \ln W$ the rise in the cost of a student. An elasticity of one means doctorates fell in proportion to cost, and zero means they did not respond.

Figure 5: Doctorates Awarded Around First-Contract Ratification (SED)



Notes: Stacked event-study coefficients on doctorates awarded (SED, Poisson), times 100, with 95 percent intervals, on the first-contract ratification clock, reference $\tau = -1$. All-cohort stacks: every contract cohort contributes its available event years up to $\tau + 8$, never-organized R1 controls, institution×stack and cohort×year fixed effects, clustered by bargaining unit (34 treated, 21 observed at $\tau \geq 6$). The red line marks ratification; the gray band marks $\tau \in [0, 5]$, whose graduating cohorts were admitted before the contract. Treated institutions: 34, 21 of them observed at $\tau \geq 6$. The pre-campaign slope in the text weights the pre-campaign coefficients by their covariance, as the sensitivity bounds do.

The cost increase comes from the contracts, not the stipend alone. A contract also sets the employer share of the health premium and, at some universities, waives mandatory fees. Not all of that is new cost. Of the 48 agreements, 19 waive no fees and only 3 waive them in full, and 50 percent write a benefit as the continuation of one already in place, which locks in a cost without raising it. Figure 6 builds the cost increase from these provisions. On a stipend base of $33,500, the stipend alone raises the cost of a student 6.1 percent. The health premium and the fee waivers raise it to 16.0 percent. Netting out what contracts merely continue gives 11.2 percent, the shock we use.

Dividing the fall in doctorates by that shock gives an elasticity of 1.08, with a 95 percent interval from 0.5 to 1.7, and from 0.72 to 2.01 across the ways of pricing the contract (Table 2). Table A19 in Appendix I sets it beside published labor-demand elasticities. It is above every public-employer and minimum-wage estimate there and close to the long-run response of firms once capital adjusts, though demand for skilled workers is usually less elastic (Lichter et al., 2015). Fields differ widely (Section 5.2).

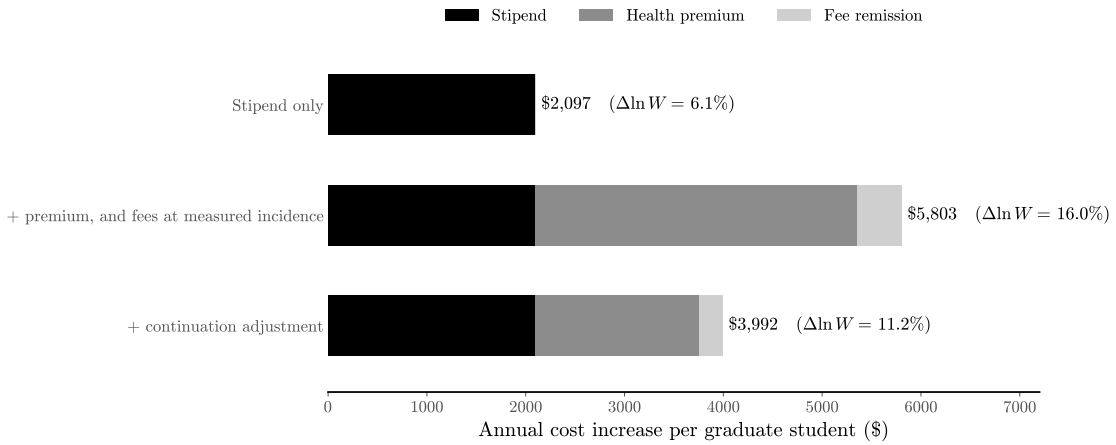
A contract could instead have raised employment. An employer with monopsony power faces an upward-sloping supply of labor, so it hires fewer workers and pays them less than they are worth. A union wage can then raise hiring, because paying more to the workers already employed is no longer a cost of adding one more (Ashenfelter et al., 2010). That power does not need a single employer: it arises wherever workers bear a cost of changing jobs, and a student partway through a doctorate bears a large one. Supply to a single employer can be very inelastic, about 0.1 for nurses at individual hospitals (Staiger et al., 2010). Doctorates fall instead, most likely because funded places are rationed. A program that turns applicants away at the going stipend is not short of students, so a higher stipend reaches it as a higher cost, not as a larger pool of applicants.

Stipend data begin in 2012, so the cost increase is priced only on contracts from 2014 on. Using it for every contract assumes that earlier contracts raised the cost of a student by a similar share. On the contracts ratified 2014 to 2018, the only ones both data sets observe, stipends rise 5.2 percent and doctorates fall 15.2 log points, an elasticity of 1.47. The stipend estimate there rests on 3 universities and the doctorate one on 4, too few to test the assumption (Table A18).

4.4 Research Assistantships Fall in Science and Engineering

Research assistantships fall after ratification (Figure 7). We count them in the GSS, which records graduate students by type of support in science, engineering and health. At the 25 treated universities, research-assistant counts are about 12 percent lower in the first

Figure 6: What a First Contract Adds to the Annual Cost of a Graduate Student



Notes: Annual cost increase per graduate student under each pricing row of Table 2, split into the stipend increase (the cost-of-living-adjusted stipend effect applied to the stipend base), the employer health premium and mandatory-fee remission, with the implied $\Delta \ln W$. The premium is each university's coded employer share of its measured single-coverage premium, and fees are IPEDS required graduate fees, applied at the fee incidence and continuation shares coded from the agreements.

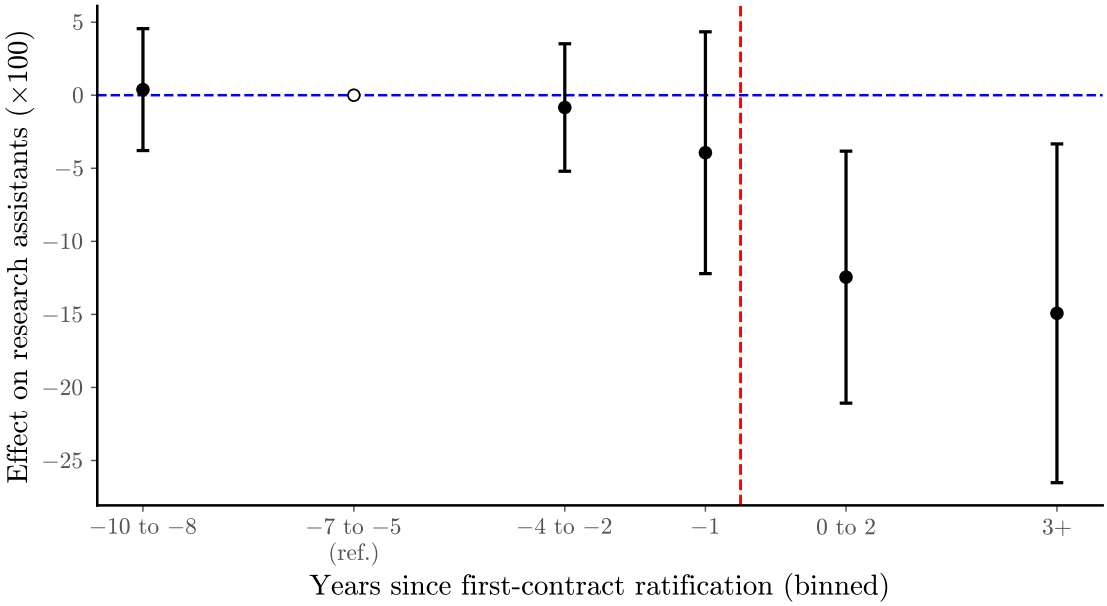
Table 2: Implied Elasticity of Doctoral Production to the Cost of Graduate Labor

	Annual cost	$\Delta \ln W$	Implied $-\mathrm{d}\ln N/\mathrm{d}\ln W$	
Components priced	increase ($)	(%)	All fields	Outside S&E
Stipend only	2,097	6.1	2.01	3.44
+ premium, and fees at measured incidence	5,803	16.0	0.76	1.31
+ continuation adjustment	3,991	11.2	1.08	1.85
Upper bound: every provision new and full	6,167	16.9	0.72	1.23

Notes: The stipend base is $33,504, the median at treated institutions in the four years before ratification (phdstipends, 386 cells), and the stipend increase is the +0.063 of Section 4.1 (net of metropolitan price growth). The fee bill, $1,366, is the mean IPEDS required fee for full-time in-state graduate students over the four years before ratification at the 35 coded contracts with data (median $954). The employer premium, $3,251, is the mean across the 29 coded contracts with a sourced single-coverage premium within two years of the year before ratification of the coded employer share times that premium (median $3,116). The rows apply multipliers coded from the 45 priced agreements, one per treated institution (the agreements collected less the three joint graduate-undergraduate units): the first contract where it could be found (25), otherwise a successor. Fee incidence: 18 agreements contain no fee-remission provision, 24 a partial or conditional one, and 3 a full one, so the measured row remits 33 percent of a full fee bill on average, weighting a conditional provision at one half. The continuation adjustment multiplies both components by 0.51: 22 of 45 agreements state the benefit as a continuation of an existing one rather than a new grant. That count is a floor, since it reads coded excerpts rather than full contract text. Employer premium shares are high and flat where stated: mean 95.4 percent and median 100 percent of the individual premium across the 28 agreements stating a numeric share. Completions responses are the fully treated coefficients on the field panel, pooled across contract cohorts; the cost shock is priced on contracts from 2014 on, the ones the stipend panel observes. All ratios are of log changes.

three years and about 14 percent lower from the third year on, and before ratification they move with those at control universities. A research assistant is paid from the grants of the faculty member who employs them, and once start-up funds are spent the stipend is that faculty member's to cover (Stephan, 1996). A contract that raises the stipend raises the cost of an assistant to a grant of fixed size. Teaching-assistant counts and the full-time enrollment stock do not detectably change. Every specification we estimate on this panel is negative for research assistants (Appendix F).

Figure 7: Research Assistants Around First-Contract Ratification (GSS)



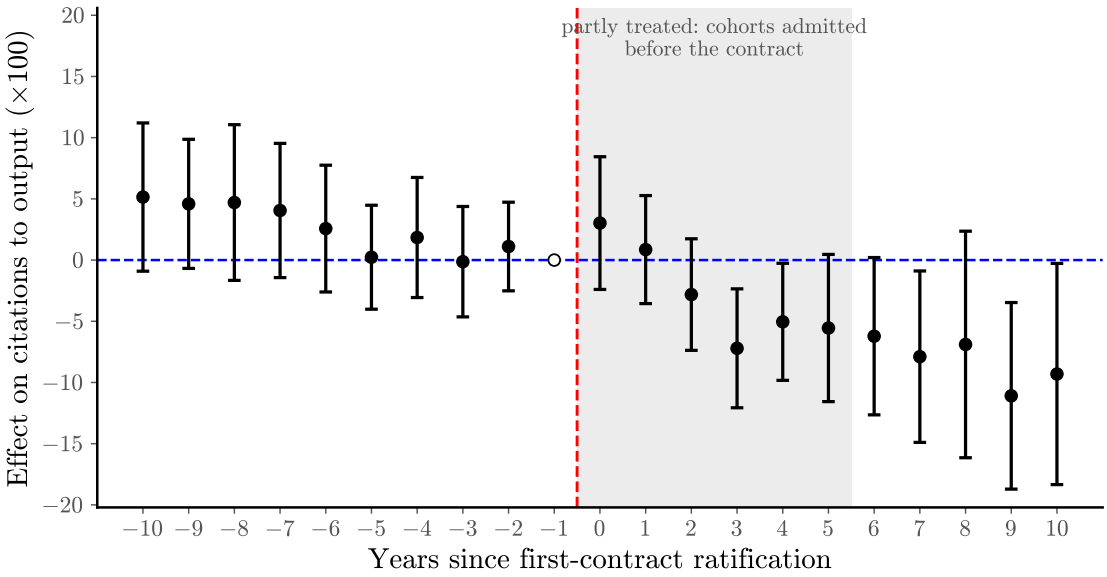
Notes: Poisson event-bin coefficients on research-assistant counts, times 100, with 95 percent intervals, relative to $\tau \in [-7, -5]$ (hollow marker). Stacked ratification-clock design, R1 comparison group, institution×stack and cohort×year fixed effects, clustered by bargaining unit; 25 treated institutions. The GSS covers science, engineering and health fields only. Treated institutions: 25, 21 of them observed at $\tau \geq 3$.

4.5 Treated Universities Publish Less in Total

Six or more years after ratification, treated universities publish less, and their work is cited less. We count every work OpenAlex records under an institution's name in a year, in every field, and the citations those works have received since. Here and in the faculty counts that follow, the comparison group adds R2 universities to R1, because only 17 treated R1 universities are observed that late. Citations do not move before ratification and are 11 percent lower from the sixth year on (Figure 8). On R1 universities alone the decline is

similar, 8 percent, but not distinguishable from zero. Articles fall about 6 percent, which is not distinguishable from zero on either comparison group.

Figure 8: Citations to Institutions' Output Around First-Contract Ratification



Notes: Stacked event-study coefficients on log citations received by all works published under the institution's name in the year, all fields, times 100, with 95 percent intervals, ratification clock, reference $\tau = -1$; institution×stack and cohort×year fixed effects, R1 and R2 comparison group, clustered by bargaining unit. The gray band marks $\tau \in [0, 5]$. Treated institutions: 47, 28 of them observed at $\tau \geq 6$.

4.6 They Also Employ Fewer Ladder Faculty

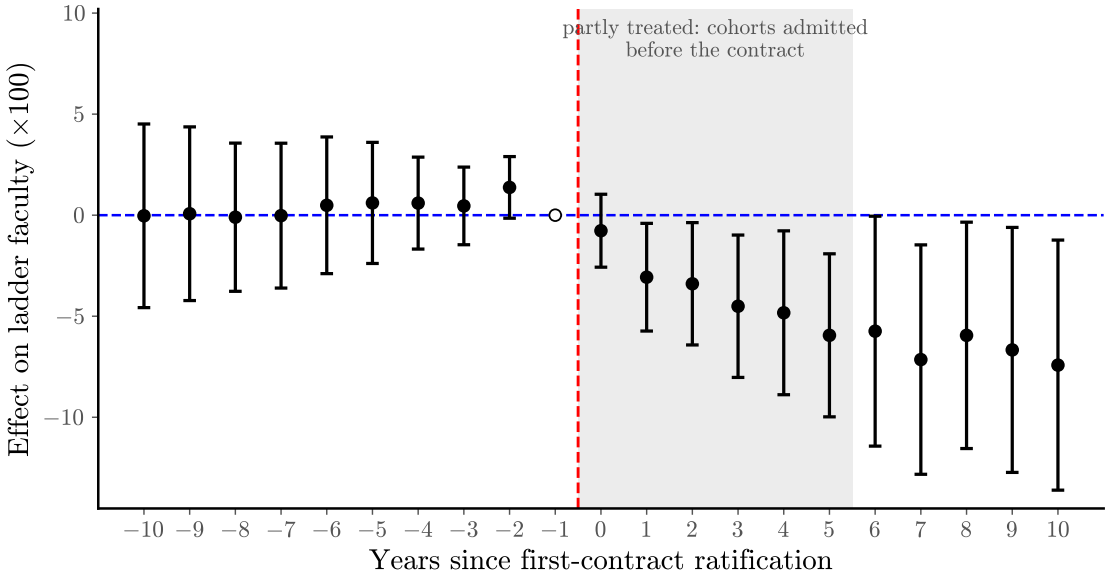
Treated universities also employ fewer ladder faculty, the full-time full, associate and assistant professors counted in the IPEDS Faculty Salaries survey. Before ratification, ladder faculty move with those at control universities. After it they fall, by about 4 percent over the first five years and about 6 percent six to ten years out (Figure 9). Averaged over every year from the sixth on, the decline is similar but just short of distinguishable from zero at the 5 percent level. The decline does not come from a change in which universities the survey reaches. On the 20 treated universities observed in every surveyed year from $\tau - 5$ to $\tau + 10$, ladder faculty fall about 6 percent over $\tau + 6$ to $\tau + 10$. The count is by rank, so it mixes tenure-system and non-tenure-track professors. Tenure-system faculty alone, counted from 2002, fall by a similar amount, though their pre-campaign test rejects (Appendix F.1).

By rank, the decline is among full and associate professors (Figure 10). Six to ten years after ratification, full professors fall about 9 percent and associate professors about 9 percent, and both move with those at control universities before ratification. Assistant professors do

not detectably change. Unreplaced departures and retirements would shrink the senior ranks, and so would slower promotion, and headcounts cannot tell these apart. A hiring freeze at entry would show up first among assistant professors, and it does not.

Contingent faculty do not detectably rise in their place (Figure 11). Six to ten years after ratification, neither full-time non-tenure-track faculty nor part-time instructors are distinguishable from zero, and the part-time series fails its pre-campaign test. Both come from the IPEDS Human Resources survey, which begins in 2002, so fewer treated universities are observed six or more years out.

Figure 9: Ladder Faculty Around First-Contract Ratification

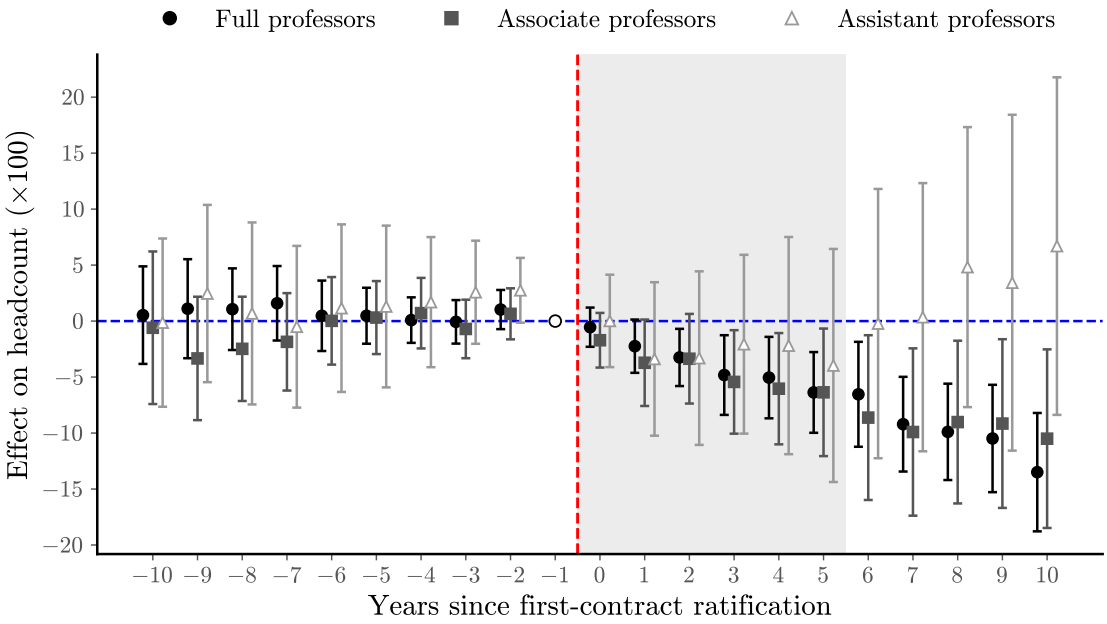


Notes: Stacked event-study coefficients on log ladder faculty (full, associate and assistant professors, IPEDS Faculty Salaries survey), times 100, with 95 percent intervals, ratification clock, reference $\tau = -1$; design as in Figure 8. Treated institutions: 47, 24 of them observed at $\tau \geq 6$.

4.7 Output per Faculty Member Does Not Detectably Change

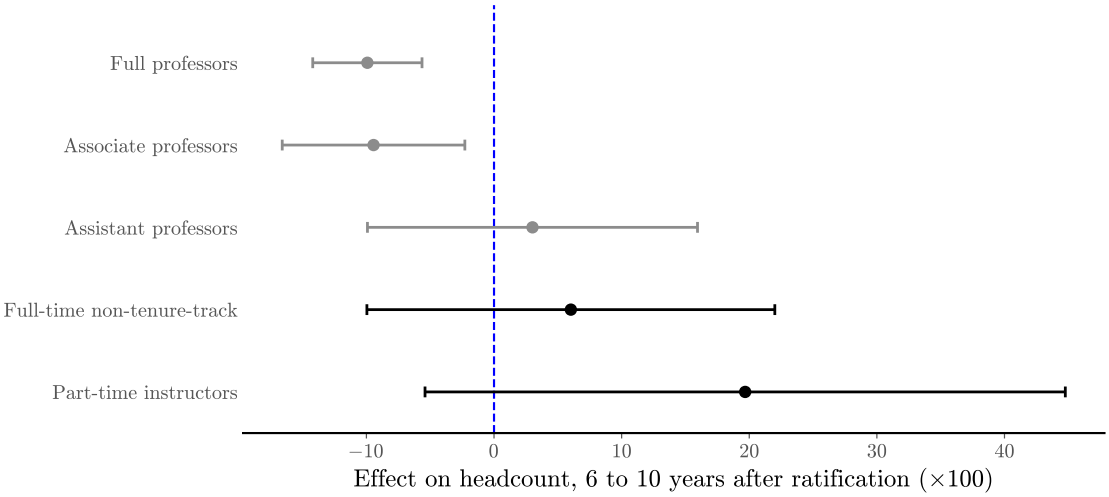
Total output falls, but output per ladder faculty member does not detectably change. We divide citations to a year's output by that year's ladder headcount. Six or more years after ratification, the ratio rules out a decline of more than about 14 percent, and its path is flat before and after ratification (Figure 12). The citation rank of the university's articles within their field and year, a measure of impact rather than volume, does not detectably change either. Treated universities publish less, but with fewer faculty, and we cannot detect that each faculty member's output falls.

Figure 10: Faculty Headcount by Rank Around First-Contract Ratification



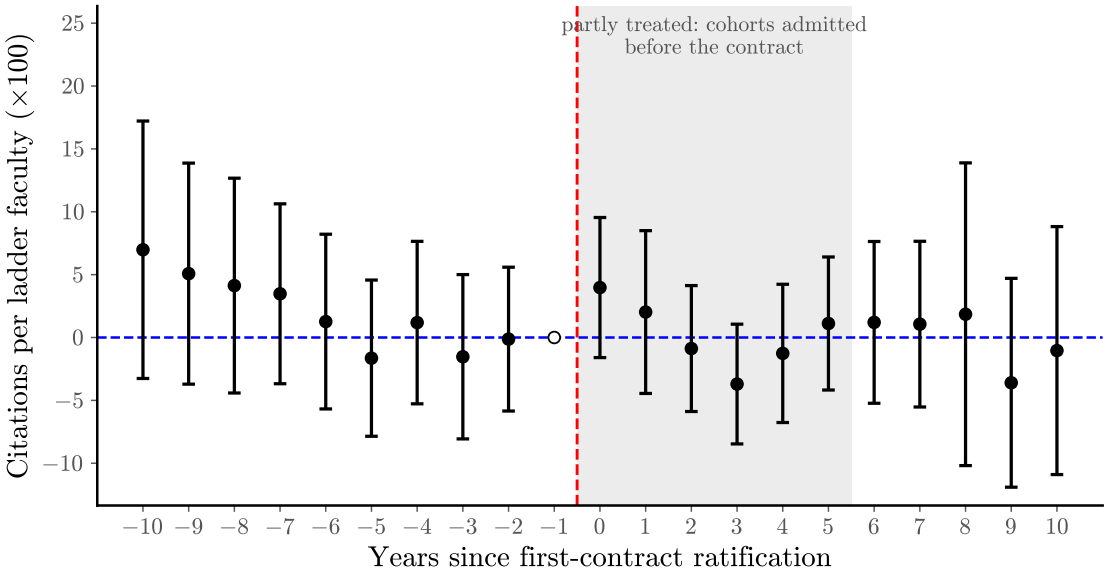
Notes: Stacked event-study coefficients on log headcount by rank, times 100, with 95 percent intervals, ratification clock, reference $\tau = -1$, markers offset horizontally for legibility; same data, sample and design as Figure 9. Treated institutions: 47, 24 of them observed at $\tau \geq 6$.

Figure 11: Faculty Headcount by Type, Six to Ten Years After Ratification



Notes: Event-bin coefficients for $\tau \in [6, 10]$ on log headcount, times 100, with 95 percent intervals. Ladder ranks (gray) from the IPEDS Faculty Salaries survey; full-time non-tenure-track faculty and part-time instructors (black) from the IPEDS Human Resources survey, 2002 onward. Design as in Figure 10.

Figure 12: Citations per Ladder Faculty Member Around First-Contract Ratification



Notes: Stacked event-study coefficients on log citations to a year's output per ladder faculty member, times 100, with 95 percent intervals, ratification clock, reference $\tau = -1$; design as in Figure 8. Treated institutions: 47, 24 of them observed at $\tau \geq 6$.

4.8 Graduates Place and Publish as Before

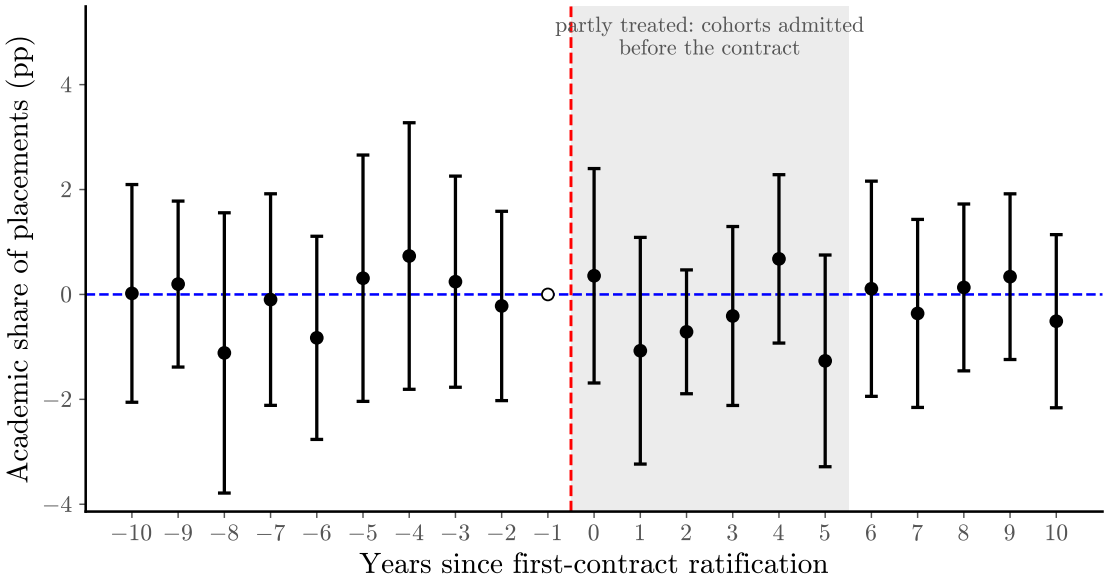
Where graduates of treated universities go, and how much their work is cited, do not detectably change. We follow each degree cohort by matching ProQuest dissertations to OpenAlex author records, which name a graduate's first employer and record their articles and citations. Against withheld ORCIDs, the match names the right person in 96 percent of top-tier cases (Appendix G). We see a placement only once a graduate publishes under an employer's name, so every sector share is conditional on a research record. Twenty-two treated universities enter, eighteen of them observed six or more years after ratification. The design is the institution design by degree cohort: shares in levels weighted by graduates, and per-graduate counts in Poisson with cohort size as exposure.

