University Research and the Market for Higher Education*

Titan Alon Damien Capelle Kazushige Matsuda

Draft Date: March 2023

Abstract

This paper develops a framework in which university research depends endogenously on competition for tuition and talented students in the market for higher education. When students are highly stratified across colleges, or when tuition rises sharply with school rank, universities spend on R&D even if the direct contribution of research to teaching is small. The model is consistent with causal evidence from a natural experiment and matches new features of the administrative microdata. It also explains why universities internally fund research with tuition, despite negligible returns to patenting. Calibrated simulations suggest quantitatively important implications for federal research and student-aid policies.

*We thank Todd Schoellman, Alexander Monge-Naranjo, Hannes Malmberg, Tommaso Porzio, Enrico Berkes, Christian Hellwig, David Lagakos, Nobuhiro Kiyotaki, Paul Niehaus as well as seminar participants at UCSD, USC, the 2022 Econometric Society's Winter Meeting, the SED, the NBER Economic Growth Workshop, and the Minnesota Macro Workshop for useful comments which improved the paper. Alon: Department of Economics, University of California San Diego (email: talon[at]ucsd.edu), Capelle: International Monetary Fund (email: dcapelle[at]imf.org), Matsuda: Department of Economics, Kobe University (email: matsuda[at]econ.kobe-u.ac.jp). All errors are our own. The views expressed therein are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

1 Introduction

The modern research university is a unique and complex institution. While primarily focused on higher education, universities also contribute to the advancement of knowledge through investment in research. Between 2000 and 2018, universities in the United States accounted for 13% of aggregate spending on research and development (R&D) and 53% of all spending on basic scientific research. Discoveries arising from university research contribute to innovation in nearly every field of science and have given rise to numerous general purpose technologies that drive long-run growth.

The coincidence of education and research within the university has multiple historical and economic roots which have been analyzed in the literature. Early American universities–especially land grant colleges–were established explicitly to provide training and conduct research that would benefit their local economy. During World War II, the federal government drastically increased funding for university research to support the war effort. The success of these war-time partnerships gave rise to the modern system of publicly funded university research.[1] While initially focused on defense, by the 1970s the federal government accounted for over 75% of university R&D funding with projects spanning healthcare, energy, and the environmental sciences. By the 1980s, patenting and the commercialization of academic research–made possible in part by the Bayh-Dole Act and a proliferation of technology transfer offices (TTOs)–emerged as important catalysts in select areas of university research, such as the biomedical and pharmaceutical sciences.

This paper proposes a new, additional motivation for university research which derives from their education mission. It develops a model in which university research expenditure depends on the competition for tuition and talented students in the market for higher education. By spending on research, universities can improve the quality of education they offer prospective students. Students benefit directly from research at their university by being exposed to frontier knowledge, and new scientific techniques. They also benefit indirectly, as top research universities attract wealthier and higher ability students, who further augment education quality through peer effects and higher spending. Consequently, the incentives for university research, and the resources available to undertake it, depend in part on equilibrium outcomes in the market for higher education.

The model explains important new features of the data on university research. It can quantitatively replicate the fact that universities which spend more on R&D are higher

[1]The case for the modern system of publicly funded academic research after World War II is famously articulated in Vannevar Bush's (1945) *Science: The Endless Frontier*.

ranked, attract wealthier and more able students, and charge higher tuition. It rationalizes why universities fund research with tuition revenue and quantitatively matches the degree of cross-subsidization observed in the data.[2] The model also explains why universities continue to spend internal funds on research despite waning government support and consistently low returns to patenting. In 2018, over 25% of university research expenditure was funded internally, while between 1991 and 2018 the median university earned patent royalties totaling less than 2% of their expenditure on research.[3]

By exploiting a natural experiment, the paper provides causal evidence in support of the model's core mechanism. A key feature of the framework is that investing in research enables a university to charge higher tuition in the future. To assess this channel, we employ plausibly exogenous variation in university R&D created by a rapid doubling of the National Institute of Health's (NIH) research budget between 1998 and 2003. We instrument the change in each university's R&D during this period with the share of all federal life science research grants they were awarded just prior to the NIH funding expansion. The results provide evidence supportive of the model's core mechanism. In the benchmark specification, a one standard deviation increase in university R&D leads to a \$1,090 increase in tuition, equivalent to roughly one-third of the cross-sectional variation in university tuition growth in the sample period.