Among graduates with a named first placement, the share placed in academia stays near 72.9 percent. The bootstrap interval rules out a fall of more than 1.4 points or a rise of more than 1.7 (Figure 13), and the path is flat on both sides of ratification. So the students universities stop producing were not mainly those headed for industry. Six or more years after ratification, treated cohorts have about 8 percent fewer graduates with a named placement. Holding that fall at its estimate, between 11 and 47 percent of the missing graduates would have gone outside academia, against 27 percent of a cohort before ratification (Appendix G). The share of each cohort with a named placement rises slightly, even as their number falls.

Bounding that selection puts the effect between 0.1 and 0.5 points, or −2.2 to 3.8 with sampling error (Appendix G).

Their research records do not detectably change either. Within field and cohort, a graduate's lifetime citations rank where they did before ratification: the percentile moves by −0.2 points, a change not distinguishable from zero, and the path is flat on both sides (Figure 14). Citations to a graduate's whole record rule out a decline of more than about 8 percent. Matching errors occur equally at treated and control universities, so they scale each estimate and its bound toward zero together. Appendix G reports what match quality implies for each bound, along with the other research-record outcomes.

Figure 13: Academic Share of Graduates' First Placements Around First-Contract Ratification

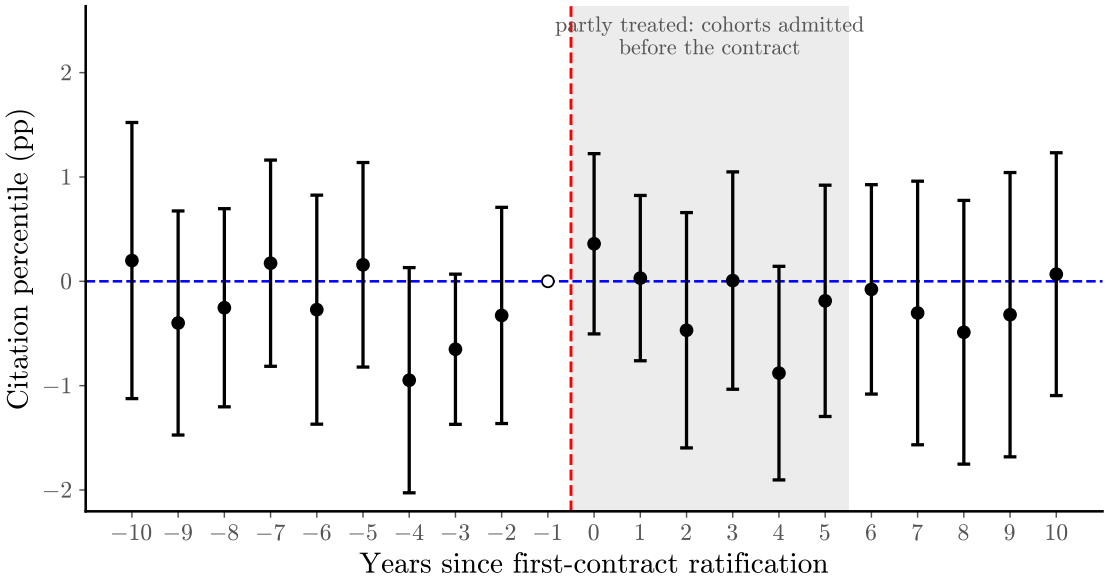


Notes: Stacked event-study coefficients on the academic share of named first placements (graduating cohort, year of degree), in percentage points, with 95 percent intervals, ratification clock, reference $\tau = -1$; cell×stack and cohort×year fixed effects, weighted by named placements, R1 comparison group, clustered by bargaining unit. The gray band marks cohorts of $\tau \in [0, 5]$, admitted before the contract. Treated institutions: 22, 18 of them observed at $\tau \geq 6$.

4.9 Undergraduates

Treated universities also award fewer bachelor's degrees, but we cannot attribute the decline to the contract. We count first majors in IPEDS Completions on the doctorate estimate's stacks and clock (Figure 15). Degrees are about 4 percent lower in the campaign window and about 12 percent lower from the fourth year after ratification, when students

Figure 14: Graduates' Citation Percentile Around First-Contract Ratification



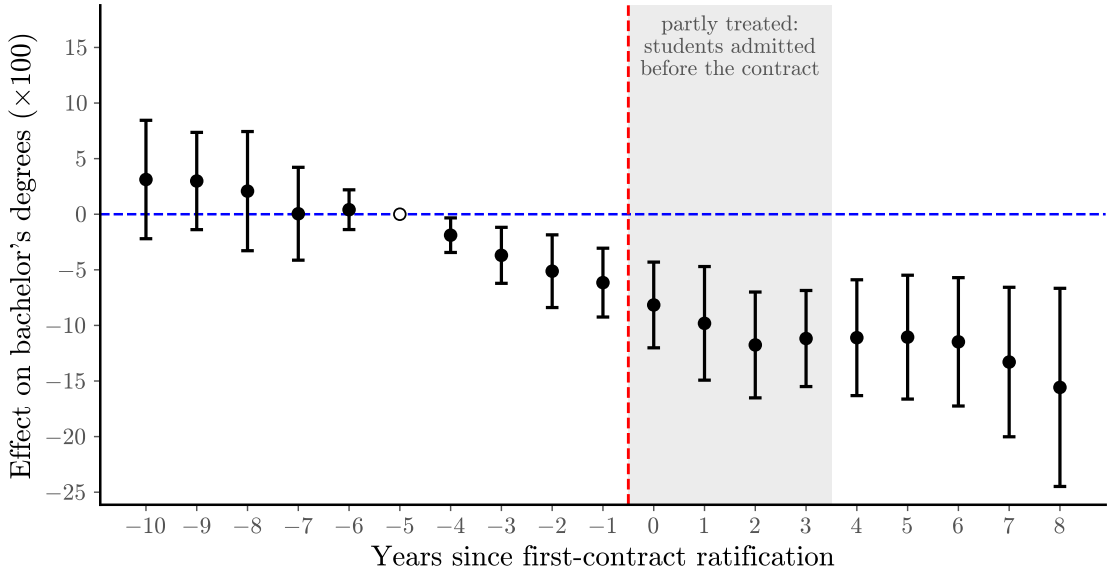
Notes: Stacked event-study coefficients on the mean percentile of graduates' lifetime citations within SED field and degree cohort, in percentile points, with 95 percent intervals; design as in Figure 13, weighted by graduates. Treated institutions: 22, 18 of them observed at $\tau \geq 6$.

admitted under a contract begin to graduate. The decline starts before any of those students could finish. The same path appears on the 14 treated universities observed in every year from $\tau - 10$ to $\tau + 8$, so it does not come from a change in which universities are observed. Entering classes shrink too. Six or more years after ratification, treated universities enroll about 8 percent fewer first-time undergraduates and about 9 percent fewer freshmen, though applications do not detectably fall. Degrees and entering classes alike slope down before the campaign, degrees by about 0.8 log points a year, though not by enough to reject a flat pre-period. Extending that slope leaves 4.1 of the 11.1 log points at $\tau + 4$, and the smoothness bounds of Rambachan and Roth (2023) include zero at $\tau + 2$, $\tau + 4$ and $\tau + 6$. With no break in timing to set it apart from a trend, we cannot establish that these declines are effects of the contract.

Among students who do enroll, graduation rates do not detectably change, and first-year retention dips. We follow undergraduates by entering cohort and compare cohorts entering after ratification with those that finished before the contract arrived (Table A1, Appendix A.2). The 150 percent-time graduation rate rules out a change of more than about 5.7 points either way, and its pre-campaign terms are flat (Figure 16). First-year retention falls about 3.0 points on a base of 93 percent, an interval from 6.1 points down to just below zero. Both results rest on few universities, 7 for graduation and 8 for retention,

so every interval is bootstrapped.

Figure 15: Bachelor's Degrees Awarded, by Year Relative to Ratification



Notes: Stacked event-study coefficients on bachelor's degrees awarded (IPEDS Completions, award level 5, first majors), times 100, with 95 percent intervals, from a Poisson regression with stack×institution and cohort×year effects; hollow marker is the reference year $\tau - 5$; the gray band marks years in which graduates were admitted before the contract. SED institution stacks, calendar years 1984 on, contract cohorts 1989 on; R1 comparison group, clustered by bargaining unit; 30 treated institutions, 19 observed at $\tau \geq 4$.

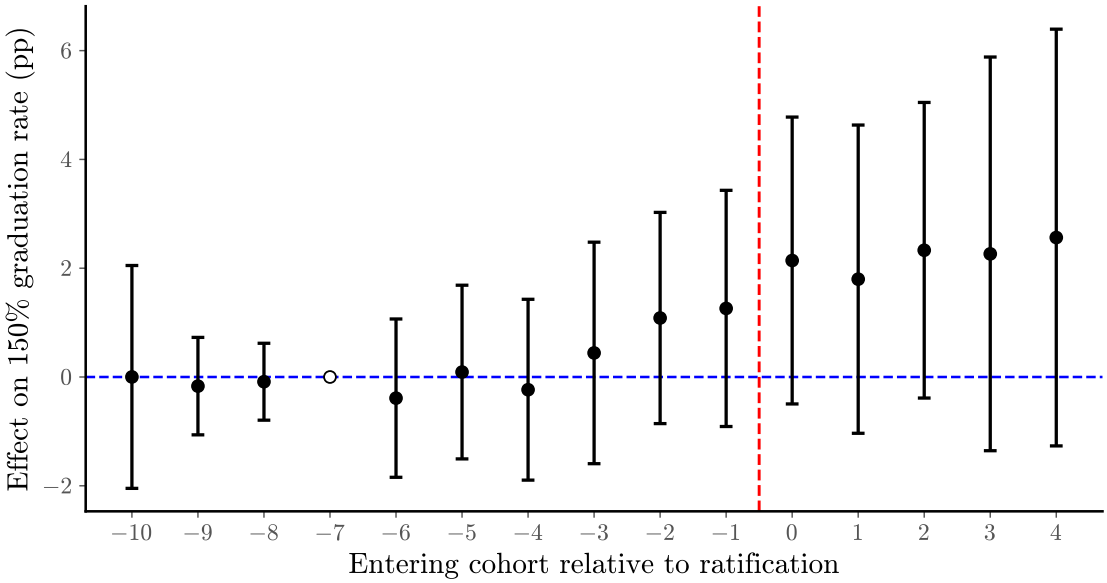
5 Where the University Adjusts: Field by Field

Stipends rise by similar amounts in every field, but doctorates fall far more outside science and engineering: 18.8 percent against 9.6 percent. We estimate the same design on institution×field cells. For institution i, field k and stack c,

$$y_{ikct} = \alpha_{ikc} + \gamma_{kct} + \beta_k^{\text{full}} D_{ict}^{[6,+]} + \beta_k^{\text{part}} D_{ict}^{[0,5]} + \beta_k^{\text{camp}} D_{ict}^{[-4,-1]} + \varepsilon_{ikct}. \tag{4}$$

Each field k has its own coefficients and its own field×cohort×year effects γ_{kct}, so a treated cell is compared with control cells in the same field and year. The panel is zero-filled, so a program that stops awarding doctorates stays in the risk set. Fields are the SED broad fields, split at the NSF science-and-engineering line, which leaves education, the humanities and arts, and other non-S&E fields outside.

Figure 16: Undergraduate Graduation Rates by Entering Cohort



Notes: Stacked event-study coefficients on the 150%-time graduation rate (IPEDS GR), in percentage points, with 95 percent intervals, by entering cohort relative to the ratification year; hollow marker is the reference cohort. R1 comparison group, clustered by bargaining unit. Treated institutions: 7, 6 of them observed at $\tau \geq 6$.

5.1 Stipends Rise by Similar Amounts Across Fields

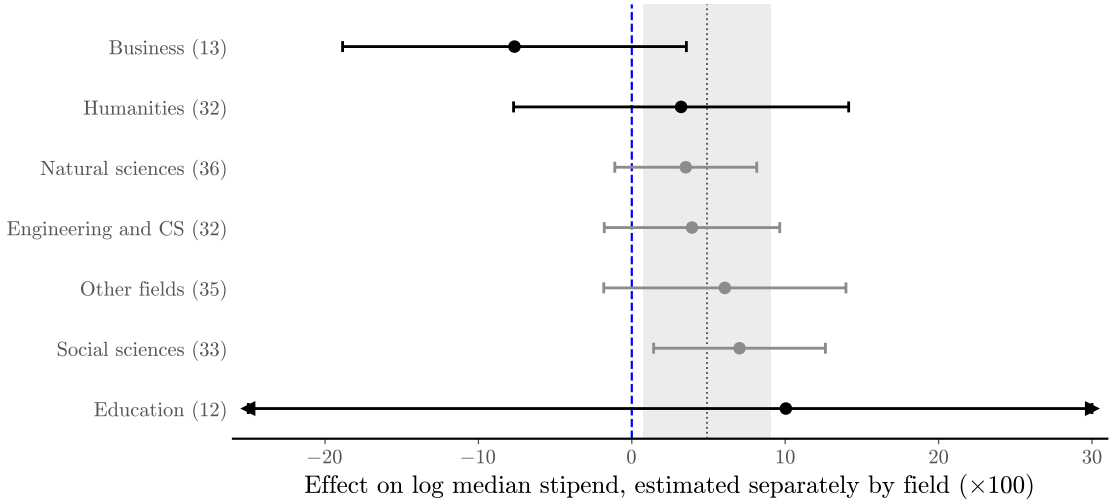
In every field, the stipend gain is close to the pooled gain of about 5 percent, and every field's interval covers it (Figure 17). We estimate each field separately on its own cells, so its institution effects are school×field effects. The gains run from about 3 percent to 7 percent across fields with at least 32 treated universities. Business is the one negative point estimate, and education the least precise, but neither is distinguishable from the pooled gain. Compared within a university, the social sciences gain less than engineering and computer science, and fields that started low gain no more than the rest (Appendix C).

Union wage standards usually raise pay most where it started lowest (Freeman, 1980; Card, 1996), as faculty unions do within Canadian universities (Baker et al., 2024). Graduate contracts do not detectably compress stipends. Within a school, a field that started below the school's own pre-period median gains no more than the others. Schenk (2010) likewise finds no differential return by field across institutions.

5.2 The Cut Is Deepest in Education and the Humanities

The doctorate decline is deepest in education and the humanities (Figure 18; Table A5). In SED, education falls 24.8 percent and the humanities and arts 17.4 percent. Psychology

Figure 17: Stipend Gains by Field, Estimated Separately



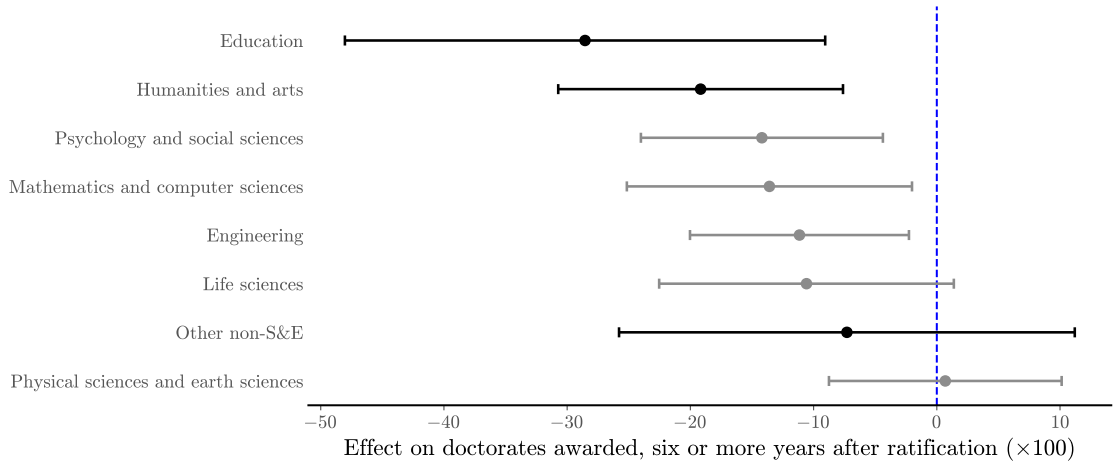
Notes: Post-ratification coefficients on log median stipends, times 100, with 95 percent intervals, estimated separately for each field bucket on its own institution×year cells with institution and year fixed effects, weighted by reports, clustered by bargaining unit; treated institutions in parentheses; intervals clipped at the axis limits are marked with arrows. The dotted line and gray band are the pooled estimate on the same cells and its interval.

and the social sciences, mathematics and computer sciences, engineering and the life sciences fall 10 to 13 percent, and other non-S&E fields 7.0 percent. The physical and earth sciences do not detectably change. Outside science and engineering doctorates fall 18.8 percent, against 9.6 percent inside, a ratio of 2.1 in SED and from 1.1 to 2.1 across the two censuses and the two treatments of the health professions. The decline in science and engineering is still distinguishable from zero. Part of the gap comes from New York University: without it, the SED ratio is 1.7.

No field's pre-campaign years differ detectably from its controls (Table A5), though the life sciences and education, the largest decline, come closest ($p = .13$ and .16). Appendix D reports moderator tests, Rambachan–Roth bounds by field, and whether the campaign-window movements of several non-S&E fields predate the campaign.

Per unit of cost, universities cut doctorates most in education and least in the physical sciences (Figure 19). Each field's elasticity divides its doctorate response by the same cost shock (equation (3)). One shock serves every field because stipend gains do not differ detectably by field and the contracts do not price benefits by field. Outside science and engineering the elasticity is 1.9, with a 95 percent interval from 0.7 to 3.0. It is highest in education, at 2.5, and the humanities, at 1.7. In the physical sciences it is not distinguishable from zero, and the other five fields lie between 0.6 and 1.3. Pooling all fields on the same field panel gives 1.2, a little above the 1.08 of Section 4.3, which divides the institution-level

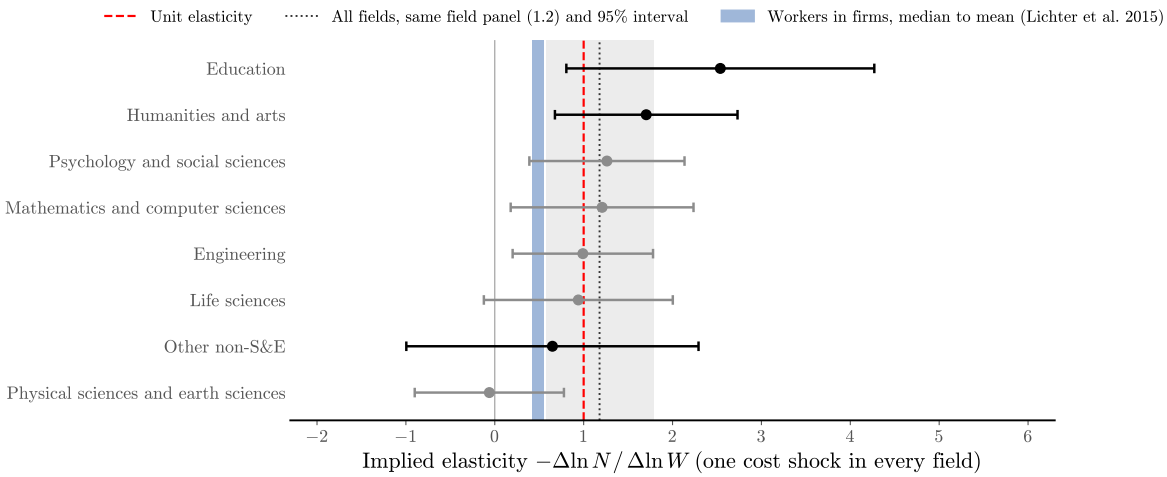
Figure 18: Doctorates Awarded by Field, Six or More Years After Ratification (SED)



Notes: Fully treated coefficients from Table A5 Panel B, times 100, with 95 percent intervals, fields sorted by the size of the coefficient. Black markers are the three broad fields the NSF convention places outside science and engineering; gray markers are the five science fields. Health-professions doctorates are inside "Life sciences" and cannot be separated in SED.

decline instead. The field intervals are wide, so we read the ordering, not the magnitudes.

Figure 19: Implied Elasticity of Doctorates Awarded with Respect to Cost, by Field

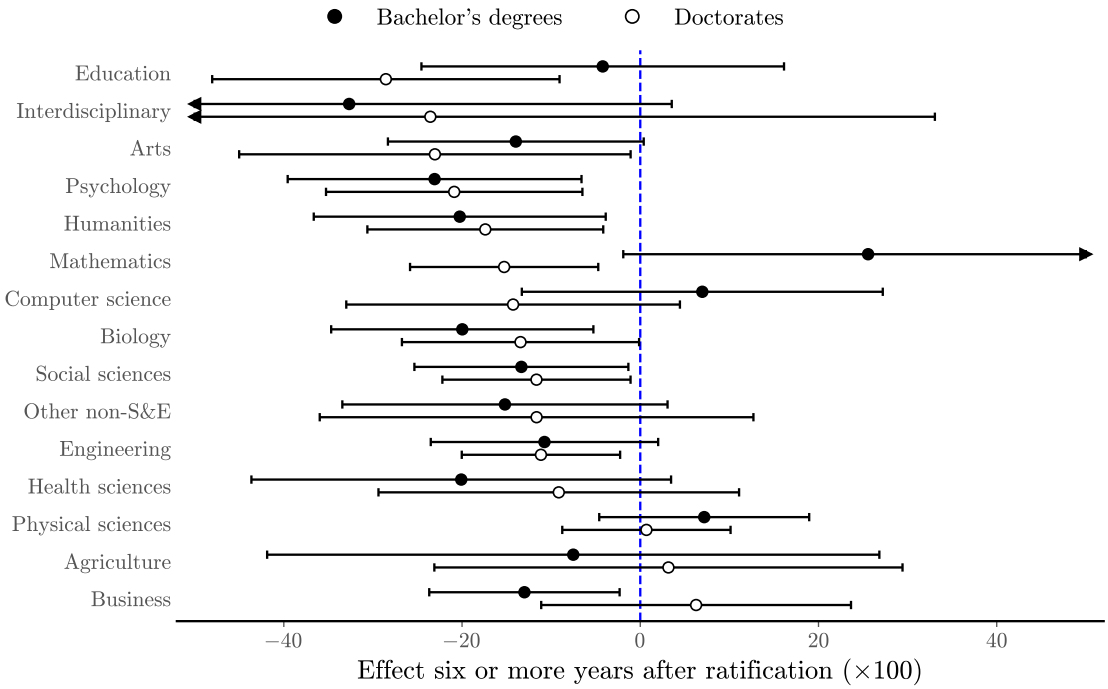


Notes: Fully treated coefficient on doctorates awarded in each SED field, pooled across contract cohorts, from the field panel of Table A5, divided by the continuation-adjusted cost shock of Table 2 and sign-reversed, with the coefficient's 95 percent interval over the same fixed shock; intervals clipped at the axis limits are marked with arrows. The red dashed line marks an elasticity of one; the dotted line and gray band are the all-field coefficient of the same field panel over the same shock, and its interval; the blue band runs from the median to the mean labor-demand elasticity in the meta-analysis of Lichter et al. (2015). Black markers are fields outside science and engineering.

5.3 Bachelor's Degrees by Field

Within a university and field, doctorates and bachelor's degrees fall by amounts we cannot tell apart, but across fields the two orderings agree only loosely (Figure 20). We map IPEDS first majors by CIP code to the fifteen fields of the SED's current taxonomy and estimate both on the same field stacks. The doctorate response minus the bachelor's response is not distinguishable from zero six or more years after ratification. Across fields the rank correlation is 0.29, which is not distinguishable from zero either. Psychology and the humanities lose among the most of both, and the physical sciences lose neither. Education loses the most doctorates but few bachelor's degrees, and the health sciences the reverse. Mathematics' bachelor's degrees rise, but their pre-campaign test rejects. As with the decline in Section 4.9, we cannot establish that these field patterns are effects of the contract.

Figure 20: Bachelor's Degrees and Doctorates by Field, Six or More Years After Ratification

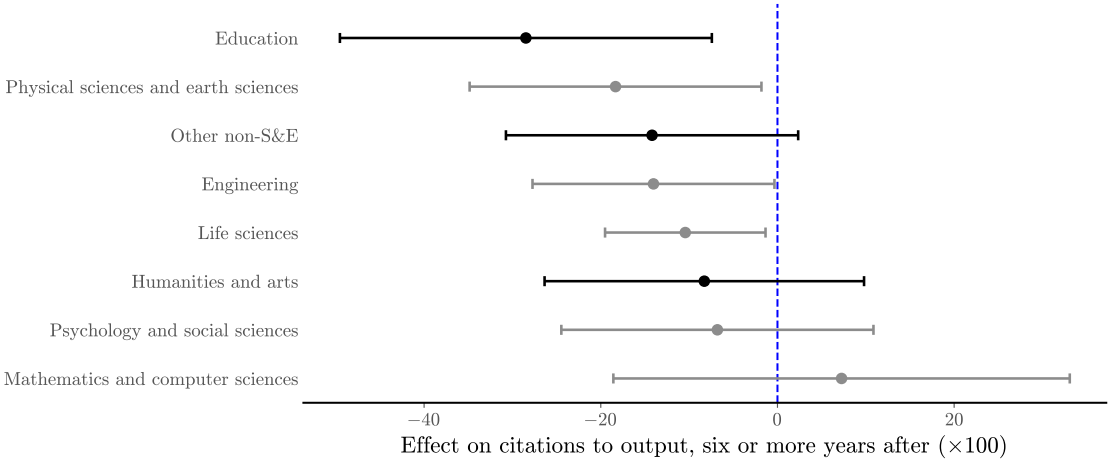


Notes: Coefficients at $\tau \geq 6$, times 100, with 95 percent intervals, fields ordered by the doctorate coefficient; intervals clipped at the axis limits are marked with arrows. Filled markers: bachelor's degrees (IPEDS award level 5, first majors, mapped to fields by CIP code), each field estimated alone on the SED institution stacks crossed with the fields, with stack×institution and cohort×year effects, terms for $\tau - 10$ to $\tau - 6$, the campaign window, $\tau \in [0, 5]$ and $\tau \geq 6$ terms, reference $\tau - 5$. Hollow markers: doctorates on the fifteen-field doctorate stacks, estimated as the eight-field coefficients of Table A5. Poisson, R1 comparison group, clustered by bargaining unit.

5.4 Research Output Falls in Most Fields, but Not in Step with Doctorates

Research output falls on both sides of the science line, by similar amounts, although doctorates fall more outside it (Figure 21). Like the university-level research measures, these field estimates compare treated universities with R1 and R2 universities. Six or more years after ratification, citations to a year's output are about 15 percent lower outside science and engineering and about 10 percent lower inside it. Across the eight SED fields, a field's citation decline does not track its doctorate decline. The rank correlation between the two, with doctorates also measured on R1 and R2 universities, is −0.26 and not distinguishable from zero, though sampling error in both series pulls it toward zero. The physical and earth sciences, where doctorates do not fall, show one of the largest citation declines, and the humanities, where doctorates fall steeply, a decline not distinguishable from zero. Unlike the university-level results, these field estimates have no pre-campaign test.

Figure 21: Citations to Institutions' Output by Field, Six or More Years After Ratification



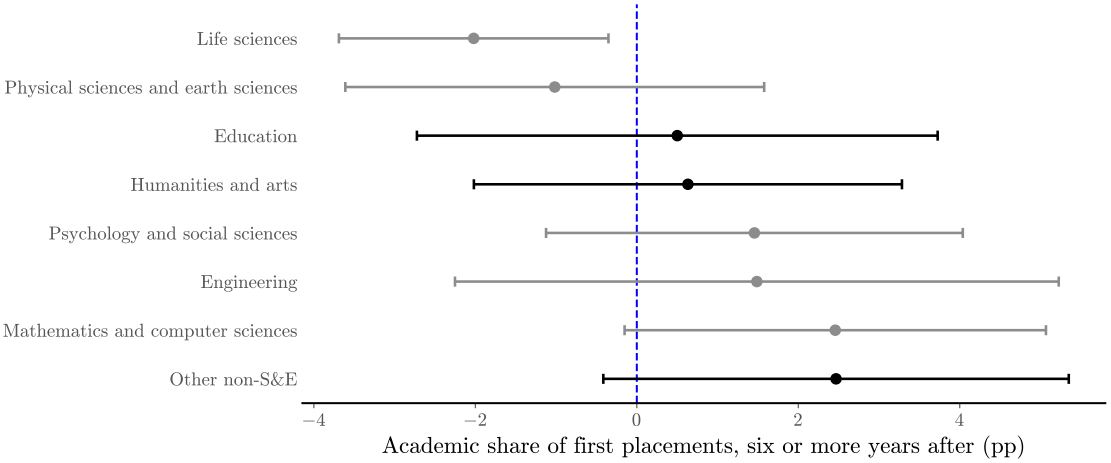
Notes: Fully treated coefficients on log citations to all works an institution publishes in a year within each SED field, times 100, with 95 percent intervals, on the institution×field×year panel with cell×stack and field×cohort×year fixed effects, R1 and R2 comparison group, clustered by bargaining unit. Black markers are fields outside science and engineering.

5.5 The Graduates, by Field

Where the contraction is deepest, the graduates still produced do not detectably lose academic placements or publish less (Figures 22 and 23). Outside science and engineering, six or more years after ratification, neither the academic share of named first placements nor the share who publish after the degree falls detectably, and each rules out a decline of

more than about 1 point. The academic share falls in one field, the life sciences, by 2.0 points. In no other field is it distinguishable from zero, and the publishing share does not detectably fall in any field (Appendix Table A11). The point estimates outside science and engineering are slightly positive, as selection toward stronger students would produce. When humanities departments in the Mellon Foundation's Graduate Education Initiative admitted smaller cohorts, their entrants had higher verbal GRE scores and left their programs less often (Groen et al., 2008).

Figure 22: Academic Share of Graduates' First Placements by Field



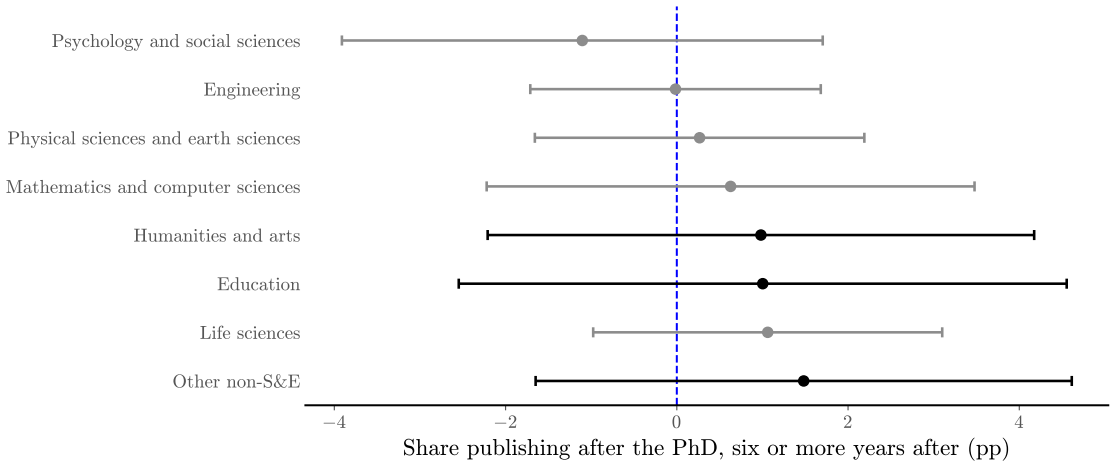
Notes: Fully treated coefficients ($\tau \geq 6$) on the academic share of named first placements within each SED field, in percentage points, with 95 percent intervals, on the institution×field×cohort panel with design as in Figure 13. Black markers are fields outside science and engineering.

5.6 Three Field Traits Line Up with the Ordering

Three traits of a field line up with the ordering of its doctorate response, and each is a different reason a university would cut that field's doctorates. These tests feed the model of Appendix E, so like it they use R1 and R2 universities (Section 3.1). On R1 universities alone every correlation below has the same sign and is smaller, and none is distinguishable from zero at the 5 percent level. The ordering is already visible in the campaign window, before ratification, and no field's pre-campaign years show it, so it begins with the campaign (Appendix D). The first trait, self-funding, is how much students pay their own way. Ranked by the share of students who support themselves from their own resources (SED), the eight fields contract in a similar order, a rank correlation of -0.667 with exact $p = .042$.

Across the eight fields the self-funded share nearly coincides with the science-and-engineering line, so it could merely mark that line. Within the five science and engineering fields it

Figure 23: Share of Graduates Publishing After the Doctorate, by Field



Notes: Fully treated coefficients ($\tau \geq 6$) on the share of a graduating cohort with any publication after the degree, within each SED field, in percentage points, with 95 percent intervals; design as in Figure 22.

orders the contraction exactly, and so does the second trait, outside money. National Science Foundation Graduate Research Fellowships arrive with the student, and no contract reprices them. The 2005–2024 fellowship rate per doctorate awarded also ranks the five science fields' contraction exactly, and all eight fields' at 0.76. Within science, then, the contraction is ordered by how much of a student's cost comes from outside the university, and the two measures cannot be told apart.

The third trait is what a field's undergraduate alumni are worth to the university, through what they later give. Fields whose majors earn more in mid-career (American Community Survey, full-time workers aged 40 to 54) lose fewer doctorates, a rank correlation of 0.74, though not detectably within science. The three traits move together. Across the fifteen fields of Section 5.3, fellowships and majors' earnings correlate at +0.85, and the self-funded share correlates with them at −0.61 and −0.69, so rank tests on one trait at a time cannot say which reason is at work. Appendix E gives the first trait a mechanism, a contract's rise in the tuition self-funded students pay, and separates the three within the model.

6 How Long a University Takes to Settle Lines Up with Its Response

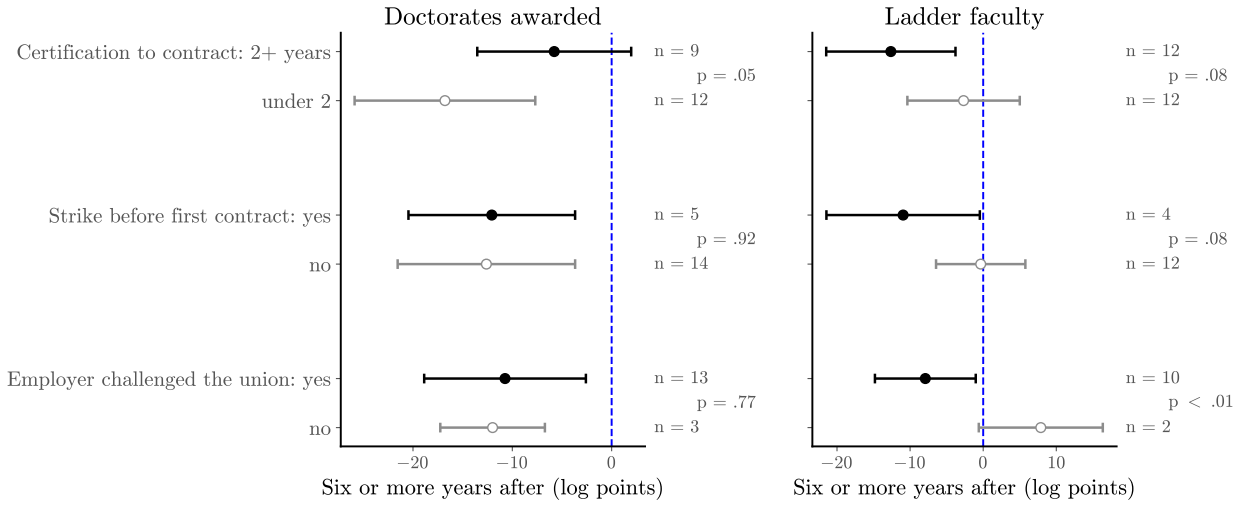
How long a university took to reach a first contract lines up with its doctorate response, and how it fought the union does not. We split the fully treated coefficient by how each university met its union. Some recognized a union and signed a contract within a year.

Others contested the unit before a labor board or in court, and at some the union struck before a contract. We coded each path from primary sources, as it stood at ratification (Appendix B.3). How a university meets its union is its own choice, so these splits describe its behavior rather than identify what contesting a union causes.

Where certification took two years or more to reach a first contract, doctorates fall less six or more years after ratification. They fall about 6 percent at these long-lag universities, against about 15 percent where the contract came sooner (Figure 24, on 9 and 12 universities observed that late). Part of the gap is mechanical. At a long-lag university the years between certification and ratification fall in the campaign window, so a response that began at certification is estimated there, not in the fully treated term. Giving each group its own campaign-window coefficient moves the two declines to about 8 and 16 percent, which narrows the gap without closing it. Ladder faculty point the other way, falling more at long-lag universities, but that difference is not distinguishable from zero.

Neither a strike nor an employer challenge makes a detectable difference to doctorates. Where the union struck before its first contract, doctorates fall about as much as elsewhere. With only 5 struck universities observed six or more years out, though, the comparison could miss a large difference. Only the time from certification to contract lines up with the size of the doctorate response, and part of that gap is mechanical.

Figure 24: Doctorates and Ladder Faculty by the University's Response to Its Union



Notes: Fully treated coefficients ($\tau \geq 6$), in log points, with the treatment terms split by the stated measure: SED doctorates on the all-cohort stacks and IPEDS ladder faculty on the department stacks, Poisson with stack and cohort×year fixed effects, R1 controls, 95 percent intervals clustered by bargaining unit. Filled points: the institutions with the stated characteristic; open points: the rest. n counts the treated institutions observed at $\tau \geq 6$ in each group; p tests the equality of the two coefficients. Treated institutions with the measure unknown are dropped. Measures are defined in Appendix B.3.

7 Conclusion

Graduate student unions do what unions do: they raise the price of covered labor, by about 6 percent net of local cost-of-living growth. Universities then buy less of it. Once the first cohorts admitted under a contract reach graduation, they award about 11 percent fewer doctorates, a response about 2.6 times as elastic as the median for workers in firms. They fund fewer research assistants in science and engineering, employ fewer ladder faculty and publish less in total. The doctorate response is 1.1 to 2.1 times as deep outside science and engineering, though the size of that gap rests partly on one large university. Output per faculty member, the placement and publication records of the doctorates still produced, and undergraduate graduation rates show no detectable change, and each interval excludes large effects. First-year retention dips slightly. Bachelor's degrees and entering classes also fall at treated universities, but they were already falling before the campaign, so we cannot attribute those declines to the contract. The outcomes we can bound are also the ones through which the voice face of unionism (Freeman and Medoff, 1979, 1984) would reach our data, and none detectably improves. Voice can still work through advising, working conditions and well-being, which these data do not record.

These results are not, on their own, a case against graduate unions. A first contract raises what a university pays for a funded doctoral place, and the university then chooses how many places to keep. That choice depends on who pays. Stipends rise by similar amounts across fields, yet doctorates fall least in the physical sciences, where grants and outside fellowships pay much of a student's stipend, and most in education and the humanities, where the university and the students themselves do. Grant support that paid graduate students at contract rates could offset part of the response, but such support is thinnest in the fields that respond most. Before a contract, graduate workers bore part of the cost of training and research through low stipends, and a contract shifts that cost to the university. Whether the scaling back that follows should be offset depends on how much a doctorate is worth beyond what the student and the university capture.

References

Ashenfelter, O. C., Farber, H., and Ransom, M. R. (2010). Labor market monopsony. *Journal of Labor Economics*, 28(2):203–210.

Baker, M., Halberstam, Y., Kroft, K., Mas, A., and Messacar, D. (2024). The impact of unions on wages in the public sector: Evidence from higher education. Working Paper 32277, National Bureau of Economic Research. Revised February 2025.

Bergström, P., Dahlberg, M., and Mörk, E. (2004). The effects of grants and wages on municipal labour demand. *Labour Economics*, 11(3):315–334.

Bradley, D., Kim, I., and Tian, X. (2017). Do unions affect innovation? *Management Science*, 63(7):2251–2271.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3):414–427.

Card, D. (1996). The effect of unions on the structure of wages: A longitudinal analysis. *Econometrica*, 64(4):957–979.

Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, 134(3):1405–1454.

Desrochers, D. M. and Hurlburt, S. (2016). Trends in college spending: 2003–2013. Report, Delta Cost Project at American Institutes for Research.

DiNardo, J. and Lee, D. S. (2004). Economic impacts of new unionization on private sector employers: 1984–2001. *Quarterly Journal of Economics*, 119(4):1383–1441.

Dube, A. and Zipperer, B. (2024). Own-wage elasticity: Quantifying the impact of minimum wages on employment. Working Paper 32925, National Bureau of Economic Research.

Ehrenberg, R. G. (1973). The demand for state and local government employees. *American Economic Review*, 63(3):366–379.

Frandsen, B. R. (2016). The effects of collective bargaining rights on public employee compensation: Evidence from teachers, firefighters, and police. *ILR Review*, 69(1):84–112.

Frandsen, B. R. (2021). The surprising impacts of unionization: Evidence from matched employer-employee data. *Journal of Labor Economics*, 39(4):861–894.

Freeman, R. B. (1980). Unionism and the dispersion of wages. *Industrial and Labor Relations Review*, 34(1):3–23.

Freeman, R. B. and Medoff, J. L. (1979). The two faces of unionism. *The Public Interest*, 57:69–93.

Freeman, R. B. and Medoff, J. L. (1984). *What Do Unions Do?* Basic Books, New York.

Ganguli, I. and Murciano-Goroff, R. (2025). Exposure to science and scientific careers: Evidence from minimum wage increases and university lab employment. Working Paper 34244, National Bureau of Economic Research.

Groen, J. A., Jakubson, G. H., Ehrenberg, R. G., Condie, S., and Liu, A. Y. (2008). Program design and student outcomes in graduate education. *Economics of Education Review*, 27(2):111–124.

Hamermesh, D. S. (2014). Do labor costs affect companies' demand for labor? *IZA World of Labor*, 2014(3).

Harasztosi, P. and Lindner, A. (2019). Who pays for the minimum wage? *American Economic Review*, 109(8):2693–2727.

Hausman, N. (2022). University innovation and local economic growth. *Review of Economics and Statistics*, 104(4):718–735.

Herbert, W., Apkarian, J., and van der Naald, J. (2024). Directory of bargaining agents and contracts in institutions of higher education. Technical report, National Center for the Study of Collective Bargaining in Higher Education and the Professions, Hunter College, CUNY.

Hirsch, B. T. (2004). What do unions do for economic performance? *Journal of Labor Research*, 25(3):415–455.

Julius, D. J. (2023). The current status of graduate student unions: An employer's perspective. *Journal of Collective Bargaining in the Academy*, 14(1). Article 15. Republished in the volume *Learning From the Past to Enhance Our Future*; the underlying census of graduate student locals is current as of November 1999.

Kantor, S. and Whalley, A. (2014). Knowledge spillovers from research universities: Evidence from endowment value shocks. *Review of Economics and Statistics*, 96(1):171–188.

Krueger, A. B. and Mas, A. (2004). Strikes, scabs, and tread separations: Labor strife and the production of defective Bridgestone/Firestone tires. *Journal of Political Economy*, 112(2):253–289.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, 76(3):1071–1102.

Lee, D. S. and Mas, A. (2012). Long-run impacts of unions on firms: New evidence from financial markets, 1961–1999. *Quarterly Journal of Economics*, 127(1):333–378.

Lichter, A., Peichl, A., and Siegloch, S. (2015). The own-wage elasticity of labor demand: A meta-regression analysis. *European Economic Review*, 80:94–119.

Liu, A. Y.-H. (2005). Union and union threat premiums among graduate student stipends. Cornell Higher Education Research Institute Working Paper, SSRN #869485.

Lovenheim, M. F. (2009). The effect of teachers' unions on education production: Evidence from union election certifications in three midwestern states. *Journal of Labor Economics*, 27(4):525–587.

Lovenheim, M. F. and Willén, A. (2019). The long-run effects of teacher collective bargaining. *American Economic Journal: Economic Policy*, 11(3):292–324.

MacKinnon, J. G., Nielsen, M. Ø., and Webb, M. D. (2023). Cluster-robust inference: A guide to empirical practice. *Journal of Econometrics*, 232(2):272–299.

MacKinnon, J. G. and Webb, M. D. (2018). The wild bootstrap for few (treated) clusters. *The Econometrics Journal*, 21(2):114–135.

MacKinnon, J. G. and Webb, M. D. (2020). Randomization inference for difference-in-differences with few treated clusters. *Journal of Econometrics*, 218(2):435–450.

Mas, A. (2008). Labour unrest and the quality of production: Evidence from the construction equipment resale market. *Review of Economic Studies*, 75(1):229–258.

Paglayan, A. S. (2019). Public-sector unions and the size of government. *American Journal of Political Science*, 63(1):21–36.

Rambachan, A. and Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, 90(5):2555–2591.

Rogers, S. E., Eaton, A. E., and Voos, P. B. (2013). Effects of unionization on graduate student employees: Faculty–student relations, academic freedom, and pay. *ILR Review*, 66(2):487–510.

Schenk, T. (2010). The effects of graduate-student unionization on stipends. Iowa State University Working Paper, SSRN #1831975.

Sojourner, A. J., Frandsen, B. R., Town, R. J., Grabowski, D. C., and Chen, M. M. (2015). Impacts of unionization on quality and productivity: Regression discontinuity evidence from nursing homes. *ILR Review*, 68(4):771–806.

Staiger, D. O., Spetz, J., and Phibbs, C. S. (2010). Is there monopsony in the labor market? evidence from a natural experiment. *Journal of Labor Economics*, 28(2):211–236.

Stephan, P. E. (1996). The economics of science. *Journal of Economic Literature*, 34(3):1199–1235.

Toivanen, O. and Väänänen, L. (2016). Education and invention. *Review of Economics and Statistics*, 98(2):382–396.

Valero, A. and Van Reenen, J. (2019). The economic impact of universities: Evidence from across the globe. *Economics of Education Review*, 68:53–67.

Walters, C. R. (2024). Empirical Bayes methods in labor economics. Working Paper 33091, National Bureau of Economic Research.

Wang, S. and Young, S. (2023). Unionization, employer opposition, and establishment closure. Working Paper 23-35, U.S. Census Bureau, Center for Economic Studies.

Winston, G. C. (1999). Subsidies, hierarchy and peers: The awkward economics of higher education. *Journal of Economic Perspectives*, 13(1):13–36.

A Identifying Assumptions and Partial Exposure

This appendix states the identifying assumptions behind equations (1), (2) and (5), and how the entering-cohort outcomes treat partial exposure. The specification of each other exercise sits in the appendix that reports it.

A.1 Identifying Assumptions

Five conditions let the coefficients of equation (1) be read as effects of the contract. We test three of them, estimate around a fourth rather than assume it, and sign the bias from the fifth.

Assumption 1 (parallel trends conditional on stack). Treated and never-treated institutions would have evolved in parallel absent unionization. Table A14 reports joint Wald tests on the pre-campaign years, $\tau - 10$ through $\tau - 6$, against a $\tau - 5$ reference. The organizing campaign gets its own column, because the design estimates campaign responses rather than assuming them away (Assumption 2). A test that also spanned $\tau - 4$ through $\tau - 1$ would ask whether a coefficient we expect to be non-zero is zero. SED doctorates pass the test ($p = .81$), as do research assistants, first-time enrollment and the full-time enrollment stock. In IPEDS the all-field panel passes ($p = .49$) and the research-doctorate panel nearly rejects ($p = .054$).

Three outcomes fail the pre-campaign test, and none carries a claim of its own. Part-time enrollment is not used. Teaching-assistant counts reject ($p = .002$) on a drift of $+1.19$ percent a year, so they serve only as a comparison for research assistants, whose test passes. Fellowships reject ($p < .001$) on a drift of -1.75 percent a year. Appendix F uses them only to ask whether research assistants were relabeled as fellows. The drift alone could produce the fall reported there, so that check also rests on total funded positions, which fall too.

Assumption 2 (no anticipation, replaced rather than assumed). We estimate the years of partial exposure rather than assume there is no anticipation. Labor disputes give reason to expect a response before the contract. Tire defects at Bridgestone/Firestone's Decatur plant rose when the company demanded concessions, months before the strike (Krueger and Mas, 2004). Mas (2008) dates the resale discount on Caterpillar equipment to the vintage built as the contract expired, and Lee and Mas (2012) allow a short anticipation window before the election. A graduate campaign and certification come before ratification, and degree counts mix cohorts admitted before and after it. So the specifications carry separate terms for the campaign window, $\tau - 4$ through $\tau - 1$, for the years 0 to 5 in which degrees go to students admitted before the contract, and, for undergraduate outcomes, for

the cohorts enrolled when it arrived. The main coefficient compares fully treated years with fully untreated ones, and the partly treated terms are reported rather than folded into either side. The stipend panel begins in 2012, too late for a campaign window, so its event study measures against the year before ratification.

Assumption 3 (SUTVA). Spillovers to control universities can bias our estimates in either direction. Universities do raise pay to head off a union: Liu (2005) finds teaching-assistant stipends higher where students had organized but no union was yet recognized. A control that raised stipends for the same reason would move toward treatment, so the stipend estimate would be too small. The raise would also move the control's costs toward treatment and should shrink its doctoral production, so the doctorate estimate would understate the decline. Applicants push the other way. Students a treated university turns away may enroll at a control instead. That raises the control's doctorates, so the doctorate estimate overstates the decline at treated universities. Research spillovers also run both ways. A control that responds as treated universities do pushes the research estimates toward zero, so the research bounds are less informative and the assistantship declines of Section 4.4 understate the employment response. Faculty who leave treated universities for controls push the faculty and research declines the other way.

Assumption 4 (correct treatment classification). The sign and timing of the contraction, and the sign of the stipend rise, hold across treatment classifications, but the magnitudes move. Under two codings of the joint units the SED estimate is not distinguishable from zero, and under two codings the stipend rise is significant only at the 10 percent level (Section 4.2, Appendix B, Figure A3). Classification errors need not only attenuate. Lovenheim (2009) finds that the Census of Governments union indicator misclassifies a substantial share of school districts, with errors correlated with the outcomes. It yields a significant union pay effect where hand-collected certification dates, less precisely, yield none. So we validated our roster against the Hunter directory (Herbert et al., 2024). We corrected the roster's dates, removed advocacy organizations without bargaining rights, and moved certified institutions we had missed from the controls to treatment (Appendix B.2). The directory lists some public systems as one row, and we mapped those rows to campuses by hand, so the mapping cannot be reproduced from the directory alone.

Assumption 5 (no differential regional trends). Treated universities are concentrated geographically: about 86 percent sit in Democratic-leaning states. An outcome that tracked wage, cost-of-living or admissions trends in those states would pass for a union effect. We check the stipend estimate against metropolitan cost-of-living growth, and the doctorate estimate against year effects specific to Democratic-leaning states and census regions

(Appendix H.2, Table A12). The stronger answer is timing, since a regional trend would have to break when each university ratified (Appendix H.2).

A.2 Entering-Cohort Outcomes and Partial Exposure

Section 4.9 follows undergraduates by entering cohort and drops the partly exposed cohorts entering in $[c-6, c-1]$, who are enrolled when the contract arrives:

$$y_{iec} \;=\; \alpha_{ic} + \gamma_{ce} + \beta^{\text{full}}\,\mathbf{1}\{e \geq c\} + \varepsilon_{iec}, \qquad e \notin [c-6,\, c-1] \text{ for treated } i, \tag{5}$$

where e indexes entering cohorts.

On entering-cohort clocks, an apparent pre-trend in graduation rates or alumni earnings can be partial exposure, since cohorts that entered before ratification were still enrolled when the contract arrived. So in equation (5), β^{full} compares fully exposed entering cohorts with cohorts that graduated before ratification, and the event studies keep every cohort with the partly exposed ones marked. Lovenheim and Willén (2019), with 33 treated states, instead use partly exposed birth cohorts as dose variation. With three to eight treated universities a slope across the partly exposed cohorts would be imprecise, and it would rest on an assumed mix of how long their students take to finish or leave. So we drop those cohorts and compare fully exposed cohorts with unexposed ones.

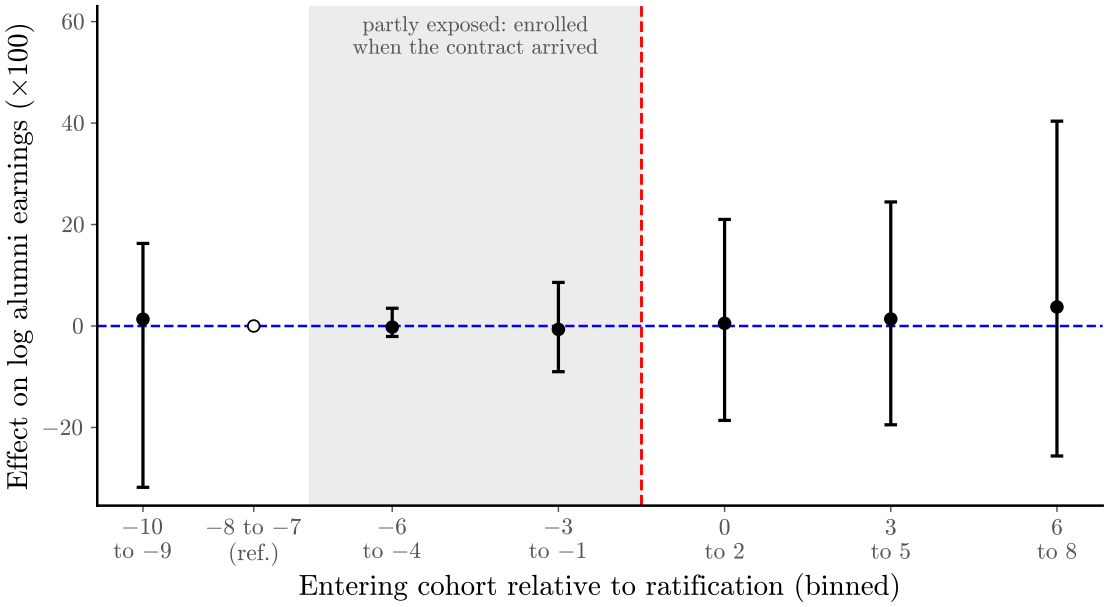
Table A1 reports the full-exposure contrasts for retention, graduation and alumni earnings. Alumni earnings rest on 3 treated universities. By entering-cohort bin their bootstrap intervals are too wide to rule out large effects of either sign, and no bin, before or after ratification, is distinguishable from zero (Figure A1).