We formalize the theory by jointly modelling university instructional and research expenditures in a general equilibrium model of the higher education sector. The model builds on existing equilibrium models of higher education, such as Epple, Romano, and Sieg (2006) and Cai and Heathcote (2022), by endogenizing university R&D decisions alongside their pedagogical ones. In each period, overlapping generations of heterogeneous households transmit human capital to their children and decide which college to attend subject to a financial friction, as in Lochner and Monge-Naranjo (2012). Universities are endogenously differentiated in the stock of knowledge–a form of institution-specific intangible capital that universities accumulate by investing in research–which they impart to students.[4] This knowledge, together with student peer effects, teacher quality, and

[2]While universities may spend on research because they care about research per se, there is still the question of how they fund such spending. The mechanism here provides an *endogenous* motivation for universities to spend on research and rationalizes why–in a competitive higher education sector–students would allow universities to divert tuition revenue to support research activities.

[3]The share of internally funded university R&D has also grown drastically relative to publicly funded R&D. While the share of publicly funded university R&D fell from around 75% in the 1970s to just over 50% today, over the same period the share of internally funded university research grew from 10% to 25%. Section 2 provides further details.

[4]The intangible capital created by research primarily includes frontier knowledge and ideas. Exposure to frontier knowledge has been shown to improve education quality and the human capital and earnings of

spending on instructional equipment determines the quality of education a university offers. Finally, a government provides need-based tuition subsidies, as in Benabou (2002), and meritocratic research grants to universities. The meritocratic nature of grant awards captures an additional motive for universities to invest in research.

The model demonstrates the interdependence of university R&D and education outcomes. Equilibrium in the market for higher education features an endogenous hierarchy of colleges differing in education quality, with two-dimensional sorting of students by ability and family income. University intangible capital induces an ordering among institutions which influences their position in the hierarchy of colleges. Those higher-up in the hierarchy charge higher tuition and attract better students and faculty. By spending on research, universities are able to improve their position in the hierarchy. The greater the positive assortative matching of students to schools, and the more tuition rises with college rank, the stronger is the incentive for universities to spend on research. An important implication is that, when students are highly stratified across colleges, universities face strong incentives to spend on research even if the direct contribution of research to teaching quality is small, as it appears in the data. Conversely, the incentive to spend on research is diminished if universities are ex ante highly unequal, since top schools are insulated from competition while lower ranked institutions face large hurdles to improve their position.

The model has important implications for public policy. It demonstrates new channels through government education and innovation policies interact to jointly shape education and research outcomes. For example, federal need-based student aid can have large effects on university research by altering the demand for education and the sorting of students across colleges. Similarly, government research grants can affect tuition and educational outcomes by altering the distribution of university intangible capital and the allocation of resources within the higher education sector. We investigate the importance of these interdependencies using a quantitative version of the model calibrated to replicate important characteristics of the U.S. higher education system, including institution level heterogeneity in tuition, student ability, government grants, and research expenditure. With the quantitative model, we compute the short-run and long-run consequences of federal R&D and tuition policies on university research and education.

The results provide evidence that the model's mechanism is quantitatively important for understanding the effects of federal research and tuition policies on university research and the market for higher education. They show that while existing student need-based

graduates (Biasi and Ma 2021). More broadly it also represents professional networks, industry recruiting, and access to advanced labs, computing, scientific methods.

financial aid programs boost education spending and decrease educational inequality, they also increase university R&D expenditures by 8.2%. Part of the rise in research spending is due to a level effect, as public subsidies increase university revenue by stimulating demand for education. The remaining effect is due to an increase in the research intensity of universities, measured by the share of resources a university allocates to R&D. The increase in research intensity is due to the indirect effects of need-based student aid which alter the nature of competition between colleges. Since children from poor families disproportionately attend lower quality schools, progressive tuition subsidies redistribute financial resources across institutions and shrink the dispersion in financial resources across schools. As a result, colleges become more similar to one another in the long-run, which leads university research intensity to rise as schools seek to differentiate themselves from competing institutions.