B Setting, Roster and Data

B.1 Certification and Ratification Clocks

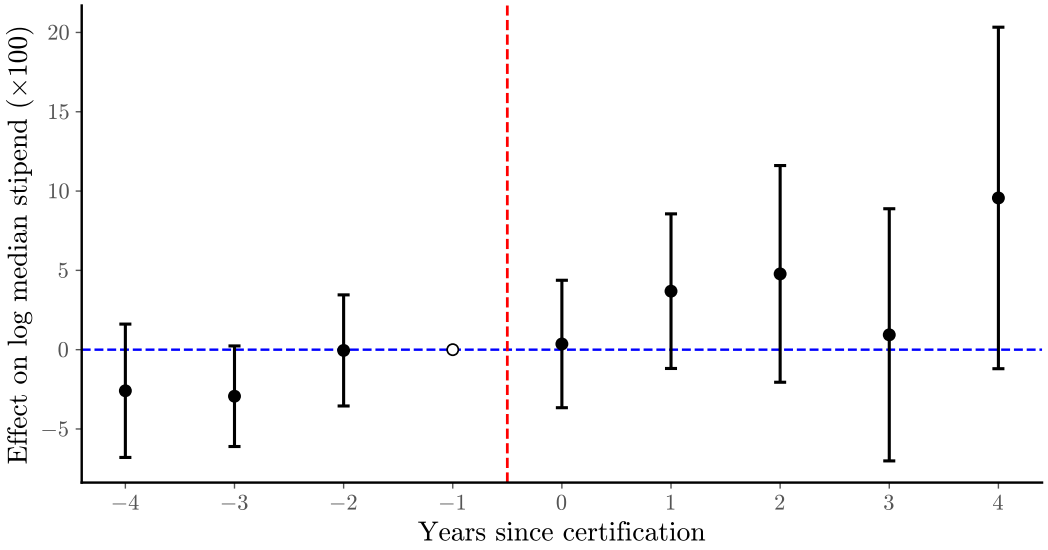
Ratification is the first date at which the stipend floor, the funding guarantees and the grievance procedure all bind, so cost responses date to it. Certification creates a duty to bargain but sets no terms of pay. The establishment literature dates treatment to certification because its designs use close elections. Not every certification produces a contract: DiNardo and Lee (2004) put the share that ever does at about 55 percent, and 6 certified graduate units in our roster have none yet. A design that treated every certified unit would estimate an intent-to-treat effect diluted by those units. We define the treated set by ratification instead, so the certification clock in Table A4 changes only the date (Figure A2).

Figure A1: Alumni Earnings by Entering Cohort



Notes: Coefficients on log median earnings of federally aided students (College Scorecard, pooled post-entry horizons 6–11), times 100, by bins of entering cohort relative to ratification, with null-imposed wild-cluster-bootstrap 95 percent intervals (Webb weights, 9,999 replications); reference bin $[-8, -7]$; the gray band marks cohorts enrolled when the contract arrived. Same data, fixed effects, weights and control pool as the pooled estimate in Table A1; 3 treated institutions.

Figure A2: PhD Stipends: Certification Clock



Notes: As Figure 4, but with event time measured from certification.

Table A1: Undergraduate Outcomes: Full-Exposure Cohorts vs. Pre-Union Cohorts

Outcome	Estimate	95% bootstrap interval	WCB p	RI-t p	Treated	Pre-period mean
Panel A. Levels (contract clock)						
First-year retention (pp)	-3.035^{*}	[−6.11, −0.05]	0.045	—	8	93.3
	(0.968)					
150%-time graduation rate (pp)	−0.000	[−5.49, 5.66]	1.000	—	7	69.5
	(2.118)					
log alumni earnings (pooled)	0.013	[−0.23, 0.23]	0.702	—	3	49,307
	(0.021)					
Panel B. Alumni earnings by parental-income tercile						
INC1 (∼ \$0–30k)	0.074	[−0.25, 0.43]	0.157	0.222	3	57,101
	(0.032)					
INC2 (∼ \$30–75k)	0.008	[−0.22, 0.28]	0.756	—	3	54,261
	(0.025)					
INC3 (\$75k+)	−0.019	[−0.23, 0.16]	0.462	0.640	3	57,500
	(0.023)					

Notes: Pre-period mean: the outcome over treated institutions' pre-ratification cohorts in the estimation sample, weighted as the regression, in natural units (retention and graduation in percent, earnings in dollars). Fully-treated entering cohorts against cohorts that completed before ratification, contract clock, partially treated cohorts absorbed. Retention and the graduation rate are reported in percentage points, the earnings rows in log points. Earnings are median earnings of federally aided students (College Scorecard), pooled across the six published post-entry horizons P6–P11, count-weighted. The control group is restricted to never-organized R1 universities (2015 Carnegie classification). Standard errors clustered by bargaining unit in parentheses. WCB = null-imposed wild cluster bootstrap (Webb weights, 9,999 replications); RI-t = studentized randomization inference (placebo ratifications assigned to random control institutions, 499 draws), reported for the rows where it was run. Placebos are drawn from all control institutions; drawing only from the largest quarter of them gives p-values that differ from these by between −0.038 and +0.022 and leaves each on the same side of .10. Stars reflect WCB p-values. The final column counts treated institutions in each row's estimation sample. Every row here identifies from at most 8 treated institutions and the earnings rows from 3, too few for cluster-robust standard errors (Appendix H.4). $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{\dagger}p < .10$.

B.2 Union Events and Roster Construction

The roster records each institution's union status by year. Certification and first-contract dates start from the Hunter directory (Herbert et al., 2024), corrected and extended with ratifications verified from union and university sources through 2026. The replication files list every treated institution with its certification year, first-contract year, sector, state, and the analyses it enters.

Read row by row, the directory would place some of the country's largest doctorate producers among the never-organized controls, because it lists each public system (the University of California, the State University of New York, the California State University) as a single row. So we build the roster down from system rows to campuses and correct errors from parsing the directory and matching its names to institutions. The replication files keep a versioned record of every change, with its source and its reason. The roster carries 140 organized institutions, of which 66 have a dated first contract.

The treatment rule must decide on two kinds of unit. The first is a first contract that covered teaching assistants only, with research assistants added in a later contract or never. The second is a joint unit of graduate and undergraduate student employees. Two of these are the California Academic Student Employees unit of 2000 and the California State University unit of 2005, which the directory itself types as covering undergraduates. The others are at The New School, Harvard, Columbia, Washington State and Western Washington, whose recognition clauses name undergraduate teaching assistants, graders, tutors or student workers. Coverage is coded for all 75 certified or contracted units, from the first contract's recognition clause (31 units), a later contract's (26), the certification order (9), or a union or university statement or news report (9). Each unit's code carries a confidence grade (53 high, 20 medium, 2 low).

Our rule treats every institution at the first ratified contract of the bargaining unit that represents its graduate workers today, whatever share of research assistants that contract covered, and holds out the joint graduate-undergraduate units. A joint contract prices undergraduate student labor along with graduate labor, so it need not make a graduate student dearer than the undergraduates who could do some of the same work. Being organized, the joint units are also out of every control pool, and they appear on neither clock. A rule applied to large systems can move an estimate weighted by doctorates, so Figure A3 gives both doctorate censuses under four codings, and the replication files carry the rest. Holding out first contracts that excluded research assistants gives -0.137 (0.043) in SED, and the reported estimate is neither the largest nor the smallest. Treating the joint units as treated moves the estimates most: SED to -0.058 (0.047) against -0.122 (0.035), and IPEDS to -0.049 (0.042) against -0.097 (0.041). The University of California campuses,

in the largest joint unit, are among the biggest doctorate producers in the roster, and a count estimator weights institutions by their doctorates.

New York University ratified the first graduate contract at a private university in 2002, withdrew recognition in 2005 after *Brown* (2004), and signed a new first contract in 2015, the date we use. Dating it to 2002 instead gives −0.124 (0.035) in SED (Table A12). On the certification clock it is dated from the December 2013 election behind that contract.

Table A2: From the Roster to the Estimation Samples

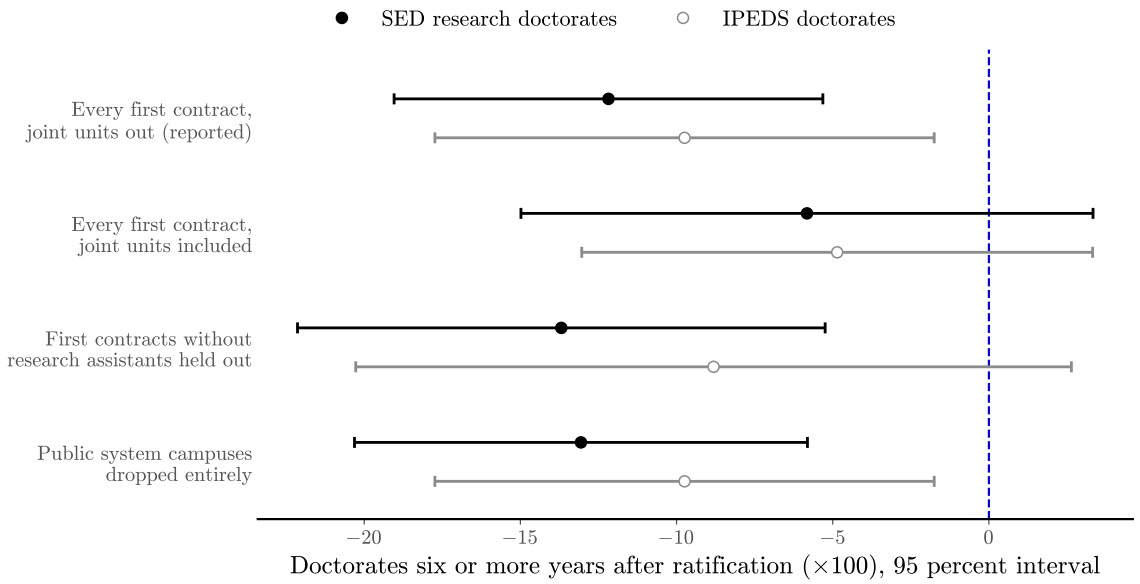
A. The roster: organized institutions by status			
All organized institutions	140		
First contract ratified, date known: treated	66		
Joint graduate-undergraduate unit: on neither clock, out of every control pool	38		
First contract ratified, date unknown: out of both groups	6		
Certified, bargaining a first contract	4		
Certified, no contract	2		
Certification nullified	1		
Advocacy organization without certification	23		
B. Estimation samples: institutions			
Outcome (source)	Treated	Fully treated	Control
Stipends (net of metro prices) (phdstipends.com)	19	19	35
Stipends, pooled table (phdstipends.com)	37	36	37
Doctorates awarded (SED)	34	21	43
Doctorates awarded, by field (SED)	37	21	43
Doctorates awarded (IPEDS)	26	13	43
Doctorates, research fields (IPEDS)	26	13	43
Research and teaching assistants (NSF GSS)	25	21	43
Institution research output (OpenAlex)	47	28	114
Ladder faculty (IPEDS)	47	24	114
Graduating cohorts (ProQuest, OpenAlex)	22	18	43
First-year retention (IPEDS)	8	1	43
Six-year graduation rate (IPEDS)	7	6	42
Alumni earnings (College Scorecard)	3		

Notes: Panel A counts every organized institution on the roster by status (Appendix B). Panel B counts the institutions in each estimation sample, read from each fitted main specification: treated institutions, those observed in the fully treated years ($\tau \geq 6$ on the ratification clock; $\tau \geq 3$ for assistantships; any year after ratification for stipends), and controls. Treated and control universities are R1 universities (2015 Carnegie classification), except in the institution research output and ladder faculty rows, which compare R1 and R2 universities on both sides (Section 4.5). The main stipend sample treats contracts ratified 2014–2025, the ratifications its 2012–2025 panel observes; the pooled stipend table counts every contracted institution with a first-contract year. Alumni earnings: treated count of the pooled-cohort estimate.

Treated counts differ across analyses through data coverage, exposure timing and the comparison group, since every analysis uses the same roster and event clocks. Outcomes dated by calendar year need years before the campaign, and outcomes dated by entering cohort need cohorts fully exposed to the contract. Most analyses compare R1 universities with R1 universities, while the institution research measures and the model inputs add R2 universities on both sides. Counts run from 47 for institution research output to 3 for alumni earnings (Table A2). Alumni earnings are observed at their post-entry horizons only

for cohorts that entered well before the post-2013 contracts, which is why so few institutions identify them.

Figure A3: Doctoral Production Under Every Coding of the Treatment Roster



Notes: Fully treated coefficients ($\tau \geq 6$) with 95 percent intervals, from the specification of Section 4.2: zero-filled PPML on the stacked institution panel, stack×unit and cohort×year fixed effects, partially treated years 0–5 absorbed, R1 controls, clustered by bargaining unit. Each row re-estimates it on a differently coded treatment roster. The reported coding treats every institution at its first ratified contract except bargaining units that also cover undergraduate student employees, which are organized, held out of every control pool and placed on neither clock. Leave-one-cohort-out under the reported coding runs from −0.107 to −0.142 in SED. Appendix B.2 defines each coding.

B.3 Coding Each University's Response to Its Union

Section 6 splits the fully treated coefficient by how each university met its union. For each of the 66 dated first contracts we coded the path from certification to ratification from primary sources, as it stood at ratification. The lag is the ratification year minus the certification year (median 1 year, and 32 took two years or more). A strike before the first contract (9 of the 55 where it is known) counts only for the campaign that produced the ratified contract. The recognition strikes of earlier, failed campaigns at Yale and Chicago are therefore coded no. A no that rests only on a state ban on public-employee strikes and on silence in the sources is coded unknown. An employer challenge (28 of 51) contests the unit or employee status before a labor board or a court. Sources are NLRB case records, state labor-board orders, court opinions, the contracts' recognition clauses and campus newspapers. The SUNY campuses, which bargain under one contract, are one unit.

B.4 Outcome Data

Doctoral production. Degrees by institution and field from IPEDS Completions (1984–2024; research-doctorate award levels, 15 broad fields, with CIP detail) and the NCSES Survey of Earned Doctorates (1958–2024; eight broad fields; research doctorates only). IPEDS also counts practice doctorates such as the EdD and the DNP, so SED is the estimate of record (Section 4.2). The two censuses agree in the aggregate, and by field IPEDS shows a smaller gap between fields outside and inside science and engineering (Table A5).

Graduate support structure. The NSF–NCSES Survey of Graduate Students and Postdoctorates (GSS), 1972–2024, by institution and discipline. It counts graduate students by type of support (research and teaching assistantships and fellowships) and by source of support, including students who support themselves, along with enrollment stocks and first-time students. It covers science, engineering and health only, so it is silent in the fields where doctorates fall most.

Stipends. A full extraction of phdstipends.com (17,958 crowd-sourced observations, dense 2014–2025), collapsed to institution × field × year medians and matched to IPEDS identifiers (88 percent match rate). Reporting is voluntary and unionization is a salient event. Appendix C.1 therefore checks the measure against administrative payroll from the University of California Office of the President annual wage disclosure (2012–2024, 10 campuses, by title and campus), used as salary-band counts rather than individual records.

Undergraduate outcomes. First-year retention (IPEDS EF-D, 2002–2023), 150%-time graduation rates by entering cohort (IPEDS GR, cohorts 1991–2018), median earnings of federally aided students 6 to 11 years after entry (College Scorecard, pooled entry cohorts 1997–2012 plus 2014), overall and by FAFSA family-income tercile.

Placements and research records. A ProQuest corpus of 3.75 million doctoral dissertations. The 2.07 million from US institutions with degrees from 1970 onward are matched to OpenAlex author profiles and, in science and engineering, to PatentsView inventors. For each graduate the match gives the first post-PhD employer named in the publication record, its sector under the SED convention, and whether any publication follows the degree. Institutions are keyed to IPEDS through the corpus's own crosswalk (114 of the 140 organized institutions, 98.6 percent of US records). For universities that stopped depositing with ProQuest we add dissertations from 7 institutional repositories, matched by ORCID where the record carries one. The panels cover the degree cohorts of 1970–2021, because a graduate is observed only once they publish and the observed share falls mechanically for recent cohorts. Cohorts are years of degree, so $\tau \geq 6$ dates the first contract-admitted graduating cohort here as in the doctorate estimates. Against hand-coded truth the match names the correct institution in about 55 percent of cases and

the correct sector in about 70. We did not check accuracy separately at treated and control institutions. If errors are the same at both, they pull the placement contrasts toward zero, so the placement bounds understate how large an effect they could hide.

Institution research output. Research output comes from the OpenAlex works file. Each work is dated by its own publication year and carries its citations and, for articles and reviews, OpenAlex's normalized citation percentile within field, year and type. Linked to the graduates, these give articles and their citation rank in fixed windows from the degree. Linked to the institutions on their bylines, they give an institution×field×year panel (1980–2024) of works, citations, authors and citation rank. It is the paper's only research measure that covers fields outside science and engineering. Output per head divides by ladder faculty from the IPEDS Faculty Salaries survey, a headcount from outside the publication record.

C Stipends: Measurement and Field Detail

C.1 Validating the Stipend Measure Against Administrative Payroll

phdstipends.com is crowd-sourced, so unionization could change who reports rather than what they are paid. A ratification is a salient public event, and it mobilizes exactly the pay-transparency activism that drives submissions to such a site. Cells are weighted by report count, so a treatment-induced surge in reports from any one department would move the estimate mechanically. Appendix C.2 tests for such a surge directly.

The University of California allows a direct test of the measure. California discloses the gross annual pay of every UC employee by title and campus. The Graduate Student Researcher (GSR) unit ratified its first contract in December 2022, and Academic Student Employees at the same campuses have been unionized since 2000. We can therefore compare the crowd-sourced measure with administrative payroll on the same campus-years.[1]

The crowd-sourced measure is accurate in levels (Table A3). Across 109 campus-years its median sits within 2 percent of the administrative median. The ratio of the two is 0.99 before 2023 and 0.95 after, so if anything the measure understates pay after the contract, which would bias the stipend estimate toward zero.[2] UC submissions per year also fell, from

[1]Data are drawn from the UC Office of the President annual wage disclosure. The interface caps any result set at 2,899 records and any page at 20 rows, so we query salary bands and record only counts; no individual records enter the analysis. UC reports no graduate-student titles before 2012, which bounds the panel at 2012–2024.

[2]The log correlation across campus-years is 0.47, so individual cells are noisy even though the central tendency is right, which argues for weighting cells by report count as we do rather than treating thin cells as equally informative.

154 in 2019–2022 to 42 in 2023–2025. If a contract changes reporting at UC, it suppresses reports rather than drawing them in from students who gained.

A within-UC difference-in-differences, comparing the newly unionized research titles with the long-unionized teaching titles, cannot pin down a magnitude, because UC publishes gross annual pay, which mixes appointment fractions. The exercise therefore validates the measure rather than giving a second estimate of the effect.

Payroll for other public universities is scarce, because most states leave student employees out of their salary portals. Beyond UC, we found usable graduate-assistant pay for one treated campus, the University of Connecticut, whose first contract dates from 2015, and for 7 never-contracted public campuses in Georgia, Iowa, Arizona and Indiana. At Connecticut the ratio of the crowd-sourced median to payroll is 0.97 before the contract, on only 4 reports, and 0.99 after. At the controls it rises over the same years, from 0.85 in 2014–2018 to 0.96 in 2019–2024. At these campuses the crowd-sourced median sits below payroll by more than at UC, by up to 17 percent, so the close match in levels is a UC result. Against the controls' rise, with campus and year effects, the ratio at Connecticut and UC together falls 12.1 log points after their contracts (s.e. 4.4). With campus effects alone it moves −4.3 (s.e. 3.3). So outside UC too, if the measure drifts after a contract, it understates pay, which would bias the stipend estimate toward zero. The size of that drift rests on one treated campus beyond UC and on how the controls' own rise is modeled.

Table A3: UC Administrative Payroll and the phdstipends Measure

	Ratio	IQR	Campus-years
Panel A: phdstipends against administrative payroll			
Median across campus-years	0.98	[0.87, 1.04]	109
Campus-years before 2023	0.99		
Campus-years 2023 onward	0.95		
Correlation of logs	0.47		
Panel B: phdstipends submissions from UC campuses, per year			
2019–2022	154		
2023–2025	42		

Notes: University of California Office of the President annual wage disclosure, 2012–2024, ten campuses; UC reports no graduate-student titles before 2012. Pay statistics are percentiles interpolated from salary-band counts, so no individual records enter the analysis. Panel A compares the phdstipends campus-year median to the administrative median conditional on gross pay above $20,000, across campus-years observed in both sources. Panel B counts phdstipends reports from UC campuses before and after the December 2022 contract for Graduate Student Researchers.

C.2 Who Reports Before and After Ratification

Reports to phdstipends.com dip in the ratification year and the year after, but the field mix of reports does not move and the stipend estimate does not rest on those two years. The check matters because if a contract changes who reports, the stipend coefficient mixes a change in pay with a change in reporters. We put report counts through the stipend design of Section 4.1: institution-year counts, zero-filled over 2012–2025, on the 19 treated universities and the never-organized controls of the stipend sample, by Poisson with institution and year effects and bargaining-unit clusters. After ratification, reports average 19.0 log points fewer than before. In the ratification year and the year after they are −23.1 and −24.9 log points lower, each significant only at the 10 percent level, and from the second year on they are not distinguishable from their earlier level. The share of reports from the humanities, education and the social sciences does not detectably change. The stipend gain is +6.3 log points as reported, +8.4 without those two years, and +5.8 with every cell weighted equally rather than by its reports, so it comes neither from the years of the dip nor from heavily reported cells.

C.3 Cross-Field Stipend Differentials and Within-School Compression

Within an institution, stipend gains differ little by field, and buckets that started low do not gain detectably more (Table A4, Panels B and C). The one exception is the social sciences, which gain 6.2 points less than engineering and computer science on the ratification clock ($p = .026$). Let i index institution, k field bucket and t year, and let y_{ikt} be the log median stipend in the cell. Regressions weight cells by their number of reports and cluster by bargaining unit. Panel B is estimated on each clock in turn,

$$y_{ikt} \;=\; \alpha_i \;+\; \gamma_{kt} \;+\; \beta^{\text{post}} D_{it}^{[0,+]} \;+\; \sum_{k' \neq k_0} \delta_{k'} \left(D_{it}^{[0,+]} \times \mathbf{1}\{k = k'\} \right) \;+\; \varepsilon_{ikt}, \tag{6}$$

where k_0, engineering and computer science, is the omitted bucket. Panel C replaces the field interactions with a single indicator for a *low-start* bucket, one whose *pre-period* median pay sits below the school's own pre-period cross-bucket median. It reports β^{post} for high-start buckets together with the low-start interaction.

Identifying assumption. Field-by-year fixed effects γ_{kt} absorb national pay trends specific to a discipline, so $\delta_{k'}$ is identified from the contrast between fields *within* the same institution and year. This is weaker than Assumption 1, since a shock that raises all stipends at a treated institution differences out. A shock that falls on one field at a treated institution, such as a large new grant program in a single discipline, would bias $\delta_{k'}$ only if its timing were correlated

Table A4: The Effect of Unionization on PhD Stipends

	Certification clock		Ratification clock	
	(1)	(2)	(3)	(4)
Panel A: Post-event level effect				
Unionized × post	0.031†	0.032*	0.055**	0.054*
	(0.016)	(0.015)	(0.020)	(0.021)
Panel B: Field heterogeneity (base: engineering/CS)				
Unionized × post		0.045*		0.076**
		(0.020)		(0.026)
× natural sciences		0.011		0.006
		(0.020)		(0.020)
× social sciences		-0.046†		-0.062*
		(0.027)		(0.028)
× humanities		-0.000		-0.042
		(0.041)		(0.040)
× education		0.055		0.039
		(0.076)		(0.074)
× business		-0.056		-0.051
		(0.051)		(0.050)
Panel C: Within-school compression				
Unionized × post (high-start fields)		0.028		0.044*
		(0.019)		(0.022)
× low-start field		0.022		0.040
		(0.028)		(0.027)
School FE	Yes	Yes	Yes	Yes
Year FE	Yes	—	Yes	—
Field × year FE	No	Yes	No	Yes
Treated schools	37	37	37	37
Pre-ratification stipend, treated ($)		33,500		

Notes: Pre-ratification stipend: the median at treated institutions in the four years before ratification. Outcome: log median stipend in institution×field-bucket×year cells (phdstipends.com, 2012–2025), weighted by cell report counts. Treated schools are those that ratified a first contract; columns (1)–(2) date treatment from certification and columns (3)–(4) from ratification. Panel B interacts the post indicator with field buckets against an engineering and computer-science base. Panel C interacts it with an indicator for buckets whose own pre-event median pay sat below that school's pre-event median across buckets, a classification fixed before treatment. Standard errors clustered by bargaining unit in parentheses. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{\dagger}p < .10$.

with ratification.

The low-start classification is fixed before treatment. A contract that compresses stipends would mechanically change which buckets sit at the bottom, so a classification measured over the full sample would be an outcome rather than a covariate. Section 5.1 discusses the pattern.

D Doctorates by Field: Tests and Detail

The field panel of Section 5 is the institution-by-field counterpart of the institution-level estimate, on a wider frame. It has 37 treated institutions against the institution panel's 34, because it needs only ten pre-years per institution-field cell, and its pooled coefficient is a little larger, -0.133 against -0.122. Fields where many doctoral students pay their own way contract most, and the sciences funded by outside grants contract least (the rank tests below). On the same design and window, the non-S&E contraction is 2.1 times the S&E one in SED, 1.8 times in IPEDS with health-professions doctorates counted inside S&E, as SED counts them, and 1.1 times under the NSF convention that places them outside.

Two shares from SED Table 4-1 rank the contraction, and they point the same way. One is the share of a field's doctoral students who support themselves, whom the university neither pays nor bargains over. The other is the share holding an assistantship, which a contract covers. Ranking the eight broad fields by each share and correlating the ranks with the fully treated coefficients on R1 and R2 universities, the sample of the model in Appendix E, the self-funded share gives $\rho = -0.667$ (exact one-sided $p = .042$ across all 40,320 orderings) and the assistantship share $\rho = +0.786$ ($p = .014$). So the fields whose students a contract covers most contract least. The model of Appendix E predicts an ordering rather than a magnitude, so a rank test matches it. It assumes no functional form and needs no asymptotics, which suits eight fields.

The second trait, outside money, ranks the fields too. The outside-money measure of Section 5.6 is National Science Foundation Graduate Research Fellowships per SED doctorate in the same field. We took every award offer NSF lists, 78,807 from 1952 to 2026, mapped each award's field of study to the SED broad fields, and divided by doctorates over the same years. Over 2005 to 2024, the years with consistent field labels, the rate runs from 62 per thousand doctorates in the physical and earth sciences to 0.5 in the humanities and 3.5 in education. It gives $\rho = +1.00$ ($p = .008$) within science and $+0.76$ ($p = .018$) across all fields. Longer windows, which reach into years with different field labels, give weaker correlations: $+0.62$ ($p = .058$) from 1990 and $+0.48$ ($p = .121$) from 1958.

As a continuous moderator on R1 and R2 universities, the self-funded share is not

distinguishable from zero. Ten points of self-funded share deepen the fully treated coefficient by -0.016 (s.e. 0.019, $p = .42$), and assistantship exposure moves it the other way, by $+0.015$ (0.013, $p = .24$). The gradient is already there in the campaign window, at -0.016 (0.008), as large as after ratification, so the change between the two is $+0.000$ (0.022). On R1 alone the post-period gradient is -0.033 ($p = .08$), about twice the campaign-window one (-0.017), but the change, -0.016 (0.021), is not distinguishable from zero. The rank tests above use the same fully treated coefficients, so they share this timing. A fixed field trait would produce such a gradient if fields had been diverging before the campaign, but no field's pre-campaign years differ detectably from its controls (Table A5). So the field ordering begins with the campaign, a response Assumption 2 counts as part of the effect.

Several non-S&E fields fall during the campaign window, and no field's pre-campaign years show a decline already underway. Humanities and arts is -7.7 log points over $\tau \in [-4, -1]$ ($p = .006$) and the non-S&E aggregate is -7.6 ($p = .003$). These are roughly 40 and 36 percent of the respective post-ratification effects. Education's campaign coefficient is -12.7 log points ($p = .007$). The S&E aggregate's campaign window is flat ($p = .35$), as are those of engineering and of mathematics and computer sciences. Assumption 2 allows movement here, but not through admissions: students admitted during the campaign graduate after ratification, so a campaign-window decline in degrees must come from students already enrolled, who finish less often or later. A decline underway before any campaign would show up in the pre-campaign years, $\tau - 10$ through $\tau - 6$. For the non-S&E aggregate those years are jointly indistinguishable from zero ($p = .426$), and so are those for humanities and arts ($p = .527$), whose campaign coefficient is among the largest. The life sciences ($p = .135$) and education ($p = .163$) pass by the smallest margins of any field.

The sensitivity analysis of Rambachan and Roth (2023) measures how large a violation of parallel trends the field estimates survive. Taking the pre-campaign years as the benchmark and counting the campaign window as part of the response, the SED non-S&E contraction survives post-ratification violations up to $\bar{M} = 1.00$ times the largest observed before the campaign. The S&E estimate survives up to 1.25 and the pooled field-level estimate up to 1.50. Counting the campaign window as a violation instead changes both the benchmark and the reference year. It leaves the non-S&E and pooled values unchanged and raises the S&E value to 1.75. The IPEDS non-S&E estimate is not distinguishable from zero even with no violation, under either benchmark ($\bar{M} = 0.00$).

Table A5: Completions by Field: Selective Contraction

	Fully treated ($\tau \geq 6$)	Implied change	Pre-campaign p	Campaign p	Treated (panel / $\tau \geq 6$)	Pre-period mean
Panel A: IPEDS, 15 fields						
All fields (pooled)	-0.119**	-11.2%	[0.33]	[0.12]	32 / 16	
	(0.039)					
Non-S&E (health outside S&E)	-0.125	-11.8%	[0.32]	[0.16]	32 / 16	
	(0.076)					
S&E (health outside)	-0.117**	-11.0%	[0.16]	[0.46]	32 / 16	
	(0.038)					
Non-S&E (health inside S&E)¶	-0.169*	-15.6%	[0.47]	[0.14]	32 / 16	
	(0.081)					
S&E (health inside)¶	-0.095*	-9.1%	[0.59]	[0.31]	32 / 16	
	(0.044)					
Panel B: SED, 8 broad fields (health necessarily inside S&E)						
All fields (pooled)	-0.133***	-12.4%	[0.62]	[0.02]	37 / 21	42.4
	(0.035)					
Non-S&E (aggregate)	-0.209**	-18.8%	[0.43]	[<0.01]	37 / 21	30.4
	(0.065)					
S&E (aggregate)	-0.101**	-9.6%	[0.43]	[0.35]	37 / 21	49.6
	(0.034)					
Education	-0.285**	-24.8%	[0.16]	[0.01]	37 / 21	32.2
	(0.099)					
Humanities and arts	-0.192**	-17.4%	[0.53]	[0.01]	37 / 21	40.6
	(0.059)					
Psychology and social sciences	-0.142**	-13.2%	[0.95]	[0.05]	37 / 21	55.8
	(0.050)					
Mathematics and computer sciences	-0.136*	-12.7%	[0.49]	[0.12]	37 / 21	22.1
	(0.059)					
Engineering	-0.111*	-10.5%	[0.75]	[0.32]	37 / 21	54.4
	(0.045)					
Life sciences	-0.106†	-10.0%	[0.13]	[0.39]	37 / 21	75.8
	(0.061)					
Other non-S&E	-0.073	-7.0%	[0.45]	[0.77]	37 / 21	18.4
	(0.094)					
Physical sciences and earth sciences	0.007	0.7%	[0.31]	[0.16]	37 / 21	40.1
	(0.048)					

Notes: Zero-filled PPML on stacked institution×field panels, cell×stack and cohort×field×year FE. The coefficient is the fully-treated DiD ($\tau \geq 6$, partially treated years 0–5 absorbed), the SE clustered by bargaining unit beneath it. The campaign p tests this regression's $\tau \in [-4, -1]$ coefficient (reference $\tau \leq -5$). The pre-campaign p is a joint Wald test on separate terms for $\tau - 10$ through $\tau - 6$ against a $\tau - 5$ reference, in an event study on the same panel. Both panels share a design and window, so the censuses differ only in what they count. Panel A: IPEDS (32 treated, 43 controls, 19 cohort stacks, 15 fields). Panel B: SED (37 treated, 43 controls, 24 cohort stacks, 8 broad fields); fields sorted by the size of the contraction. The treated column reports institutions entering the panel and, after the slash, those reaching $\tau \geq 6$ and identifying the coefficient. Pre-period mean: SED doctorates per treated institution, field and year before ratification (Panel B). Panel B runs on a wider stack frame than the institution-level estimate of Section 4.2.

¶The NSF convention places health-professions doctorates outside S&E; SED cannot split them out of "Life sciences" and counts them as S&E. Panel A is estimated under both, so the difference between its two blocks is the reclassification alone. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{\dagger}p < .10$.

E Why Some Fields Contract More: A Model and Its Calibration

A single contract sets one minimum stipend and one raise, yet fields within a university cut their doctorates by different amounts. We model a university that funds some doctoral students and enrolls others who pay their own way, and calibrate it to the field responses of Appendix D. The model is static, so it describes the whole response, including the part that begins during the campaign. Appendix E.6 gives the assumptions and proofs, and Appendix E.7 the estimation.

E.1 The Provost and the Prospective Student

The provost chooses how many doctoral students each field funds, A_k as teaching assistants and B_k as research assistants, and the tuition T that unfunded students pay, to maximize the university's net financial resources,

$$\max_{\{A_k,\,B_k\}_k,\,T} \quad \Pi = \sum_k \Big[\, p_k^U \, H_k(A_k + \phi S_k) + T\, S_k - w_k A_k - (w_k - f_k)\, B_k - C_k(G_k) \Big], \qquad (7)$$

summed over fields k. Each funded student is paid the stipend w_k. Another S_k enroll without funding, and $G_k = S_k + A_k + B_k$ is doctoral enrollment. The field teaches $U_k = H_k(A_k + \phi S_k)$ undergraduates, an increasing and concave function of the teaching done, where ϕ is the share of a teaching assistant's teaching that an unfunded student does. Each undergraduate is worth $p_k^U = P_k(R) + \psi_k(R_k)$ in tuition and alumni gifts, which rise with the university's reputation $R = \sum_j \varsigma_j(Q_j)$ and the field's own reputation $R_k(Q_k)$. Both are built by research, $Q_k = B_k + \omega_A A_k + \omega_S S_k$, where ω_A and ω_S are the shares of a research assistant's research done by a teaching assistant and by an unfunded student. A research assistant brings grant support f_k. $C_k(G_k) = \kappa_k^0 G_k + \kappa_k^1 G_k^\nu$ is the program's cost beyond stipends, with economies of scale ($0 < \nu < 1$). Funded places $L_k = A_k + B_k$ are filled from applicants, so unfunded enrollment S_k depends on them through the admission cutoff below. The provost makes the two choices separately: funded places taking unfunded enrollment as given, and tuition taking funded places as given. An allocation is a fixed point of the two (Appendix E.6). The contract sets a minimum and a raise for students already above it, $w_{1k} = \max\{\underline{w}, (1+r)\, w_{0k}\}$, the most common rule in the coded agreements (17 of the 42 with a coded pay rule, against 13 with a pay schedule). It covers funded students only. In the model, unfunded students teach as part of their training, unpaid, which is why their teaching enters H_k at the share ϕ.

A prospective student in field k takes the best of a funded place, enrolling without funding

and the outside option,

$$\max\Big\{\ \underbrace{v+w_k}_{\text{funded, if admitted}}\ ,\ \underbrace{v-T_e+\theta\,w^{\text{out}}}_{\text{pay own way}},\ \underbrace{w^{\text{out}}}_{\text{outside}}\ \Big\},\qquad T_e=T-s^E T_0. \tag{8}$$

Here v is the yearly dollar value of the degree beyond its pay, and w^{out}, the outside option, is what the student would earn without enrolling. Across prospective students, (v, w^{out}) is jointly normal with means μ_v and μ_{out}, standard deviations σ_v and σ_{out}, and correlation $\rho_{v,\text{out}}$. T_e is tuition net of employer support, which covers a share s^E of pre-contract tuition T_0, and θ is the share of the outside option a student keeps by working while enrolled without funding. A student applies for funding only if the funded place is the best option, $v + w_k \geq w^{\text{out}}$ and $w_k \geq \theta w^{\text{out}} - T_e$. Funded places are rationed: programs admit the applicants with $v \geq v_k^*$, the cutoff that fills the L_k places. More students apply than there are places, so the university fills each place at the stipend. Students who pay their own way are either applicants who were turned down and pay anyway or direct choosers, who never apply because enrolling unfunded while keeping their outside earnings beats a funded place. Only a student who keeps earning while enrolled can be a direct chooser (Lemma 1). Education shows why the model needs them. If all its self-funded students were rejected applicants, 71 percent of the applicants it turned down would have paid their own way (93 percent counting employer support), against at most 12 percent in any other field, and 70 percent of its doctoral students work while enrolled, against 12 percent in the field where the fewest do.

E.2 What a Contract Does

We represent a contract as a rise in funded students' stipends, $\Delta w_k \geq 0$ in field k. The provost then funds fewer places. By revealed preference, funded places, weighted by each field's stipend increase, fall for given unfunded enrollment, however the fields' reputations interact (Appendix E.6). One field's places can grow only if others fall by more. In the model, teaching assistants fall in every field whose stipend rises, and research assistants fall unless the loss of shared reputation raises the value of research by more than the stipend rise.

In the model, the contract also moves unfunded enrollment through three channels. Fewer funded places turn away more applicants, and those who prefer the unfunded place to the outside option enroll without funding anyway (displacement). We call these students persisters. If tuition rises, it pushes out unfunded students (tuition). A higher stipend draws some who would have paid into the queue for a funded place (switching). Displacement raises

unfunded enrollment, and the other two lower it. Switching is zero in a field whose students keep none of their outside option while enrolled. Displacement creates the most persisters where many rejected applicants just below the cutoff value the unfunded place. The tuition channel pushes out the most students where many self-funded students sit at the payment margin, as in fields where most of them work while enrolled.

To first order, enrollment responds by

$$\Delta \ln G_k \approx \frac{Q_k^F}{G_k} \Delta \ln Q_k^F + (1 - \omega_A) \frac{A_k}{G_k} \Delta \ln A_k + s_k \Delta \ln S_k, \tag{9}$$

where Q_k^F is the research funded students do and s_k the self-funded share. A field's enrollment can rise only if the higher stipend draws new funded entrants from outside the program.

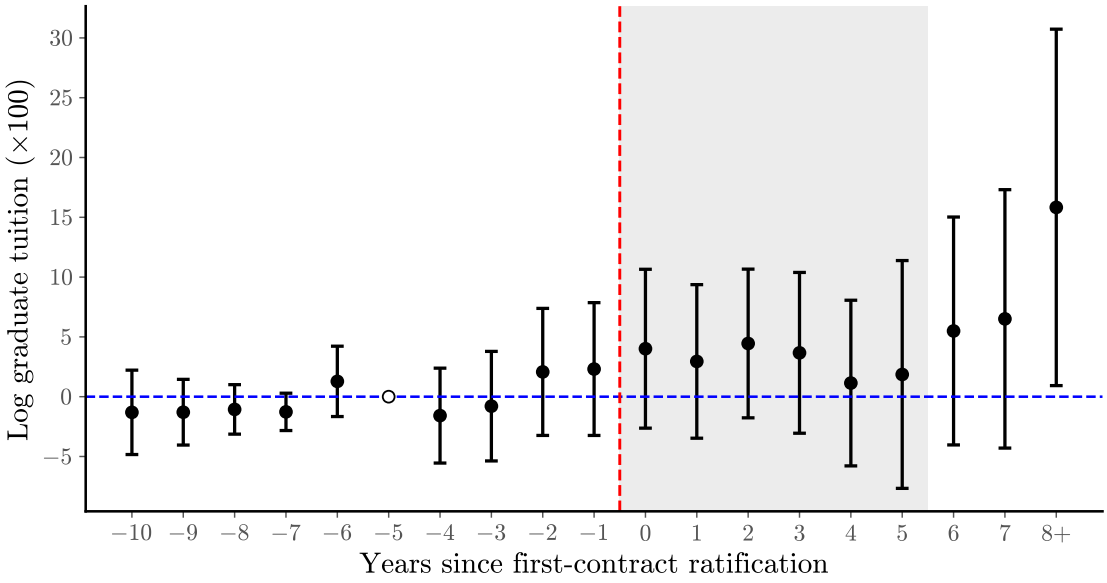
Because H_k is concave, undergraduates served change by a smaller percentage than the teaching input, $d \ln U_k = \eta_k^U \, d \ln(A_k + \phi S_k)$ with $\eta_k^U \in (0, 1)$. Where teaching input falls with doctoral enrollment, the model therefore has bachelor's degrees falling less than doctorates within a field. The difference estimated in Section 5.3 has the other sign but lies within a standard error.

E.3 Tuition Rises After a Contract

Universities charge more for graduate study after a first contract (Figure A4). Effective graduate tuition, the in-state and out-of-state rates weighted by the share of students from in state, is flat before ratification and only slightly higher in the first five years after it. Six or more years out it is about 8 percent higher. Like the other inputs to the model, this estimate uses R1 and R2 universities. On R1 universities alone the rise is similar but less precise, and so is the rise with separate year effects for public and private universities. In-state tuition, a narrower measure, rises less and is not distinguishable from zero.

The rise falls on students who pay their own way, and they cannot be counted where the contraction is deepest. The GSS counts self-supported students only in science, engineering and health. There, holding fixed each department's share of master's students, self-supported doctoral students change by -3.5 percent six or more years after ratification (s.e. 21.2) on R1 and R2 universities, and by -10.4 percent (s.e. 27.6) on R1 alone. Neither is distinguishable from zero, and the intervals are too wide to confirm or rule out the fall in self-funded enrollment the model implies.

Figure A4: Graduate Tuition Around First-Contract Ratification



Notes: Stacked event-study coefficients on log effective graduate tuition, times 100, with 95 percent intervals, against $\tau = -5$; $\tau \leq -10$ and $\tau \geq 8$ pooled. Effective tuition weights the in-state and out-of-state rates by the university's mean share of first-time undergraduates from in state. OLS with stack-by-university and cohort-by-year effects, standard errors clustered by bargaining unit; IPEDS Institutional Characteristics. Gray band: $\tau = 0$ to 5.

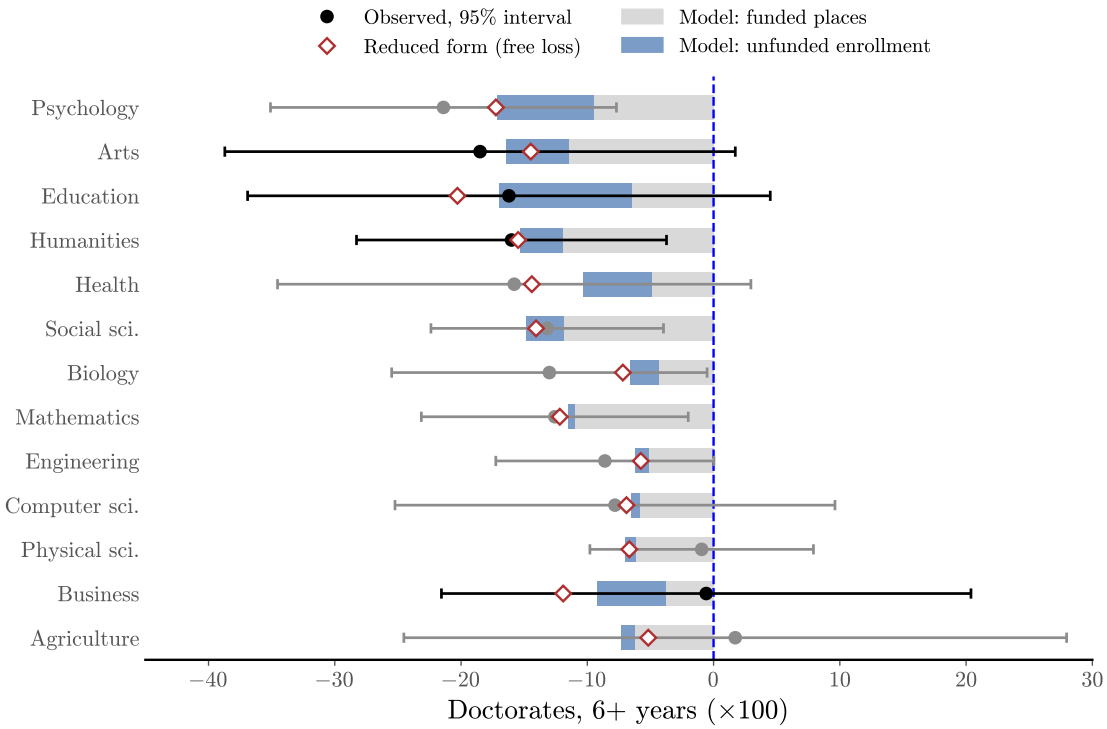
E.4 Calibration

Every input on the students' side comes from data other than the field responses, and each is a proxy for its quantity in the model. SED gives the self-funded share and employer support, measured among doctorate recipients rather than everyone enrolled. GradCafe decisions, scaled to CGS acceptance rates, give admission rates. The American Community Survey gives the outside option, as the earnings of bachelor's degree holders aged 23 to 30. Revelio employment histories of doctoral students give the share who work while enrolled, and IPEDS gives tuition and its response to the contract. The spread of the degree's value and its correlation with the outside option come from how admitted and rejected applicants on GradCafe differ in the mean and spread of their GRE scores. These gaps identify each field's pair only weakly, so each pair is shrunk by empirical Bayes toward the mean of similar fields (Walters, 2024). The stipend rise Δw_k comes from each treated contract's coded first-year minimum and raise against the university's pre-contract stipends in the field. On the funded side, one research curvature, one teaching curvature and a shared-reputation term are fitted to the field responses $\hat{\beta}_k$: the doctorate coefficients at $\tau \geq 6$ on the fifteen-field stacks, on R1 and R2 universities, for the 13 fields of Table A6.

The calibrated fit has a weighted R^2 of 0.55 and a rank correlation of 0.78 (Figure A5).

That is about as good as a single free parameter does: replacing the measured students' side with one share of self-funded students lost, fitted to the same responses, gives 0.57 and 0.89. Unfunded enrollment matters to the fit. With no response of it the fit falls to 0.30 and 0.52. Education is fitted at −17 log points against −16 observed, −11 of it through unfunded enrollment. At the estimates the spread of the degree's value is \$2,500 to \$15,000 a year across fields, against \$17,000 to \$49,000 for the outside option, so in the calibration the outside option does most of the sorting into applying, paying and leaving. That is a property of the estimates, not something the data show directly. The fit holds when the least certain inputs change: R^2 is 0.53 with education's admission rate pooled with EdD programs and 0.51 with GRE scores from applicants who post three or more results. The individual parameters move, because the GRE gaps separate the spread of value from its correlation only loosely (Appendix E.7). The largest misses are biology, which falls more than predicted, and business, the physical sciences and agriculture, which fall less.

Figure A5: Field Responses, Observed and Calibrated



Notes: Markers: fully treated ($\tau \geq 6$) SED doctorate coefficients on the fifteen-field stacks, times 100, with 95 percent intervals; black markers are education, the humanities, the arts and business. Bars: the fitted response of the model of this appendix, split into the part through funded places and the part through unfunded enrollment, with the funded side fitted to the thirteen responses by weighted least squares with weights $1/\text{se}^2$ and every other input measured. Diamonds: the same funded side with a free share of self-funded students lost.

E.5 Where the Model Puts the Gap

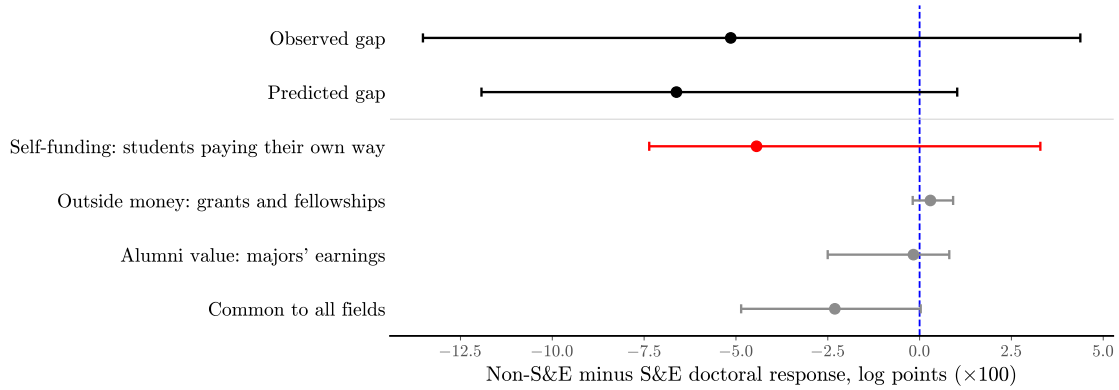
In the calibrated model the gap between fields outside and inside science and engineering runs mostly through the tuition self-funded students pay. Here non-S&E means education, the humanities, the arts and business. The model is fitted on R1 and R2 universities (Section 3.1). There the observed gap, the doctorate-weighted difference of field responses, is -5.1 log points, with an interval of $[-13.5, +4.4]$ that includes zero. It is larger among R1 universities alone, -10.8 ($[-19.2, -1.4]$). The predicted gap is -6.6 on R1 and R2 universities and -7.4 on R1 alone, inside the observed interval in both (Figure A7). Funded places contribute -2.2 and unfunded enrollment -4.4, of which the tuition channel is -7.2. Its interval, $[-11.1, +1.8]$, also includes zero. Tuition rises after a contract, by 7.8 log points (s.e. 3.8; Appendix E.3), and outside science many more students pay it. So the gap comes mostly from how many students pay their own way rather than from how fast they leave: net of displacement, self-funded enrollment falls by 6 to 20 log points in the science fields and by 6 to 24 across all fields. Displacement works the other way, $+2.8$ ($[+2.0, +3.5]$), because applicants turned away in education and the humanities become persisters. Switching contributes -0.9, all of it in education, and the interaction of the three $+0.9$. The observed gap less the predicted one is $+1.5$.

Split by Shapley values among the three field traits of Section 5.6, self-funding carries the largest part of the predicted gap (Figure A6). Self-funding, the unfunded response through the self-funded share, work while enrolled and tuition, accounts for -4.4 log points ($[-7.4, +3.3]$) of a predicted -6.6. Outside money, the grants and outside fellowships that pay part of a research student's stipend and that a contract does not reprice, accounts for $+0.3$. Alumni value, majors' mid-career earnings, which set how much a field's reputation is worth through what its alumni give, accounts for -0.2. The part common to all fields is -2.3. No interval excludes zero. Outside money and alumni value enter the fit itself, so the four parts add up to the fitted gap.

In the model, self-funding carries the gap because self-funded students amplify the drop. With fewer funded students, each remaining student carries more of a program's fixed costs. Tuition, one rate across the university, rises to cover them and pushes out students who pay their own way, which raises cost per student again and cuts funded places again. Each round is smaller than the last. The rounds describe an equilibrium, not a sequence in time. In the model they resolve at once, so the data show only the fixed point they converge to. The calibration takes the size of the tuition rise from the data (Appendix E.3) rather than from this loop. Enrollment settles lowest in the fields where the most students pay their own way.

The model's explanation of the gap rests on the tuition response: without it the predicted gap is only -1.3. That response comes from a reduced-form event study, and its size depends

Figure A6: The Three Field Traits in the Calibrated Model



Notes: Shapley values of the doctorate-weighted non-S&E minus S&E gap in the model's prediction, log points times 100, with 90 percent intervals from 500 draws of the field responses and the tuition response. Self-funding: the unfunded response, through the self-funded share, work while enrolled and tuition. Outside money: outside support per funded research student and the share of fellowships a contract reprices (Table A6), which set the research and teaching cost terms. Alumni value: majors' mid-career earnings (American Community Survey) relative to their mean, which scales the reputation term. A trait that is off takes its doctorate-weighted mean across the 13 fields; coefficients stay at the fit. Common to all fields: the predicted gap with every trait off. Observed gap as in Figure A7.

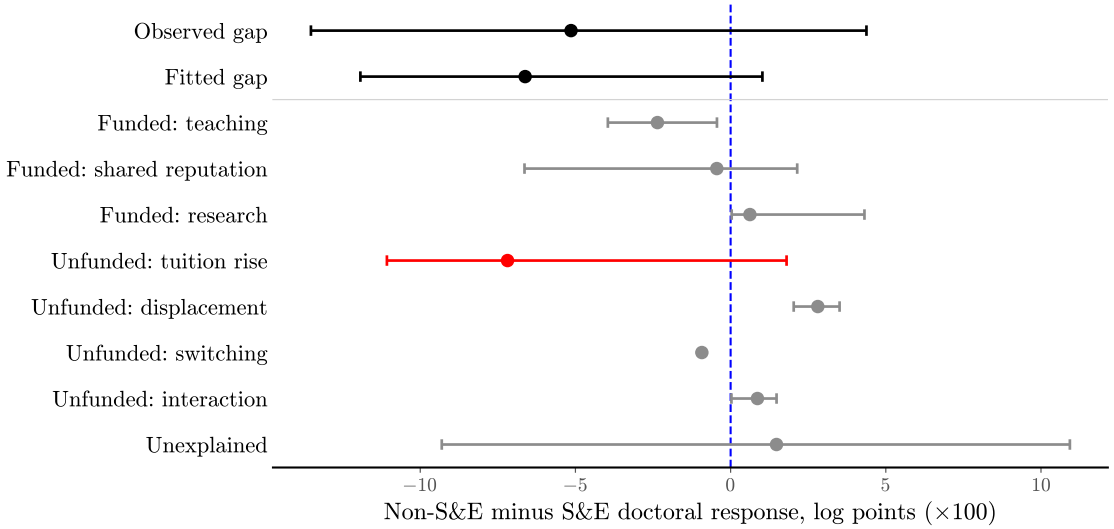
on the measure and the comparison. On the main measure the rise passes its pre-campaign test and holds with sector-specific year effects (Appendix E.3), and in the calibrated model it produces the predicted gap. The observed gap it would explain, though, is not distinguishable from zero on R1 and R2 universities. Nor can the data confirm the fall in self-funded enrollment the calibration implies. In science the calibrated decline is 6 to 20 log points, and the one direct count, the GSS, is too imprecise to confirm or rule it out (Appendix E.3). Outside science, where the calibration puts most of the loss, no count of self-funded students exists. The intervals draw the field responses and the tuition response but hold the students' side at its estimates, so they understate the uncertainty. Finally, the self-funded gradient is as large in the campaign window as after ratification (Appendix D). The model has no campaign, so it describes the whole response but cannot say why part of it comes before the contract.