For federal research policies, the model predicts that public R&D grants boost university research by 68.8%, but also contribute to educational inequality by concentrating funding at top universities. In the long-run, top schools capture the majority of intangible capital associated with publicly funded university research to the disproportionate benefit of their own students. Moreover, the model predicts that current R&D subsidies lead to an increase in university research expenditure that is 7.2 percentage points below the government's share of university research funding. The crowding-out of internal university research spending occurs because the concentration of federal R&D funding at the top makes it too costly for lower-ranked institutions to compete effectively through research. The model predicts that replacing the status-quo grant system with a flat R&D subsidy would boost university research expenditure by 14.9% in the long-run by reducing the cost of research while preserving the competition between schools.

This paper contributes to the economics of science and in particular the literature investigating the determinants and contributions of university R&D (Merton 1973; Rosenberg and Nelson 1994; Stephan 1996).[5] It identifies additional incentives for university research which depend endogenously on the market for higher education. The mechanism helps explain the joint distribution of university research and teaching outcomes and the growing share of internally funded R&D cross-subsidized with tuition revenue. More broadly, by emphasizing the complementarity of university research and teaching, this paper puts forth a new perspective that may help shed light on several related economic debates, such as whether universities prioritize research to the detriment of education

[5]The recent literature focuses on the commercialization of university discoveries (Rothaermel, Agung, and Jiang 2007; Mowery et al. 2015) and the contribution of academic research to the real economy (Jaffe 1989; Mansfield 1995; Cohen, Nelson, and Walsh 2002; Ponds, Oort, and Frenken 2009; Andrews 2022).

(Willetts 2018) and what explains the large regional and international variation in university quality (Hoxby 2009; Kerr et al. 2016; Hendricks and Schoellman 2018; Urquiola 2020; Martellini, Schoellman, and Sockin 2022).

The paper also contributes to the literature on optimal public R&D subsidies by highlighting new channels through which public innovation and education policies interact (Biasi, Deming, and Moser 2020; Akcigit, Pearce, and Prato 2020). The standard view in the literature is that basic research, while socially valuable, is subject to imperfect property rights which prevent the full appropriation of the returns to new discoveries. Public subsidies are necessary since firms have little incentive to invest in basic scientific breakthroughs that others can cite and build-on (Nelson 1959; Arrow 1962; Rosenberg 1990). The recent literature seeks to quantify the extent of these market failures in order to better inform public policies supporting basic research (Akcigit, Hanley, and Serrano-Velarde 2020; Akcigit, Hanley, and Stantcheva 2022).[6] An important insight of this paper is that universities appear to have private incentives to undertake basic research and, in practice, conduct most of it. As a result, the need for government subsidies to basic research could be less than previously thought.

Finally, the model abstracts from considering ownership rights over discoveries which arise from university research. The literature shows that increasing the ownership rights and creative control of faculty researchers over their work can boost their performance, patenting, and entrepreneurship (Lach and Schankerman 2008; Azoulay, Graff Zivin, and Manso 2011; Hvide and Jones 2018). However, at the university level, the data continue to show that patent revenue is far too small to explain the majority of university spending on research (see section 2). The finding is also consistent with economic studies of the Bayh-Dole Act, which conclude that the resulting expansion in university ownership over publicly funded R&D discoveries did little to boost aggregate university research or patenting (Mowery et al. 2015). Future work should strive to better understand the joint contribution of institutional and individual incentives to university R&D and innovation.

The remainder of the paper is organized as follows. Section 2 presents important facts about university research and reports results from the natural experiment on the NIH. Sections 3 and 4 lay out the model and establish the key theoretical results. Section 5 introduces government policies and describes the quantitative model. Section 6 reviews the calibration strategy and the model fit. Section 7 contains results from the quantitative exercises. Section 8 concludes.

[6] A related literature studies the interaction of basic research subsidies, global growth, and the patterns of international development (Gersbach, Schneider, and Schneller 2013; Gersbach and Schneider 2015).