E.6 Assumptions and Proofs

The model describes one university. Each field k is one market for students, and the subscript is dropped where it is not needed.

Students. A prospective student has a value of the degree v, in dollars a year, and an outside option w^{out}, jointly normal with standard deviations σ_v and σ_{out} and correlation

Figure A7: Decomposing the Non-S&E Minus S&E Gap with the Calibrated Model



Notes: Doctorate-weighted mean field response for education, the humanities, the arts and business minus that for the nine science and engineering fields, in log points times 100. Fitted gap: the model's prediction. Funded terms: the fitted research, teaching and shared-reputation terms of equation (9), which add up exactly. Unfunded terms: the unfunded response at the fitted funded cut, moving the admission cutoff, tuition and the stipend one at a time, with the remainder as an interaction. Intervals are the 5th and 95th percentiles of 500 draws of the field responses and the tuition response from their sampling distributions, refitting each time, with the students' side held at its estimates; the switching term does not vary across draws.

$\rho_{v,\text{out}}$. A funded place is worth $v + w$, an unfunded place $v - T_e + \theta w^{\text{out}}$ and the outside option w^{out}, where w is the stipend, $T_e = T - s^E T_0$ is tuition net of employer support fixed in dollars, and $\theta \in [0, 1)$ is the share of the outside option kept while enrolled. A student applies for funding only if a funded place is the best option, $v + w \geq w^{\text{out}}$ and $w \geq \theta w^{\text{out}} - T_e$. The university fills its L funded places with the applicants who value the degree most, $v \geq v^*$ with $K(v^*, T, w) = L$, and everyone else takes the better of the unfunded place and the outside option. N, K and S denote applicants, funded entrants and unfunded entrants, $s = S/(S + K)$ the self-funded share and $a = K/N$ the admission rate.

Lemma 1 (The queue bound)**.** *If every unfunded student is a rejected applicant, which holds whenever $\theta = 0$, the share of rejected applicants who enroll without funding is $S/(N - K) = \frac{s}{1-s} \cdot \frac{a}{1-a} \leq 1$.*

Proof. $S/K = s/(1 - s)$ and $K/(N - K) = a/(1 - a)$. A student who prefers the unfunded place to a funded one needs $w < \theta w^{\text{out}} - T_e$, which with $\theta = 0$ requires $w < -T_e \leq 0$. □

If every unfunded student were a rejected applicant, the share of rejected applicants who enroll without funding would be 71 percent in education, 93 percent counting employer-supported students, and at most 12 percent in any other field (Section E).

The university. Field k funds A_k teaching and B_k research assistants at stipend w_k, so $L_k = A_k + B_k$. Unfunded students do a share ϕ of a teaching assistant's teaching and ω_S of a research assistant's research, and teaching assistants do a share ω_A of a research assistant's research. The field serves $U_k = H_k(A_k + \phi S_k)$ undergraduates, each worth $p_k^U = P_k(R) + \psi_k(R_k)$, where research $Q_k = B_k + \omega_A A_k + \omega_S S_k$ builds the field's reputation $R_k(Q_k)$ and the university's $R = \sum_j \varsigma_j(Q_j)$. Program cost $C_k(G_k) = \kappa_k^0 G_k + \kappa_k^1 G_k^\nu$, $0 < \nu < 1$, has marginal cost κ_k falling toward κ_k^0, and grants pay f_k per research student. The objective is equation (7). The provost chooses $x = (A_k, B_k)_k$ to maximize Π holding unfunded enrollment fixed, an objective we write $\Pi^F(x; w)$, and chooses T holding the funded allocation fixed. An allocation is a fixed point of the two choices.

Assumptions. (i) Funded places are rationed. (ii) Tuition is one rate per university. (iii) The pre-contract allocation is interior in every field. (iv) The contract covers funded students only. (v) H_k, R_k and ς_j are increasing and strictly concave, P_k and ψ_k increasing and bounded, all twice differentiable. (vi) Grants never cover a student's marginal cost, $\kappa_k^0 + w_k > f_k$. (vii) The marginal values $\lambda_k^Q = \partial\Pi/\partial Q_k$ and $\lambda_k^A = p_k^U H_k'$ fall in Q_k and A_k, so the curvatures η_k^Q and η_k^A are positive. Under (v) and (vi) an optimum exists, because cost grows linearly in enrollment and the value of undergraduates is bounded.

Proposition 1 (Tuition)**.** *With funded places fixed, let* $\eta_k^T = -(T/S_k)\, dS_k/dT|_L$*, which includes the movement of the cutoff, and* $\kappa_k^S = \kappa_k - \omega_S \lambda_k^Q - \phi \lambda_k^A$*, the net cost of a student who pays. At an interior optimum with* $\bar{\eta}^T = \sum_k S_k \eta_k^T / \sum_k S_k > 0$*,* $(T - \bar{\kappa}^S)/T = 1/\bar{\eta}^T$*, where* $\bar{\kappa}^S$ *averages* κ_k^S *with weights* $S_k \eta_k^T$*.*

Proof. The first-order condition is $\sum_k [S_k + (T - \kappa_k^S)\, dS_k/dT|_L] = 0$. Substitute $dS_k/dT|_L = -\eta_k^T S_k/T$ and divide by $\sum_k S_k$. □

Proposition 2 (Cost-weighted funded places fall)**.** *Let* x^0 *and* x^1 *be the provost's optimal choices at stipends* w_0 *and* $w_1 = w_0 + \Delta w$*, with* $\Delta w_k \geq 0$ *and unfunded enrollment held fixed. Then* $\sum_k \Delta w_k L_k^1 \leq \sum_k \Delta w_k L_k^0$*. With a common* $\Delta w > 0$*, total funded places weakly fall.*

Proof. At fixed x the stipend enters only through $-\sum_k w_k L_k$, so $\Pi^F(x; w_1) = \Pi^F(x; w_0) - \sum_k \Delta w_k L_k$. Add the optimality conditions $\Pi^F(x^0; w_0) \geq \Pi^F(x^1; w_0)$ and $\Pi^F(x^1; w_1) \geq \Pi^F(x^0; w_1)$. □

Proposition 3 (Assignments)**.** *For small* Δw_k*, holding* p_k^U*,* κ_k *and unfunded enrollment at pre-contract levels and writing the induced change in the value of shared reputation as a common dollar shift* ξ*,*

$$\Delta \ln Q_k^F \approx -\eta_k^Q \frac{Q_k}{Q_k^F} \frac{\Delta w_k - \xi}{w_{0k} + \kappa_k - f_k}, \tag{10}$$

$$\Delta \ln A_k \approx -\eta_k^A \frac{(1 - \omega_A)\, \Delta w_k}{(1 - \omega_A)(w_{0k} + \kappa_k) + \omega_A f_k}, \tag{11}$$

with $Q_k^F = B_k + \omega_A A_k$*. Teaching falls whenever* $\Delta w_k > 0$*, and funded research falls if and only if* $\Delta w_k > \xi$*.*

Proof. Before the contract a research assistant is hired until $\lambda_k^Q = w_{0k} + \kappa_k - f_k$ and a teaching assistant until $\lambda_k^A = (1 - \omega_A)(w_{0k} + \kappa_k) + \omega_A f_k$. Differentiate each condition in w and use the definitions of η_k^Q and η_k^A. □

The common ξ and the frozen cost and price are reduced-form restrictions, so the fitted η^Q, η^A and ξ are effective parameters, not the structural curvatures.

Proposition 4 (The unfunded response)**.** *At a given cutoff,* $\partial S/\partial v^* \geq 0$ *(displacement),* $\partial S/\partial T \leq 0$ *(tuition) and* $\partial S/\partial w \leq 0$ *(switching), and switching is zero if* $\theta = 0$*.*

Proof. S is the probability of preferring the unfunded place to the outside option without being admitted to a funded one. A higher cutoff admits fewer, a higher net tuition makes the unfunded place worse, and a higher stipend moves students into funded queues. With $\theta = 0$ everyone who prefers the unfunded place already applies. □

From one university to the field data. The calibration treats each field as one market facing the enrollment-weighted tuition of the universities where its self-funded students study, applies one proportional tuition response to every field, and reads doctorates awarded six years on as a fixed fraction of enrollment. Proposition 1 holds within a university, not across these field averages.

E.7 Estimation

In each field, μ_v and v^* are solved so that the model's s and a equal the measured values. That leaves σ_v and $\rho_{v,\text{out}}$, which come from the gaps in the mean and variance of GRE scores between admitted and rejected GradCafe applicants, assuming scores load on the outside option with one loading Λ in every field. Λ is estimated in a restricted fit with one σ_v per group of fields, $\Lambda = 0.32$ ($[0.12, 0.72]$), and each field's pair is then estimated at that loading and shrunk toward its group by bivariate empirical Bayes (Walters, 2024). With one σ_v per group instead, the weighted R^2 is 0.51. The remaining inputs are measured: s^E is the share of doctorate recipients reporting employer reimbursement, and θ the share working while enrolled.

The funded side takes Δw_k from the coded contract minimums and raises against pre-contract stipends, 21 contracts. It fits η^Q, $\eta^Q\xi$ and η^A by weighted least squares with weights $1/\text{se}(\hat{\beta}_k)^2$, iterating with the unfunded response, with base cost normalized to one stipend. The fitted η^Q is 0.26 ($[+0.02, +1.17]$), η^A is 1.04 ($[+0.22, +1.66]$), and the shared-reputation term ξ is \$1.4 thousand a year per funded research student ($[-30.0, +6.6]$).

Two field traits enter the fit, outside money and alumni value. Outside money f_k is measured by field, with a broader measure than the fellowship rate of Appendix D. Research assistants not paid by their own university (a share of 0.36 to 0.75 in the GSS fields, 2005–2016) and NSF and NIH fellows together pay 0.02 to 0.61 of a funded research student's stipend. A contract raises the university's own fellowships with its floor but not those that NSF and NIH set, so the share of fellowships it reprices runs from 0.80 to 1.00. Alumni value, majors' mid-career earnings, scales the reputation term. The Shapley split sets each trait to its mean when off. Table A6 reports every input, parameter and fitted value.

F Research Assistants and Faculty

This appendix gives the specification and checks behind the research-assistant decline of Section 4.4 and, in its subsection, the tenure-system check on the ladder-faculty decline of Section 4.6.

Table A6: The Students' Side by Field: Inputs, Estimated Parameters and Fit

	Measured								From GRE gaps		Doctorates, $\tau \geq 6$ (×100)			
	s	a	θ	s^E	T	σ_{out}	f/w_0	π	σ_v	$\rho_{v,\text{out}}$	Observed	Fitted	Unfunded	Free-loss fit
Arts	0.21	0.17	0.35	0.05	23.8	26.7	0.02	1.00	2.5	0.25	−18.5	−16.4	−5.0	−14.5
Education	0.61	0.37	0.70	0.29	14.6	16.6	0.19	0.99	5.0	0.20	−16.2	−17.0	−10.5	−20.3
Humanities	0.21	0.18	0.35	0.11	24.5	26.5	0.02	1.00	10.0	−0.25	−16.0	−15.4	−3.4	−15.5
Business	0.30	0.15	0.40	0.14	17.7	34.0	0.15	1.00	2.5	0.40	−0.6	−9.2	−5.4	−11.9
Psychology	0.36	0.09	0.35	0.08	17.9	24.3	0.34	0.88	2.5	0.30	−21.4	−17.1	−7.6	−17.2
Health	0.41	0.18	0.34	0.20	17.3	27.2	0.34	0.97	2.5	0.25	−15.8	−10.3	−5.4	−14.4
Social sciences	0.19	0.12	0.28	0.10	23.8	35.5	0.15	0.95	7.5	−0.20	−13.2	−14.8	−3.0	−14.1
Biology	0.12	0.18	0.15	0.08	20.8	26.8	0.43	0.87	2.5	0.40	−13.0	−6.7	−2.3	−7.2
Mathematics	0.06	0.20	0.17	0.08	20.7	40.1	0.40	0.87	10.0	−0.25	−12.6	−11.6	−0.5	−12.2
Engineering	0.09	0.25	0.19	0.08	20.1	33.3	0.55	0.85	7.5	−0.05	−8.6	−6.2	−1.1	−5.8
Computer science	0.11	0.20	0.21	0.10	17.3	49.1	0.60	0.80	15.0	−0.40	−7.8	−6.5	−0.7	−6.9
Physical sciences	0.04	0.23	0.12	0.07	20.3	30.2	0.61	0.82	2.5	0.35	−0.9	−7.0	−0.8	−6.7
Agriculture	0.16	0.19	0.26	0.11	13.1	23.4	0.40	0.98	15.0	−0.40	1.7	−7.4	−1.1	−5.2

Notes: s: share of doctorates enrolled without a funded place (SED Table 4-1, own resources plus employer and other support). a: research-doctorate admission rate (GradCafe PhD decisions scaled to CGS acceptance rates; education uses PhD programs only). θ: share of doctoral students working while enrolled outside graduate jobs (Revelio), used as the share of the outside option kept. s^E: share of doctorate recipients reporting employer reimbursement (SED Table 4-2), used as the expected share of tuition reimbursed. T: graduate tuition in thousands of dollars, weighted by where the field's self-funded students study (IPEDS). σ_{out}: robust standard deviation of earnings of bachelor's degree holders aged 23 to 30 not enrolled, in thousands (ACS 2023). f/w_0: outside support per funded research student as a share of the stipend, the share of research assistants not paid by the institution (GSS, 2005–2016; the social sciences' share where the GSS has no field) plus NSF and NIH fellows (GRFP, F31, T32, three funded years of a six-year degree), over research assistants and fellows. π: share of fellowships not from those programs, which a contract reprices. σ_v (thousands of dollars a year) and $\rho_{v,\text{out}}$: spread of the value of the degree and its correlation with the outside option, estimated field by field from the admit-reject gaps in GRE means and variances at a common loading estimated with one σ_v per group of fields, and shrunk toward the field group's mean by bivariate empirical Bayes. Observed: fully treated SED coefficient on the fifteen-field stacks. Fitted: the model's response with the funded side fitted to the 13 observed responses by weighted least squares; Unfunded is the part that comes through unfunded enrollment. Free-loss fit: the same funded side with a free share of self-funded students lost. Rows above the line are outside science and engineering.

Specification. Table A7 estimates research- and teaching-assistant counts on the stacked GSS panel. Windows run over $\tau \in [-10, +5]$, with at least eight pre-years and two post-years required. The specification is Poisson in levels:

$$E[\, y_{ict} \,|\, \cdot\,] \;=\; \exp\!\left(\alpha_{ic} + \gamma_{ct} + \sum_{b \neq b_0} \beta_b \, D^b_{ict}\right), \qquad b \in \big\{[-10,-8],\, [-4,-2],\, \{-1\},\, [0,2],\, [3,5]\big\}, \tag{12}$$

where D^b_{ict} indicates a treated institution with τ_{ict} in bin b and is zero for every control observation. The omitted bin is $b_0 = [-7, -5]$, the last pre-campaign years, so the campaign window and $\tau = -1$ are estimated as response. The specification has institution-by-stack and cohort-by-year fixed effects, keeps zero-count institution-years, and clusters standard errors by bargaining unit. Treated and control institutions are R1 universities in the 2015 Carnegie classification, controls are never organized, and merged universities are consolidated.

Identifying assumption. This is Assumption 1 applied to counts rather than logs, with one addition: the pre-campaign bin $[-10, -8]$ is estimated and tested against that reference, and the campaign window is estimated rather than pooled into it. With a separate coefficient for each pre-campaign year (Table A14), research assistants still pass ($p = .777$) but teaching assistants reject ($p = .002$), so teaching assistants serve only as a comparison.

Relabeling cannot account for the research-assistant decline. If research assistants had only moved onto fellowships, total funded positions (research and teaching assistantships and fellowships) would hold steady. Instead they fall -6.0 percent on impact (s.e. 1.5) and -8.4 percent at $\tau \geq 3$ (s.e. 3.9). Fellowships themselves fall too, -2.8 percent on impact and -10.5 percent at $\tau \geq 3$ (s.e. 12.2), but they were already drifting down before ratification (Appendix A.1), so their fall alone could hide some relabeling. SED doctorates in science and engineering fall about 10 percent, so in the fields the GSS covers the degree decline is about what the fall in funded positions would predict.

The estimation conventions change the size and precision of the research-assistant decline, not its sign (Table A8, which varies one convention at a time on the same panel). With two post-ratification bins against the whole pre-period, row (1), it is -14.3 percent. It runs from -12.9 under a single Poisson post indicator to -17.2 under the log of positives on a fully balanced window. Teaching assistants are flat under Poisson, at -0.8, but fall nearly as much as research assistants under the log specifications, to -20.1 on the balanced window, though not distinguishable from zero at the 5 percent level. The contrast between the two holds under the Poisson specifications the paper reports, not under every convention.

Table A7: Research and Teaching Assistantships on the First-Contract Clock (GSS)

Event bin	Research assistants (1)	Teaching assistants (2)
$\tau \in [-10, -8]$	0.004	-0.011
	(0.021)	(0.029)
$\tau \in [-4, -2]$	-0.008	-0.003
	(0.022)	(0.034)
$\tau = -1$	-0.039	0.021
	(0.042)	(0.044)
$\tau \in [0, 2]$	-0.125**	-0.009
	(0.044)	(0.051)
$\tau \geq 3$	-0.149*	-0.011
	(0.059)	(0.063)
Private − public, $\tau \geq 3$	0.026	0.210†
	(0.088)	(0.109)
Pre-campaign test p	0.86	0.71
Treated institutions	25	25
Control institutions	43	43

Notes: Zero-filled PPML on stacked institution panels, first-contract ratification clock, stack×institution and cohort×year fixed effects, reference $\tau \in [-7, -5]$. Standard errors clustered by bargaining unit in parentheses. The pre-campaign test is a Wald test on the $\tau \in [-10, -8]$ bin, shown in bold where it rejects at five percent; $\tau \in [-4, -2]$ and $\tau = -1$ fall in the campaign window. The GSS covers science, engineering and health fields only. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{\dagger}p < .10$.

Table A8: Assistantship Estimates Across Estimation Conventions

	Specification	Research assistants	Teaching assistants
(1)	Poisson event bins $\tau \geq 3$	−0.143*	−0.009
		(0.057)	(0.058)
(2)	Poisson single post indicator	−0.129**	−0.008
		(0.043)	(0.043)
(3)	Log of positives single post	−0.147**	−0.138
		(0.056)	(0.089)
(4)	Log fully balanced window	−0.172**	−0.201†
		(0.063)	(0.103)

Notes: Each row varies one estimation choice at a time on the same stacked institution-year panel, contract-ratification clock, R1 comparison group, standard errors clustered by bargaining unit. Row (1) is Poisson on zero-filled counts with two post-ratification bins ($\tau \in [0, 2]$ and $\tau \geq 3$) against the whole pre-period as reference, reporting the $\tau \geq 3$ coefficient. It differs from Table A7, which also estimates the earliest pre-campaign bin, the campaign window and $\tau = -1$ and takes $\tau \in [-7, -5]$ as the reference; there the $\tau \geq 3$ coefficient is −0.149 (0.059) for research assistants and −0.011 (0.063) for teaching assistants. Rows (2) through (4) replace it with a single post indicator, then with the log of positive counts, then with a fully balanced window. Each row changes one convention and holds the panel, the clock and the control pool fixed. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{\dagger}p < .10$.

F.1 Ladder Faculty by Tenure Status

The ladder count of Section 4.6 is by rank, so it could mask a shift in the kind of appointment. It counts full, associate and assistant professors whether or not they hold tenure-system appointments, so a university that replaced departing tenure-system professors with non-tenure-track professors of the same rank would show no ladder decline. Two IPEDS surveys from 2002 onward separate the two (Table A9). The Human Resources survey counts full-time faculty by tenure status. The Fall Staff survey crosses rank with tenure status.

Tenure-system faculty fall by about as much as the ladder count, so the ladder decline is not a change of titles. Over $\tau \in [6, 10]$, full-time faculty who are tenured or on the tenure track are -8.4 log points (s.e. 3.9), and tenure-system professors of the three ranks -6.8 (3.6), against -6.9 (4.3) for the ladder count over the same years. The decline is again among tenured full and associate professors (-6.7 and -9.8, s.e. 3.4 and 3.4). At the assistant rank the two kinds of appointment move apart: those on the tenure track are -7.9 (8.6) and those off it $+30.8$ (22.3). Across the three ranks, the point estimate for professors off the tenure track is a rise of 24.4 (17.4). None of these non-tenure-track estimates is distinguishable from zero, but they are large enough that some substitution toward non-tenure-track professors cannot be ruled out.

We do not carry these series in the body because their pre-campaign tests reject, for tenure-system faculty ($p = 0.013$), for the three tenure-system ranks ($p = 0.021$) and for tenured associate professors ($p = 0.007$). Tenured full professors pass ($p = 0.187$), as does the ladder count over the same years ($p = 0.524$). Before ratification, tenure-system faculty at treated universities were rising relative to controls by about 0.9 log points a year, and they fall after it. That could mean the decline understates the break from trend, or that part of it reverses an earlier rise. The pre-campaign years cannot tell the two apart, so the check agrees with the ladder decline in size but does not support reading it as an effect of the contract.

G The Publication Record in Detail

The publication corpus of Appendix B.4 follows the doctorates themselves, in every field, and Table A10 estimates on it the design of Table A5. The panels are stacked institution×cohort panels with cell×stack and cohort×year fixed effects, the campaign window estimated against a $\tau \leq -5$ reference, standard errors clustered by bargaining unit, and the R1 comparison group for the graduates. The institution outcomes of Panel C compare treated universities with R1 and R2 universities, as Section 4.5 explains. On the

Table A9: Faculty Headcount by Rank and Tenure Status

	Log points (×100)		Leads, p		
	$\tau \in [0,5]$	$\tau \in [6,10]$	Pre-campaign	Campaign	Trea
A. Ladder faculty by rank (Faculty Salaries survey)					
All years	−3.8	−6.6	0.938	0.241	47
	(1.5)	(2.8)			
2002 onward	−3.7	−6.9	0.524	0.334	46
	(2.1)	(4.3)			
B. Full-time faculty by tenure status (IPEDS HR)					
Tenured or on the tenure track	−3.3	−8.4	0.013	0.093	46
	(2.0)	(3.9)			
Tenured	−2.8	−9.4	0.057	0.130	46
	(2.0)	(3.4)			
On the tenure track	−4.7	−5.8	0.100	0.558	46
	(5.1)	(9.6)			
C. Professors by rank and tenure status (IPEDS Fall Staff)					
Tenure system, three ranks	−2.8	−6.8	0.021	0.009	42
	(1.8)	(3.6)			
Not on the tenure track, three ranks	15.6	24.4	0.401	0.108	42
	(13.1)	(17.4)			
Tenured full professors	−1.7	−6.7	0.187	0.040	42
	(2.2)	(3.4)			
Tenured associate professors	−3.5	−9.8	0.007	0.360	42
	(2.1)	(3.4)			
Assistant professors on the tenure track	−6.2	−7.9	0.131	0.014	42
	(4.2)	(8.6)			
Assistant professors not on the tenure track	20.2	30.8	0.155	0.089	41
	(16.3)	(22.3)			

Notes: Stacked Poisson event studies on the department stacks, ratification clock, $\tau \in [-10, 10]$ against $\tau = -1$, stack and cohort×year fixed effects, R1 comparison group, standard errors clustered by bargaining unit in parentheses. Columns 1 and 2 average the event-time coefficients over the stated years. Pre-campaign: joint test of the leads $\tau \in [-10, -6]$; Campaign: joint test of $\tau \in [-4, -2]$. Panel A counts full-time full, associate and assistant professors by rank; the second row keeps the years from 2002, when the other series begin. Panel B counts full-time instructional faculty by tenure status in the IPEDS Human Resources survey. Panel C counts full-time instructional faculty by rank and tenure status in the IPEDS Fall Staff survey (2002 to 2023); before 2012 the survey was required only in odd years. Treated: treated institutions in the estimation sample.

ratification clock 22 treated institutions enter and 18 reach $\tau \geq 6$. Graduates in the years $\tau \in [0, 5]$ were admitted before the contract, so for doctorate counts the admission margin opens only at $\tau + 6$. For placement and publishing, though, what matters most is the last years of the degree, which those cohorts spent under the contract. So we report the partially treated coefficient beside the fully treated one. Intervals for the share outcomes are from the null-imposed wild cluster bootstrap.

Where the doctorates go: no detectable change. The fully treated academic share of Section 4.8 has a bootstrap p of .80, and the industry share moves +0.2 points (0.3) on a base of 7.0 percent. Cohorts who finished under the contract but were admitted before it show no earlier shift: over the partially treated years the academic share is −0.2 points (0.6), which rules out a change of more than about 1.3 points either way.

Selection into the named set moves the academic share little. The academic share is measured among graduates with a named placement, and a contract could change who is named. The fully treated effect on the named share is +0.3 points, so we assume, in that direction, that a contract only adds graduates to the named set, and bound the effect on graduates who would be named either way (Lee, 2009). Then 0.4 percent of treated named graduates are marginal (2.7 at the upper end of the named effect's interval), and the bounds treat them first as all academic and then as all non-academic. Section 4.8 reports the bounds.

Nothing detectable happens to the graduates' own research. Figure A8 collects the estimates against zero, on a common scale within each block. The share of a cohort with a named placement changes by +0.3 points (0.9, bootstrap $p = .74$, interval $[-1.5, +2.1]$), and the share with any post-PhD publication by +0.7 (0.8, $p = .42$). Citations received by everything on a graduate's record to date, per graduate, are −0.8 percent (3.9), with career length absorbed by the cohort×year effects. Leaving out work dated before the degree, they are −1.1 percent (4.0). Articles per graduate in the five years from the degree are −0.3 percent (2.2), and −0.3 (2.4) over ten. First-authored articles are −1.6 (3.2). The share of a cohort with any article in five years is −0.4 points (1.0, interval $[-2.2, +1.7]$) on a base of 55.7 percent. Citations to those articles are +0.2 percent (4.9) at five years and −2.4 (4.8) at ten. Their mean normalized citation percentile, an article's rank among works of its field, year and type, is −0.3 points (0.7) on a base of 63.1, and −0.3 (0.7) at ten years. No estimate in the block is distinguishable from zero. They exclude declines of more than about 8 percent in citations to a graduate's whole record, 5 percent in articles per graduate, 12 percent in citations to articles from the first five or ten years, and 1.7 percentile points in their average citation rank. These bounds assume the graduates are matched to the right authors. At least 69 percent of matches are right. If wrong matches are as common at treated as at control institutions, they pull each estimate and its interval toward zero, so

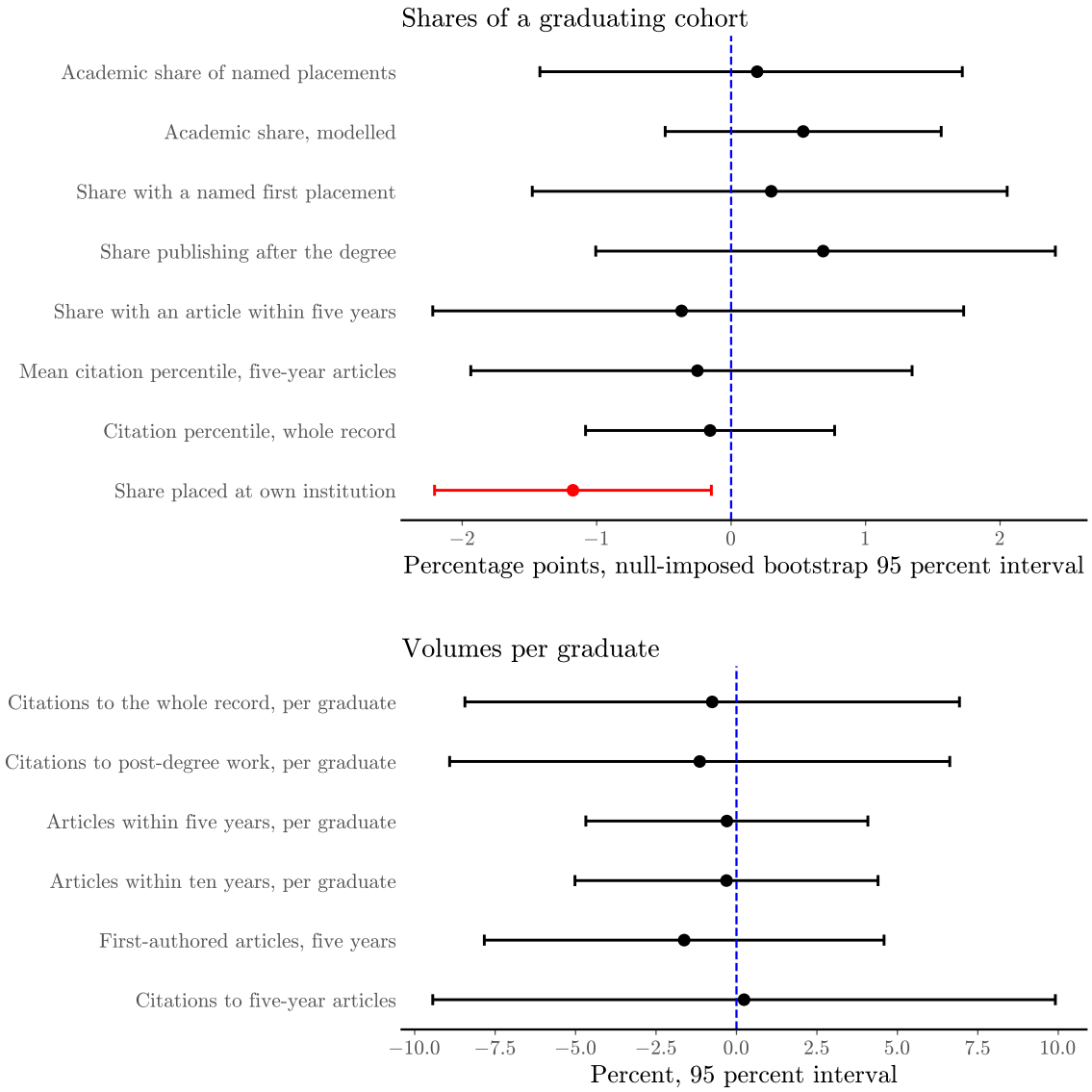
an exclusion of 8 percent is one of roughly 12 in truth. Patenting is +0.30 points (0.47) on a base of 2.2 percent, excluding declines beyond 0.6 points, the coarsest bound in the table. The inventor link observes patents only in the science, engineering and social-science fields and only through 2018.

Their institutions are cited less, while output per faculty member holds. Panel C of Table A10 reports the institution outcomes behind Sections 4.5 and 4.7, on R1 and R2 universities. Citations to a year's output fall about 11 percent at $\tau \geq 6$. Articles published fall −6.5 log points (4.4), not distinguishable from zero. Citation-weighted output per ladder faculty member and the citation rank of articles are unchanged to within about 14 percent and 1.2 percentile points.

The corpus covers most doctorates. The dissertations are mostly those deposited with ProQuest, 85 to 94 percent of SED doctorates in every year from 1975 to 2021. Several universities, Brown and Johns Hopkins among the treated, left ProQuest for their own repositories during the panel. We add the dissertations of 7 such repositories to the corpus. Eighteen institutions identify the fully treated coefficients on the cohort panels and 28 on the department panel, 24 for the rows divided by ladder-faculty headcount.

The person match is validated against withheld ORCIDs, and dropping its least reliable tier leaves every null a null. Several institutional repositories print the author's ORCID on the dissertation record. We withheld those ORCIDs, matched the dissertations by name, field and institution as for every other record, and compared the ORCID on the matched OpenAlex profile. The match is tiered by its own confidence. The top tier names the right person in 96 percent of cases (952 of 994), the middle tier in 62, and the lowest tier in 29. The last is a lower bound, because it is measured only where the matched profile carries an ORCID, and a wrong match to an established author is more likely to carry one. The lowest tier is 27 percent of graduates. Dropping it from every cell moves no placement or publication estimate by more than 0.99 of a standard error. The academic share moves from +0.0019 to −0.0024 (0.0077), the share with any post-PhD publication from +0.7 to +1.1 points, and the share with any article in five years from −0.4 to +0.6. The last moves toward more publication, as a share tied mechanically to match quality would. Own-institution placement becomes slightly larger, −1.25 points (s.e. 0.50) against −1.18, as attenuation from mismatches would imply. Restricting instead to institution-years whose ProQuest coverage of SED doctorates reaches half gives −0.0007 (0.0065) on the academic share and −1.24 points (0.42) on own-institution placement (Table A10, Panel A). On the ProQuest-only panel no coefficient moves by more than 0.16 of its standard error.

Figure A8: What the Publication Record Excludes



Notes: Fully treated coefficients ($\tau \geq 6$) on the stacked institution×cohort panels of Table A10, with cell×stack and cohort×year fixed effects and the R1 comparison group. Share outcomes (upper panel) are in percentage points with the null-imposed wild cluster bootstrap interval; per-graduate volumes (lower panel) are in percent with the conventional interval, since the bootstrap is not run on the count specifications. The two panels are on different scales. An interval that excludes zero is drawn in the darker colour. Section 4.8 reports the base rates and reads each bound.

Table A10: Where the Doctorates Go, and Whether They Publish

Outcome	Partially ($\tau \in [0,5]$)	Fully ($\tau \geq 6$)	All post ($\tau \geq 0$)	Implied ($\tau \geq 6$)	WCB p ($\tau \geq 6$)	Campaign p	Treated (panel / $\tau \geq 6$)
Panel A: Where the doctorates go							
academic share of named placements	-0.002	0.002	-0.000	+0.2 pp	[0.80]	[0.46]	22 / 18
	(0.006)	(0.007)	(0.006)	base 0.729			
industry share of named placements	0.005	0.002	0.003	+0.2 pp		[0.39]	22 / 18
	(0.003)	(0.003)	(0.003)	base 0.070			
placed at own institution, share of named	-0.007†	-0.012*	-0.009*	-1.2 pp	[0.02]	[0.29]	22 / 18
	(0.004)	(0.005)	(0.004)	base 0.076			
modelled academic probability, all graduates	0.004	0.005	0.005	+0.5 pp	[0.29]	[0.21]	22 / 18
	(0.005)	(0.005)	(0.005)	base 0.720			
Panel B: Whether they publish and patent							
named first placement (research trace)	0.004	0.003	0.003	+0.3 pp	[0.74]	[0.63]	22 / 18
	(0.006)	(0.009)	(0.007)	base 0.736			
any post-PhD publication	0.006	0.007	0.006	+0.7 pp	[0.42]	[0.49]	22 / 18
	(0.006)	(0.008)	(0.007)	base 0.700			
citations to the graduate's entire record, per graduate	0.019	-0.008	0.008	-0.8%		[0.18]	22 / 18
	(0.040)	(0.039)	(0.035)	base 1072.09			
citations to work dated from the PhD on, per graduate	0.027	-0.011	0.012	-1.1%		[0.13]	22 / 18
	(0.042)	(0.040)	(0.036)	base 965.53			
any article in the five years after the PhD	-0.001	-0.004	-0.002	-0.4 pp	[0.70]	[1.00]	22 / 18
	(0.007)	(0.010)	(0.008)	base 0.557			
articles in those five years, per graduate	0.000	-0.003	-0.001	-0.3%		[0.84]	22 / 18
	(0.018)	(0.022)	(0.018)	base 3.12			
first-authored articles in those five years, per graduate	-0.005	-0.016	-0.010	-1.6%		[0.41]	22 / 18
	(0.021)	(0.032)	(0.024)	base 1.64			
citations to those articles, per graduate	-0.002	0.002	0.000	+0.2%		[0.18]	22 / 18
	(0.039)	(0.049)	(0.041)	base 164.12			
mean normalized citation percentile of those articles	-0.001	-0.003	-0.002	-0.3 pp	[0.73]	[0.65]	22 / 18
	(0.005)	(0.007)	(0.006)	base 0.631			
articles in the ten years after the PhD, per graduate	0.009	-0.003	0.003	-0.3%		[0.33]	22 / 17
	(0.019)	(0.024)	(0.020)	base 6.55			
mean normalized citation percentile, ten-year articles	-0.003	-0.003	-0.003	-0.3 pp	[0.68]	[0.58]	22 / 17
	(0.004)	(0.007)	(0.005)	base 0.632			
all OpenAlex works in five years (datasets included), per graduate	0.003	0.019	0.010	+2.0%		[0.75]	22 / 18
	(0.017)	(0.022)	(0.017)	base 3.59			
any patent (inventor link), S&E-type fields	0.003	0.003	0.003	+0.3 pp		[0.21]	20 / 15
	(0.004)	(0.005)	(0.004)	base 0.022			
patent within ten years of the PhD	0.003	0.003	0.003	+0.3 pp		[0.21]	20 / 15
	(0.004)	(0.005)	(0.004)	base 0.022			
Panel C: Department research activity							
citations to the year's output, per ladder faculty member	-0.025	-0.028	-0.026	-2.8%		[0.15]	47 / 24
	(0.033)	(0.055)	(0.039)	base 200.89			
citations to the year's output, per byline author	-0.025	-0.044	-0.031	-4.3%		[0.45]	47 / 28
	(0.028)	(0.029)	(0.027)	base 45.56			
citations to the year's output, total	-0.055*	-0.112**	-0.075**	-10.6%		[0.20]	47 / 28
	(0.023)	(0.042)	(0.029)	base 170256.45			
articles published	-0.023	-0.065	-0.035	-6.2%		[0.81]	47 / 28
	(0.031)	(0.044)	(0.032)	base 3203.18			
authors of those articles	-0.037	-0.070†	-0.047†	-6.8%		[0.42]	47 / 28
	(0.025)	(0.037)	(0.027)	base 3461.30			
articles per publishing author	0.011	0.004	0.009	+0.4%		[0.43]	47 / 28
	(0.013)	(0.014)	(0.012)	base 0.93			
citations per article	-0.022	-0.039	-0.028	-3.8%		[0.30]	47 / 28
	(0.024)	(0.026)	(0.024)	base 50.92			
mean normalized citation percentile of the year's articles	0.001	-0.001	0.000	-0.1 pp		[0.83]	47 / 28
	(0.005)	(0.006)	(0.005)	base 0.706			
publishing authors affiliated (author file)	-0.018	-0.049	-0.027	-4.8%		[0.93]	47 / 28
	(0.024)	(0.041)	(0.027)	base 3747.35			

Notes: Stacked institution×cohort panels of first placements and research records built from the OpenAlex corpus (cohorts are year of degree, 1970–2021), on the design of Table A5: cell×stack and cohort×year FE, partially treated $\tau \in [0,5]$ and fully treated $\tau \geq 6$ estimated jointly, the all-post column from a second fit with a single $\tau \geq 0$ indicator, campaign window $\tau \in [-4,-1]$ against a $\tau \leq -5$ reference, SEs clustered by bargaining unit, R1 comparison group. Shares (Panel A, and the share rows of Panel B) are estimated in levels weighted by the graduates behind each share; the coefficient is in share points and the implied change in percentage points, with the treated pre-period mean beneath. Per-graduate rates are Poisson with the cohort size as exposure. Articles are OpenAlex works of type article or review, dated by their own publication year, from the works snapshot; citations are those to the same articles, and the normalized citation percentile is OpenAlex's rank of each article among works of its field, year and type (0 to 1), averaged over the articles and weighted by them. The first two rows of the per-graduate block count citations to every OpenAlex work on the graduate's record, of any type, to the snapshot date; career length is absorbed by the cohort×year effects. Panel C is an institution×calendar-year panel: a work counts for an institution when an author on its byline lists it. Its first three rows take every work of any type dated to the calendar year and the citations that work has received to the snapshot date; the first divides by professors, associate professors and assistant professors from the IPEDS Faculty Salaries survey, an institution-level headcount from outside the publication record, the second by the distinct authors on the institution's bylines that year, and the third is undivided. The last row comes instead from the OpenAlex author file. Per-author and per-article rates are Poisson with the denominator as exposure and counts Poisson with no exposure; for both the implied change is $\exp(\hat{\beta}) - 1$. Shares among named placements condition on a research trace (Panel B, first row), which is the selection margin for Panel A. The WCB column is the null-imposed wild-cluster-bootstrap p (Webb weights, 9,999 draws) where run; the campaign p is conventional. The treated column reports institutions entering the panel and, after the slash, those reaching $\tau \geq 6$ with the outcome observed. First-contract ratification clock. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{\dagger}p < .10$.

Table A11: Placement and Research Trace by Field

	Academic share ($\tau \geq 6$)	Campaign p	Research trace ($\tau \geq 6$)	Treated (panel / $\tau \geq 6$)
All fields (pooled)	0.002	[0.46]	0.007	22 / 18
	(0.007)		(0.008)	
Non-S&E (aggregate)	0.011	[0.62]	0.011	22 / 18
	(0.010)		(0.012)	
S&E (aggregate)	-0.001	[0.20]	0.002	22 / 18
	(0.008)		(0.006)	
Life sciences	-0.020*	[0.19]	0.011	22 / 18
	(0.009)		(0.010)	
Physical sciences and earth sciences	-0.010	[0.91]	0.003	22 / 18
	(0.013)		(0.010)	
Education	0.005	[0.78]	0.010	22 / 17
	(0.016)		(0.018)	
Humanities and arts	0.006	[0.10]	0.010	22 / 18
	(0.014)		(0.016)	
Psychology and social sciences	0.015	[0.01]	-0.011	22 / 18
	(0.013)		(0.014)	
Engineering	0.015	[0.55]	-0.000	22 / 18
	(0.019)		(0.009)	
Mathematics and computer sciences	0.025†	[0.11]	0.006	22 / 18
	(0.013)		(0.015)	
Other non-S&E	0.025†	[0.77]	0.015	22 / 18
	(0.015)		(0.016)	

Notes: The academic share of named first placements and the share of the cohort with a research trace (any post-PhD publication), estimated separately for each SED broad field on the stacked institution×field×cohort panel, with the design, weights and inference of Table A10: cell×stack and cohort×field×year FE, fully treated $\tau \geq 6$, partially treated years absorbed, campaign window against a $\tau \leq -5$ reference, SEs clustered by bargaining unit beneath each coefficient. Coefficients are in share points. Fields are the eight SED broad fields of Table A5, sorted by the academic-share coefficient; a field with fewer than 20 fully-treated cell-years is omitted. The treated column reports institutions entering the panel and, after the slash, those reaching $\tau \geq 6$. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{\dagger}p < .10$.

G.1 Placement and Research-Record Shares

Section 4.8 estimates the binned form of equation (1) on shares of a graduating cohort. Let i index institution, c the contract cohort of the stack and t the year of degree, and let y_{ict} be the share of the cohort with the outcome,

$$y_{ict} \;=\; \alpha_{ic} \;+\; \gamma_{ct} \;+\; \beta^{\text{camp}} D_{ict}^{[-4,-1]} \;+\; \beta^{\text{part}} D_{ict}^{[0,5]} \;+\; \beta^{\text{full}} D_{ict}^{[6,+]} \;+\; \varepsilon_{ict}, \tag{13}$$

weighted by the number of graduates behind the share and clustered by bargaining unit. The weight is the cohort for the research-record shares and the named placements for the sector shares. The reported coefficient is β^{full}. In the field panels α and γ carry a field subscript, so every comparison is within field-year within stack. Per-graduate rates (publication-years, works, citations) are estimated as Poisson on the cohort sum with the cohort size as exposure. Department counts are Poisson with no exposure, as in equation (2).

Identifying assumption. This is Assumption 1 applied to shares, with one addition specific to this source. A graduate is observed only once they publish under an employer's name, so the probability of being observed is itself an outcome. The sector shares are therefore conditional on being named. A contract that changed who is named without changing sector would still move the conditional shares if the newly named graduates differ in sector from the rest. For that reason we bound the effect on graduates who would be named either way (Appendix G). Table A10 also reports a modeled academic probability defined over all graduates. It treats a graduate who is never observed as outside academia, so it moves with the share who publish and is not separate evidence.

Who the missing graduates were. The fall in named graduates also bounds who was cut. Six or more years after ratification, treated cohorts have about 8 percent fewer graduates with a named placement. Call that share χ, and let χ_0 be the fraction of the missing graduates who would have gone outside academia. If the contract changes nothing else about where graduates go, it moves the academic share from $\bar{y}$ to $(\bar{y} - \chi(1 - \chi_0))/(1 - \chi)$. Inverting this at the ends of the academic share's bootstrap interval, with χ held at its estimate, puts χ_0 between 11 and 47 percent, with a point estimate of 29 percent. Before ratification, 27 percent of named graduates went outside academia. So the missing graduates look like the rest of their cohort. The evidence rules out a cut that fell mostly on students headed outside academia, or almost entirely on those headed into it. In science and engineering the range is 9 to 47 percent. Outside science fewer graduates are lost and the share is noisier, so its range, 3 to 71 percent, says little. The ranges are conditional on χ: its own interval reaches zero, and if almost no graduates were lost, the academic share could not say who they were.

H Robustness and Inference in Full

Every main outcome passes through three audits, of pre-trends, inference and treatment classification. The pre-trend audit runs joint Wald tests on the pre-campaign years of every main outcome (Table A14). Appendix A.1 lists the outcomes that fail it and how each is read, and none carries a claim of its own. The inference audit re-estimates every main coefficient under the null-imposed wild cluster bootstrap. It uses randomization inference where the estimator rules out the bootstrap or the treated count is smallest (Table A15). The treatment-classification audit re-estimates the main results under each coding of the joint units and of research-assistant coverage (Figure A3). It leaves the sign and timing of the contraction and the sign of the stipend rise intact while moving the size of both. Under two codings of the joint units, the SED estimate is no longer distinguishable from zero and the stipend rise is significant only at the 10 percent level.

Table A12 lists each threat to the doctorate estimate and what its check returns. The paragraphs below take the comparison group and selection into organizing, and Appendices H.1 to H.3 take the rest.

Matched on pre-ratification trajectory, the controls give a smaller contraction with the same sign and timing. From 1995 to 2024 the controls grew faster than the treated universities. Indexed to 1995, control institutions reach 147 in SED research doctorates by 2024 against 123 for treated institutions, and in IPEDS the gap is wider still. A stacked design can read such a divergence as a treatment effect, so we also match controls on each institution's own pre-ratification slope, keeping only controls whose slope falls inside the treated interquartile range. The SED estimate falls by about 37 percent, to -0.076 (0.030). The research-doctorate panel gives -0.105 (0.031) and the all-field one -0.069 (0.035), each still distinguishable from zero (t between 2.0 and 3.4). What survives sits where the cohort clock puts it: the partially treated coefficient collapses from -0.04 to roughly zero, leaving the decline at $\tau \geq 6$. The all-field IPEDS pre-test passes both unrestricted and under the match ($p = .491$ and $p = .899$).

We keep the estimate against all R1 controls as the main one, because the matched estimate likely understates the contraction. Institutions organize partly because of the trajectory they are on, so selecting controls to share it removes variation that belongs to treatment as well as to confounding. A slope estimated over ten years of institution-level counts is also noisy enough that matching on it attenuates by itself. And the match, imposed stack by stack, keeps every control institution but discards a large share of control observations. Against research universities on comparable pre-contract trajectories the contraction is still of the order of 8 percent, and the sign, the timing and the agreement

Table A12: Threats to the Doctorate Estimate and What Each Check Returns

Threat	Check	Result (SED, fully treated)
Regional trends	Democratic-state×cohort×year effects	−0.130 (0.043) against −0.122 (0.035)
Fiscal shocks	State appropriations per student, endowment and state revenue as controls, 1980 on	−0.113 (0.040) against −0.111 (0.039)
University-wide decline	Professional degrees (law, dentistry, pharmacy), whose schools employ few graduate workers	JDs +0.013 (0.055); other professional degrees +0.175 (0.093); bachelor's degrees −0.132 (0.033)
Controls on other trajectories	R1 and R2 universities; pre-trend-matched controls	−0.087 (0.036); −0.076 (0.030)
	All never-organized doctorate-granting universities; region×cohort×year effects	−0.092 (0.036); −0.068 (0.039)
Functional form	Least squares on log doctorates, unweighted; weighted by pre-period doctorates	−0.156 (0.073); −0.104 (0.034)
Selection into organizing	Campaigns that never won a contract	placebo +0.013 (0.037)
Few treated clusters	Randomization inference	$p = .007$
Many outcomes	Benjamini–Hochberg over 17 outcomes	$q = .028$ (stipends $q = .037$, non-S&E doctorates $q = .012$)
Who counts as treated	Joint graduate-undergraduate units treated	−0.058 (0.047)
	Undated first contracts dated; NYU at 2002	−0.120 (0.036); −0.124 (0.035)
One large university	Field panel without New York University	non-S&E/S&E ratio 1.7 against 2.1

Notes: Each row re-estimates the SED fully treated coefficient of Section 4.2 (baseline −0.122, s.e. 0.035) under the stated change, with standard errors in parentheses, except the randomization-inference row, the multiplicity row (q-value), the law row, which estimates JD degrees awarded on the same design, and the last row, which reports the ratio of the non-S&E to the S&E field coefficient. The broader comparison-group row replaces the R1 controls with every never-organized university awarding at least one research doctorate a year, and treats every first contract; the region row adds census-region×cohort×year effects to the R1 design. The text of this appendix discusses each row.

across censuses hold throughout.

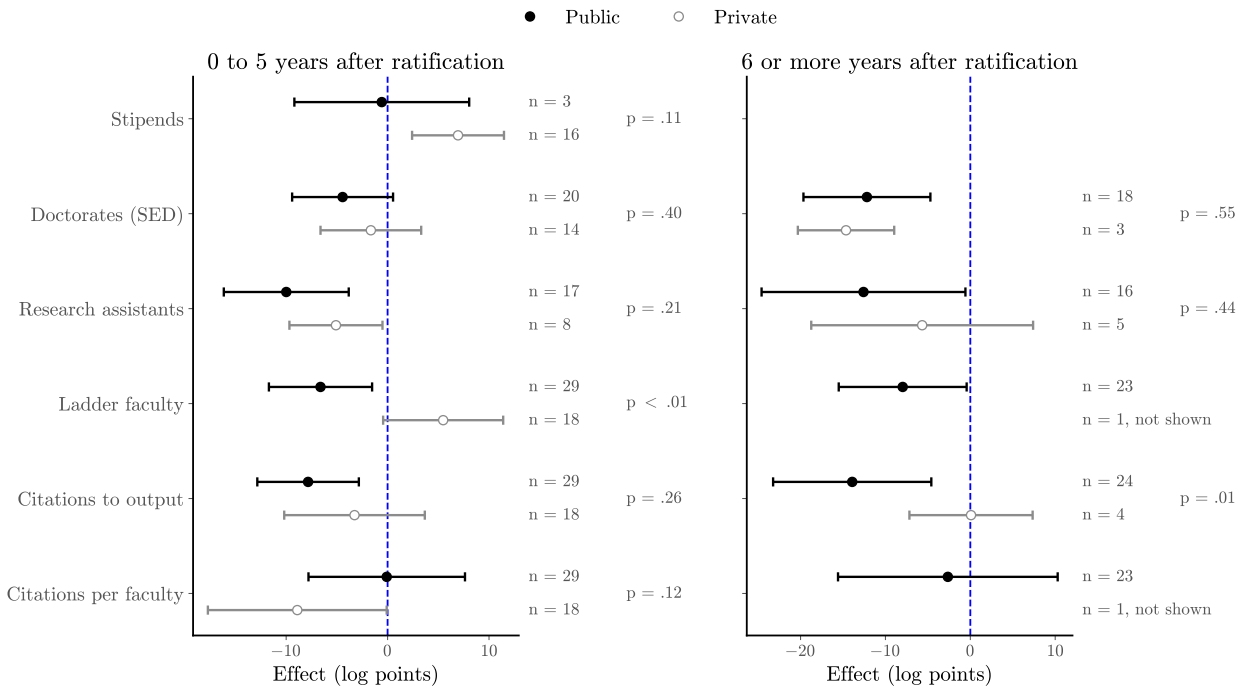
Campaigns that never produced a contract show no fresh break. The roster holds 26 institutions that organized and never signed a first contract. At Loyola Chicago, Southern Illinois Edwardsville and Boston College, certification produced no agreement, and the rest were never certified. Pseudo-treating them at their campaign year on the same design gives +0.013 (s.e. 0.037) in SED and +0.010 (0.043) on the research-doctorate panel, on the fourteen of them that are R1 universities, of which 6 reach $\tau \geq 6$. On the research-doctorate panel the placebo differs from the treated estimate at $t = 2.5$, so selection of this kind can account for at most about 58 percent of that contraction. On SED the placebo also differs from the treated estimate ($t = 2.7$). The placebo institutions fail their own pre-campaign test ($p < .001$). They drifted down at 2 to 3 percent a year before anyone organized, against a drift of 0.3 percent in the treated set. Many of the public ones are in states that give public employees no bargaining right, where state law rather than the campus decided that no contract would follow. That makes them a poor counterfactual for the selection that produced our treated set. What the exercise shows is narrower: institutions already contracting faster than ours, which organized without winning a contract, show no fresh break six years later.

H.1 Public and Private Universities

Public and private universities, which fund doctoral students from different budgets, respond alike on most main outcomes. Two differences stand out: citations six or more years out, which rest on few private contracts, and ladder faculty in the first five years. Figure A9 re-estimates each main specification with its treatment terms split by IPEDS sector, and the pooled version of each reproduces the main estimate. Six or more years after ratification, the SED doctorate coefficient is −12.2 log points (s.e. 3.8) at public universities and −14.6 (2.9) at private ones ($p = .549$ for the difference). The citation coefficient is −13.9 against +0.1 ($p = .009$), the one difference at that horizon. Few private contracts are that old, however. Of the 14 private universities in the SED panel, 3 are observed that long after ratification, and only 1 in the faculty panel. Over the first five years, stipends change by +6.9 log points at private universities and −0.6 at public ones, a gap not distinguishable from zero ($p = .109$), though the stipend panel observes only 3 public universities. Doctorates ($p = .397$), research assistantships ($p = .206$) and citations ($p = .259$) do not differ either. Ladder faculty do: −6.6 at public universities against +5.5 at private ones over the first five years ($p = .002$). Nearly every private contract dates from after *Columbia*, so this compares sectors that ratified in different eras, and with 1 private university observed past five years

we cannot tell whether the gap lasts. We do not carry it as a result.

Figure A9: Main Responses at Public and Private Universities



Notes: Each main specification with its treatment terms split by IPEDS sector (HD2024 control; private includes for-profit). Left panel: the partially treated term (τ 0 to 5; research assistantships τ 0 to 2; stipends the post-ratification indicator of the cost-of-living specification). Right panel: the fully treated term ($\tau \geq 6$; research assistantships $\tau \geq 3$). SED doctorates on the all-cohort stacks; ladder faculty, citations to the year's output and citations per ladder faculty member (ladder headcount as exposure) on the department stacks, with a campaign-window term; research assistantships on the GSS stacks in event bins; all Poisson with stack and cohort×year fixed effects and R1 controls. Stipends: log median pay on log regional price parity, school and field-bucket×year fixed effects, weighted by report counts. Intervals are 95 percent, clustered by bargaining unit. n counts the treated institutions in each group, in the right panel those observed in the fully treated window; a group's fully treated estimate is omitted where fewer than three institutions reach it. p tests the equality of the public and private coefficients.

H.2 Geographic Confounding

Treated universities are concentrated in Democratic-leaning states (Assumption 5), and the concentration predates the private organizing that followed *Columbia*. In November 1999 graduate student locals were 34 percent Midwest, 31 percent West and 25 percent Northeast against 10 percent in the South, and 86 percent were at public institutions (Julius, 2023).

The test is whether a point estimate is stable under geographic controls, not whether it stays significant. The stipend effect is stable. Netting out regional price growth, a control that exists only for pay, leaves it unchanged at +6.3 percent before and +6.3 after, still

significant.

A stipend is a dollar amount and can be adjusted for local prices, but a count of doctorates cannot, so for degrees the only geographic controls are fixed effects, and the finest of them leave too little variation. Most states hold one or two treated universities and few R1 controls. State-by-year effects leave about 14 percent of the treatment variation in the Poisson specification, against 86 percent with year effects for Democratic-leaning states. With state-by-year effects the SED estimate is -0.033 (s.e. 0.079), too imprecise to read either way. For doctorates we therefore compare each unionized university with controls in Democratic-leaning states, and rest the remaining geographic case on timing.

For doctorates, interacting cohort×year with an indicator for a Democratic-leaning state leaves the SED estimate close to where it was, -0.130 (s.e. 0.043) against a -0.122 (0.035) baseline. It shrinks the two IPEDS estimates, to -0.102 (0.035) against -0.127 on the research-doctorate panel and -0.077 (0.040) against -0.097 on the all-field one. These are estimated in the Poisson specification of Section 4.2.

Timing asks a question no control for place can answer. A confounder reproducing the event-year path of Figure 5, flat for three years and breaking from $\tau + 3$, has to key its timing to each institution's own ratification year. It has to switch on within a few years of ratification, since the pre-period is flat. It has to leave the first three post-contract years untouched, which rules out anything acting at the same time on enrolled students or institutional budgets. It also has to follow each institution's ratification year rather than the calendar, because the stacks are aligned on event time and ratification years run from 1975 to 2026. A statewide funding shock, a regional wage boom or a national trend in doctoral demand fails that condition: none of them knows when a particular university signed a contract.

What the shape rules out is a geographic shock that runs on the calendar. A confounder tied to each university's ratification date and acting with a six-year lag would survive the timing test, wherever it came from. The obvious candidate, that universities organize when their doctoral programs are already deteriorating, is what Assumption 1 and the pre-campaign tests are for.

H.3 Fiscal Shocks

A university hit by a cut in appropriations or a fall in its endowment would shrink, and its graduate workers would have reason to organize. If such shocks drove both, they would show up in the university's finances before ratification, during the campaign. They do not. On the stacks of the doctorate estimate of Section 4.2, with the same fixed effects, pre-campaign

terms and bins, the campaign-window coefficient on the log of a public university's state appropriation revenue is +1.8 log points (s.e. 2.6), with pre-campaign test $p = .373$ on 16 treated universities. Over the same window tuition revenue is −2.3 (1.6), gifts −0.9 (7.3) and endowment value −7.2 (6.6), none distinguishable from zero. Six or more years after ratification the estimates are too imprecise to rule out large changes: state revenue −5.0 (9.6), tuition revenue +6.6 (4.4), gifts −16.4 (15.1) and endowment −3.6 (21.5). State education appropriations per full-time student in the treated universities' states do not fall in the campaign window either (−1.3, s.e. 3.5), though their pre-campaign test rejects ($p < .001$), so those states' appropriations were on their own path well before any campaign.

Controlling for these measures leaves the doctorate decline in place. On the years with appropriations data, 1980 on, the SED estimate is −0.111 (0.039) without controls. Adding the log of state appropriations per student, interacted with an indicator for a public university, gives −0.112 (0.039). Adding the log of the university's endowment a year earlier gives −0.114 (0.039), and adding a public university's state revenue a year earlier gives −0.113 (0.040). Institution finances begin in the 1990s in IPEDS, so years without them carry a missing indicator.

Professional schools test whether the decline reaches the parts of the university graduate workers do not touch (Table A13). Law schools employ almost no graduate workers and usually keep their own tuition, so a contract that raises the cost of graduate labor should not reach them. On the stacks of the doctorate estimate of Section 4.2, the 21 treated universities with a law school, 11 of them observed six or more years after ratification, award JD degrees at +1.3 log points (s.e. 5.5) at $\tau \geq 6$ and −0.2 over the campaign window, with pre-campaign test $p = 0.15$. On R1 and R2 universities the first is +0.6. Treated universities' bachelor's degrees, by contrast, are 13.2 log points lower over the same years, a decline we cannot attribute to the contract (Section 4.9). Dentistry and pharmacy, which also employ few graduate workers, do not fall either: +11.9 (s.e. 10.0) and +0.5 (15.1) at $\tau \geq 6$, on 12 and 12 treated universities. Pooled with optometry, which no treated R1 university awards, the three rise +17.5 (9.3), significant only at the 10 percent level, with pre-campaign test $p = 0.14$. The check rules out a decline that reaches every unit. It does not rule out a university that spreads a contract's cost over the units that share a budget with the graduate programs, since law schools usually do not.

H.4 Inference: Wild Bootstrap and Randomization Inference

Standard errors are clustered by bargaining unit rather than by institution. This matters only for the public systems, where one contract covers every campus. The 10

Table A13: Professional Degrees Around First-Contract Ratification

	Law (JD) (1)	Dentistry (2)	Pharmacy (3)	Professional, pooled (4)
Campaign window, $\tau \in [-4, -1]$	−0.002	−0.022	−0.001	0.022
	(0.023)	(0.034)	(0.061)	(0.033)
Partly treated, $\tau \in [0, 5]$	0.028	0.044	−0.049	0.085
	(0.039)	(0.086)	(0.121)	(0.074)
Fully treated, $\tau \geq 6$	0.013	0.119	0.005	$0.175^{\dagger}$
	(0.055)	(0.100)	(0.151)	(0.093)
Pre-campaign test p	0.15	0.24	0.32	0.14
Treated universities	21	12	12	16
Observed at $\tau \geq 6$	11	9	9	12
Pre-period mean awards	215	107	58	124

Notes: Professional-practice degrees awarded by treated universities (IPEDS Completions, 1984–2024, first-professional and doctor's professional-practice award levels): JD (CIP 22.01), DDS and DMD (51.04), PharmD (51.20), and in column (4) the sum of these two with optometry (51.17), which no treated R1 university awards. Zero-filled PPML on the stacked institution panels of the doctorate estimate of Section 4.2, contract cohorts 1994 on, stack×university and cohort×year fixed effects, R1 comparison group, reference $\tau = -5$. The pre-campaign test is a joint Wald test on $\tau - 10$ through $\tau - 6$. Universities with no award of the degree in a stack drop out of that column. Standard errors clustered by bargaining unit in parentheses. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{\dagger}p < .10$.

University of California campuses bargain as one unit, as do the State University of New York and California State University campuses, so each system is one cluster and every other institution is its own. Several outcomes identify from as few as 3–8 treated institutions, where cluster-robust standard errors are known to overstate precision (MacKinnon et al., 2023). In our data, placebo-t distributions from randomly assigned event years have standard deviations of 1.8–3.5 against a nominal 1.0. So for all main coefficients we report null-imposed wild cluster bootstrap p-values (Cameron et al., 2008; MacKinnon and Webb, 2018). For the smallest treated counts we use studentized randomization inference (RI-t). Placebo ratification years are assigned to random control institutions within each stack, the full stack is rebuilt per draw, and the observed t-statistic is compared with the placebo t distribution (MacKinnon and Webb, 2020). The studentized version is preferred because coefficient-based randomization inference is too conservative when treated institutions are larger than the placebo pool. Table A15 reports both for every main coefficient.

Log coefficients are reported under two conventions, which differ by several points for the largest effects. Prose on stipend effects multiplies the coefficient by 100 and calls the result a percentage change. Prose on counts, among them doctorates, assistantships, citations and faculty, uses the implied proportional change $\exp(\hat{\beta}) - 1$, and the field tables report both. The IPEDS non-S&E coefficient of −0.125 is −12.5 log points and an implied change of

−11.8 percent. The SED non-S&E coefficient of −0.209 is −20.9 log points and an implied change of −18.8 percent. The SED estimate of −0.122 is −11.5 percent and the all-field IPEDS −0.097 is −9.3. Ratios of field effects are ratios of the coefficients, and standard errors are always on the log scale.

Table A14: Pre-Trend Audit: Fully Untreated Leads, with the Campaign Window Reported Separately

	Fully untreated leads ($\tau-10$ to $\tau-6$)			Campaign		
Outcome	χ^2 (df)	p	Slope/yr	$\tau \in [-4,-1]$	Treated	Verdict
GSS full-time enrollment (stock)	6.1 (5)	0.292	+0.2%	+0.042†	25	Flat
GSS part-time enrollment	19.4 (5)	0.002	+0.7%	-0.014	25	Level wiggle
GSS fellowships	22.8 (5)	0.000	-1.7%	+0.114†	25	**Trend**
GSS first-time enrollment	3.5 (5)	0.617	+0.4%	+0.011	25	Flat
GSS research-assistant count	2.5 (5)	0.777	+0.3%	+0.017	25	Flat
GSS teaching-assistant count	19.3 (5)	0.002	+1.2%	+0.058	25	**Trend**
IPEDS completions	4.4 (5)	0.491	-0.7%	-0.017	26	Flat
SED completions	2.3 (5)	0.806	-0.3%	-0.020	34	Flat

Notes: Joint Wald tests that the *fully untreated* event-study leads equal zero, against a $\tau-5$ reference, with the fitted linear drift through those leads in percent per year. The organizing campaign, $\tau \in [-4,-1]$ in equation (1), is reported in its own column rather than folded into the test: Assumption 2 estimates campaign responses rather than assuming them away, so a test including those years asks whether a coefficient the paper expects to be non-zero is zero. Each outcome is estimated on its main specification, contract-ratification clock, R1 comparison group, Poisson on zero-filled counts, clustered by bargaining unit. The verdict separates the omnibus test from the slope because they can disagree: a *trend* rejects and drifts at least 1 percent a year; a *level wiggle* rejects on a flatter slope; an *imprecise drift* carries a slope of at least 1 percent a year that five leads cannot resolve; *flat* is neither. Passing here means no *detectable* drift in the pre-campaign years: five leads on as few as 25 treated institutions have limited power, and Appendix H.4 documents placebo-t standard deviations of 1.8 to 3.5 against a nominal 1.0 at these counts. †Campaign coefficient significant at 10 percent.

Under the Benjamini–Hochberg false-discovery-rate correction over the 17 outcomes the paper reports (Table A16), every result the text reads as an effect of the contract clears $q < .05$. They are the stipend increase ($q = .037$), the SED doctorate estimate ($q = .028$) and its non-S&E and S&E parts ($q = .012$ and $.017$), the research-assistant decline ($q = .033$), ladder faculty ($q = .041$) and citations to institutions' output ($q = .028$). The all-field IPEDS estimate clears too ($q = .047$). The family has one fully treated coefficient per outcome and excludes anticipation terms, field and sector splits, alternate clocks and robustness variants. The doctorate p-values come from randomization inference, because the main estimate is Poisson, and the table note gives the kind of each p-value. Without the four undergraduate outcomes, the largest q among these results is .036.

Table A15: Re-Inference Audit: Bootstrap and Randomization Inference

Coefficient	Estimate	(SE)	Conventional p	WCB p	RI-t p
Completions					
SED log completions	-0.156^{*}	(0.072)	0.035	0.041	—
SED completions, Poisson	-0.121	(0.035)	0.001	—	0.007
IPEDS log completions	-0.183^{**}	(0.053)	0.001	0.004	—
IPEDS completions, Poisson	-0.099	(0.041)	0.015	—	0.025
Stipends					
Stipends (contract DiD)	0.063^{*}	(0.022)	0.006	0.015	—
Alumni earnings					
Pooled level, contract clock	0.013	(0.021)	0.542	0.702	—
Pooled level, certification clock	0.004	(0.022)	0.841	0.853	—
INC1, contract clock	0.074	(0.032)	0.026	0.157	0.222
INC2, contract clock	0.008	(0.025)	0.761	0.756	—
INC3, contract clock	-0.019	(0.023)	0.413	0.462	0.640
Undergraduate outcomes					
First-year retention	-3.035^{*}	(0.968)	0.003	0.045	—
150%-time graduation rate	-0.000	(0.021)	1.000	1.000	—

Notes: Every main coefficient re-estimated with the null-imposed wild cluster bootstrap (Webb weights, 9,999 replications, clustered by bargaining unit) and, where it was run, with studentized randomization inference. All estimates use the R1 comparison group. Poisson rows carry a dash in the WCB column because no score bootstrap is available for Poisson; those coefficients are covered by randomization inference instead, which re-estimates rather than perturbing scores and is therefore indifferent to the estimator. Earnings RI-t draws placebos from all control institutions, matching Table A1; drawing only from the largest quarter of them gives p-values that differ from these by between -0.038 and $+0.022$. The final column reports the studentized randomization-inference p-value for the rows where it was run and a dash elsewhere. Stars reflect WCB p-values where available. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{\dagger}p < .10$.

Table A16: Benjamini–Hochberg False Discovery Rate Across the Main Outcomes

Outcome	p	BH q
Bachelor's degrees	0.000	0.001
Doctorates outside S&E, SED fields	0.001	0.012
Doctorates in S&E, SED fields	0.003	0.017
Doctorates, SED	0.007	0.028
Citations to institutions' output	0.008	0.028
Research assistants, $\tau \geq 3$	0.012	0.033
Stipends	0.015	0.037
Ladder faculty, τ 6 to 10	0.019	0.041
Doctorates, IPEDS	0.025	0.047
Graduate tuition, effective	0.040	0.068
First-year retention	0.045	0.069
Citations per ladder faculty member	0.606	0.858
log alumni earnings, pooled	0.702	0.917
Graduates' academic share	0.798	0.917
Graduates' career citations	0.847	0.917
Teaching assistants, $\tau \geq 3$	0.863	0.917
150%-time graduation rate	1.000	1.000

Notes: Benjamini–Hochberg step-up false-discovery-rate q-values over a family of $m = 17$ hypotheses, one fully treated coefficient for each outcome the paper reports, contract-ratification clock. Anticipation coefficients, field and sector splits, alternate clocks, and robustness variants are excluded, since they re-estimate the same hypothesis. Input p-values are of three kinds. The null-imposed wild cluster bootstrap for Stipends; Graduates' academic share; First-year retention; 150%-time graduation rate; log alumni earnings, pooled. Studentized randomization inference for Doctorates, SED; Doctorates, IPEDS, the two Poisson doctorate estimates, as in Table A15. Conventional cluster-robust two-sided p-values for Doctorates outside S&E, SED fields; Doctorates in S&E, SED fields; Research assistants, $\tau \geq 3$; Teaching assistants, $\tau \geq 3$; Ladder faculty, τ 6 to 10; Citations to institutions' output; Citations per ladder faculty member; Graduates' career citations; Bachelor's degrees; Graduate tuition, effective. q-values are enforced monotone. The largest q among the results the text reads as effects of the contract (stipends, the three SED doctorate estimates, research assistants, ladder faculty and institution citations) is $q = 0.041$. Without the four undergraduate outcomes, on $m = 13$, it is $q = 0.036$.

I Coding the Contracts

I.1 What the Agreements Contain

We collected 48 first agreements, and large language model agents coded their provisions (Appendix I.2). They depart in three respects from the common picture of a first contract as pay plus fee relief and a funding guarantee: full fee remission is rare, half write a benefit as a continuation, and multi-year guarantees are uncommon (Table A17).

Contracts nearly always cover pay, and rarely remit fees in full. 45 of the 48 agreements set a minimum pay rate, some as an hourly or monthly rate rather than an annual stipend. Mandatory-fee remission is the opposite. 19 agreements contain no fee provision at all, 26 a partial or conditional one, and 3 remit fees in full. So the common picture holds in full for 3 of the 48. The direction of the coding rule matters here. A contract not shown to remit fees is coded as remitting none, so 19 is a ceiling on true absence and the 29 with some provision a floor. Health insurance sits in between. 31 agreements state a numeric employer premium share. Where stated it is high and flat, a mean of 95.2 percent of the individual premium and a median of 100.

Half the contracts write a benefit as the continuation of one already in place. Of the 48, 24 write a benefit as a continuation of one already in place rather than as a new grant. They use the language of continuing, maintaining or not reducing an existing rate. A clause that locks in an existing benefit binds the university in future years without raising its cost today. So the level a contract specifies is an upper bound on the increment it delivers, and Section 4.3 carries that distinction through.

Multi-year funding guarantees are real but uncommon, and easily overcounted. Of the 48, 9 commit to multi-year funding, 7 of them stating a number, between 2 and 5 years. 2 of the 9 state no number at all and are counted without one: Cornell promises funding for as long as a student is enrolled and in good standing, and Southern California defers to a multi-year offer whose length the agreement never gives. A further 7 set only a minimum length for a single appointment: 1 at a quarter, 1 at a semester, 3 at an academic year, 1 at nine or twelve months, and 1 stating a minimum term without giving its length. None of these can work as a guarantee. They bound the commitment within a funding spell rather than the number of spells, and every stated value is at or below one year. Counting them with the multi-year commitments would put the incidence at 33 percent rather than 19.

Table A17: Provisions Coded from Graduate-Employee Collective Bargaining Agreements

	Agreements	Percent
Mandatory-fee remission		
No provision	19	39.6
Partial or conditional	26	54.2
Full	3	6.2
Other provisions		
Employer premium share stated as a percent	31	64.6
Multi-year funding guarantee	9	18.8
Minimum appointment term only	7	14.6
Benefit written as a continuation	24	50.0
Document vintage and verification		
First contract (not a successor)	26	54.2
Verified, no change	25	52.1
Verified, corrected	23	47.9

Notes: Coding of one collective bargaining agreement for each of 48 treated institutions: the first contract where it could be located (26), otherwise a successor; percentages are of the 48 coded. Each field was required to carry the verbatim contract sentence supporting it, and a field without one was recorded as not stated rather than inferred, so every count here is a floor. Verification re-read the source document and checked each field against its quote in both directions, asking of each filled field whether the quote supports it and of each field coded absent whether the document in fact states it; it reached all 48 agreements and corrected 23, most often a wrong article citation attached to correct quoted language. No fee-remission value changed in verification; a later full-text pass over every fee clause recoded two. Nine agreements carried guarantee language that had been coded silent; three of those are counted in the multi-year row and six in the minimum-term row, the remaining entries in each row having been coded on the first pass. That classification is described in Appendix I. "Benefit written as a continuation" counts agreements whose premium or fee language commits the university to continue, maintain, or not reduce an existing benefit rather than to grant a new one. Employer premium shares, where stated as a number, have mean 95.2 and median 100 percent of the individual premium across 31 agreements. Agreements located online are usually the current one rather than the first; the vintage panel reports which.

I.2 How the Coding Was Done

We collected 48 agreements, one per institution with a first-contract year: the first contract where we could find it (26), otherwise a successor. Large language model agents coded each one, working in parallel, one agreement at a time. They recorded the employer share of the individual health premium, the treatment of mandatory student fees, any funding-guarantee term and the stipend schedule.

Two features of the procedure govern how the counts should be read. First, every filled field had to carry the verbatim sentence supporting it. A field without one was recorded as not stated rather than inferred from context, so every count in the table is a floor on the true incidence of a provision. The fee-remission and continuation counts move the elasticity of Section 4.3 in opposite directions. A missed fee waiver means the true cost increase is larger and the elasticity smaller, and missed continuation language means the reverse. At one extreme, with no benefit merely continued and every conditional waiver covering the full fee bill, the elasticity is 0.72. At the other, it is 1.18. The continuation floor can be sized directly. On the 37 agreements whose full documents were retrieved, a detector run over the complete text finds continuation language in 20 where the excerpt rule finds it in 16, so that rule misses about 20 percent of what it should catch. Second, a separate adversarial pass of agents checked all 48 agreements against their source documents in both directions: whether each filled field is supported by its quote, and whether each field coded absent is in fact absent. Of those, 23 were corrected, almost always for a wrong article citation attached to correct quoted language. No fee-remission value changed anywhere, so verification leaves the incidence multiplier Section 4.3 applies unaffected.

Asking the reverse question, whether each field coded absent is in fact absent, found guarantee language in 9 agreements recorded as having none. That direction was added because the first pass lacked it. Checking only whether filled fields are supported cannot find a provision that was never noticed. The University of Washington's minimum appointment term, for example, had been coded absent because the article containing it was never opened. Of those 9, 6 set a minimum length for a single appointment rather than a term of support, and Table A17 reports them separately for that reason. The other 3 were multi-year commitments, which carry that count from 6 to 9.

Two limits bound what the coding can support. First, agreements posted online are usually the current one rather than the first, so 26 of the 48 are first contracts and the rest are successors. Later contracts usually build on earlier terms, so the coded terms of a successor likely overstate what its first contract gave. Second, a contract states a level, not a change. It records that the university pays a given share of the premium, never what it paid the year before. The continuation language counted in the table is the only direct evidence

in these documents on that question. It is why Section 4.3 reports the cost shock both with and without a continuation adjustment.

I.3 The Elasticity on Matched Contracts and Against Other Employers

Restricted to the contracts that both the stipend panel and the doctorate panels observe, those ratified 2014 to 2018, the elasticity cannot be pinned down (Table A18). The stipend effect then rests on 3 R1 universities, or 7 with R2 universities added, and is not distinguishable from zero in either, so the matched elasticity has no useful interval. Its point estimates bracket the pooled one: 1.47 on R1 and 0.96 on R1 and R2, against 1.08 pooled.

Table A18: Elasticity on the Contracts Observed in Both Data Sets

	All contracts	Contracts ratified 2014–2018	
	R1	R1	R1 and R2
Stipend effect, net of local prices	+6.3	+5.2	+6.6
	(2.2)	(6.2)	(5.7)
wild-bootstrap p	0.01	0.60	0.47
universities	19	3	7
Doctorates, all fields, $\tau \geq 6$	−12.2	−15.2	−11.0
	(3.5)	(2.9)	(4.6)
Doctorates, outside S&E, $\tau \geq 6$	−20.9	−39.6	−26.7
	(6.5)	(9.9)	(14.5)
universities observed at $\tau \geq 6$	21	4	8
Cost shock $\Delta \ln W$ (log points)	11.3	10.3	11.5
Elasticity, all fields	1.08	1.47	0.96
Elasticity, outside S&E	1.85	3.85	2.31

Notes: Log points times 100, standard errors clustered by bargaining unit in parentheses. First column: the stipend effect of Section 4.1 (contracts ratified 2014–2025) and the field-panel doctorate responses of Section 4.3, pooled over every contract cohort. Other columns: both estimated on the contracts ratified 2014 to 2018, the only ones with stipend reports before ratification and a doctorate response six or more years after it, with every never-organized university in the stated comparison group. Stipends: log median pay on log regional price parity, school and field-bucket-by-year effects, weighted by reports; the bootstrap is the null-imposed wild cluster bootstrap (Webb weights, 9,999 replications). Doctorates: Poisson on the field panel with cell-by-stack and cohort-by-field-by-year effects and campaign and partly treated terms. Cost shock: the continuation-adjusted row of Table 2 with its stipend part re-priced at the column's stipend effect and its premium and fee parts unchanged. Elasticity: minus the doctorate coefficient over the cost shock.

Table A19 sets the elasticity of Section 4.3 beside the published labor-demand elasticities it is compared with there.

Table A19: The Elasticity of Demand for Graduate Workers Beside Other Labor-Demand Elasticities

Workers or setting	Elasticity	Source
A. Graduate workers, this paper		
Doctorates, all fields	1.08 [0.6, 1.8]	Table 2
Doctorates, outside S&E	1.85 [0.7, 3.0]	Table 2
Doctorates, contracts of 2014–2018	1.47	Table A18
B. Workers in firms		
Median of 1,334 estimates	0.42	Lichter et al. (2015)
Mean of the same estimates	0.55	Lichter et al. (2015)
Short run, all workers	0.3	Hamermesh (2014)
Long run, once capital adjusts	1.0	Hamermesh (2014)
C. Public employers		
U.S. state and local functions, long run	0.1–0.6	Ehrenberg (1973)
U.S. state and local education, long run	0.08–0.57	Ehrenberg (1973)
Swedish municipalities, within a year	0.5	Bergström et al. (2004)
Swedish municipalities, long run	0.9	Bergström et al. (2004)
D. Minimum wages		
Median of published own-wage elasticities	0.13	Dube and Zipperer (2024)
Hungarian exporting firms	0.84	Harasztosi and Lindner (2019)
Undergraduate research assistants in university labs	0.47	Ganguli and Murciano-Goroff (2025)

Notes: Elasticities of employment with respect to its cost, in absolute value. Panel A divides the fully treated doctorate response on the field panel by the contract-priced rise in the annual cost of a student, with 95 percent intervals in brackets; the last row uses only the contracts both the stipend and doctorate panels observe. Panels B to D report the published estimates as their authors state them. Ranges are the spread the authors report across government functions or specifications.