2 Research in the Higher Education Sector: Stylized Facts

In this section, we document several important economic characteristics of the market for higher education and how they relate to university research expenditure. We focus in particular on four important empirical facts. First, while the government is an important source of university research funds, over one-quarter of university R&D is paid for with internal funds. Second, while patenting is the traditional explanation for why universities spend their own money on research, patent licensing revenue in the data appears too small to be the primary driver of university R&D spending. Third, universities which spend more on R&D also deliver higher quality education, admit higher ability students, produce more scientific output, and charge higher tuition. Finally, we review causal evidence from a natural experiment on the NIH which implies spending on R&D can enable universities to charge higher tuition.

2.1 Main Sample and Data Sources

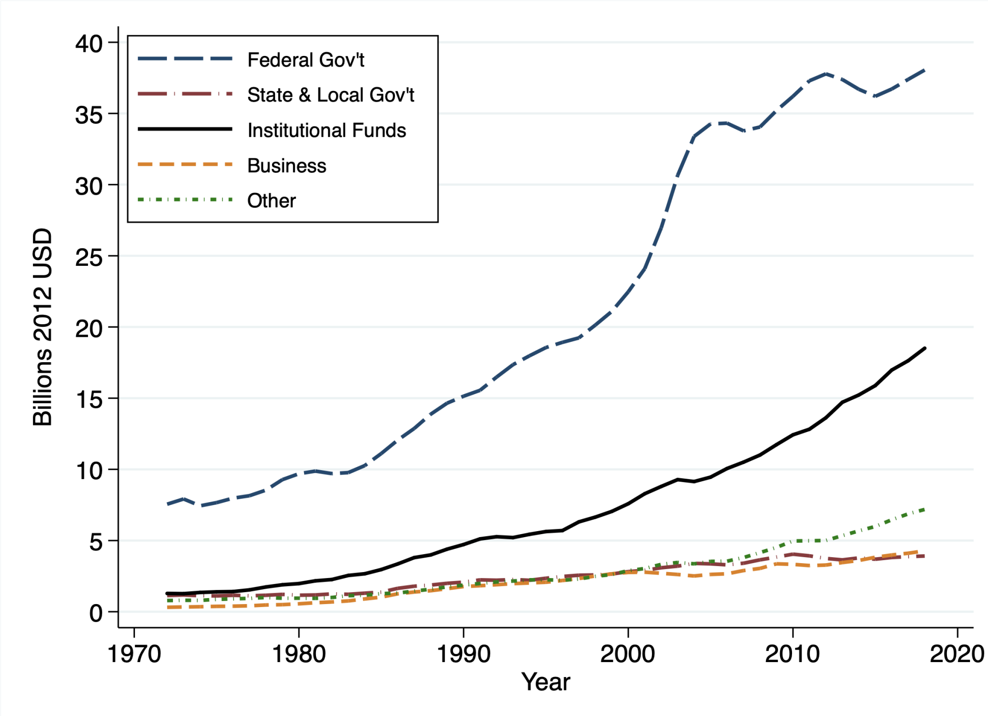
Our primary sample includes all 4-year public and private non-profit institutions in the United States. Data primarily comes from the National Center for Education Statistics Integrated Postsecondary Education Data System (IPEDS) which provides university level microdata for the universe of domestically accredited higher-education institutions. We merge in additional data from the National Science Foundation's Higher Education Research and Development Survey (HERD) on research spending by field and source of funds, the Association of University Technology Managers (AUTM) Patent Licensing Survey, and Web of Science (WoS) bibliometric data from the CWTS Leiden Rankings. Appendix C contains additional details on the underlying data.

2.2 University R&D: Stylized Facts from Administrative Microdata

The first important fact is that a large share of university R&D is conducted using internal funds. Figure 1 plots the major sources of university research funds going back to 1972, using the HERD survey of U.S. academic institutions.[7] While it confirms the well-known

[7]The HERD survey includes 916 universities representing 99.1% of the total R&D expenditure of the higher education sector and roughly 80% of full-time equivalent students. Before 2010, it includes only institutions with degree programs in science and engineering (S&E) and at least $150,000 in separately accounted for R&D expenditures, as defined by Office of Management and Budget (OMB) Circular A-21. Institutions that performed R&D exclusively in non-S&E fields were added to the survey population beginning in 2010. Universities are subject to a variety of guidelines and audits, by numerous government agencies, such as HHS, NSF, GAO, OMB, and ONR. See Council on Government Relations (2019) for additional details.

Figure 1: University Research Spending by Source of Funds



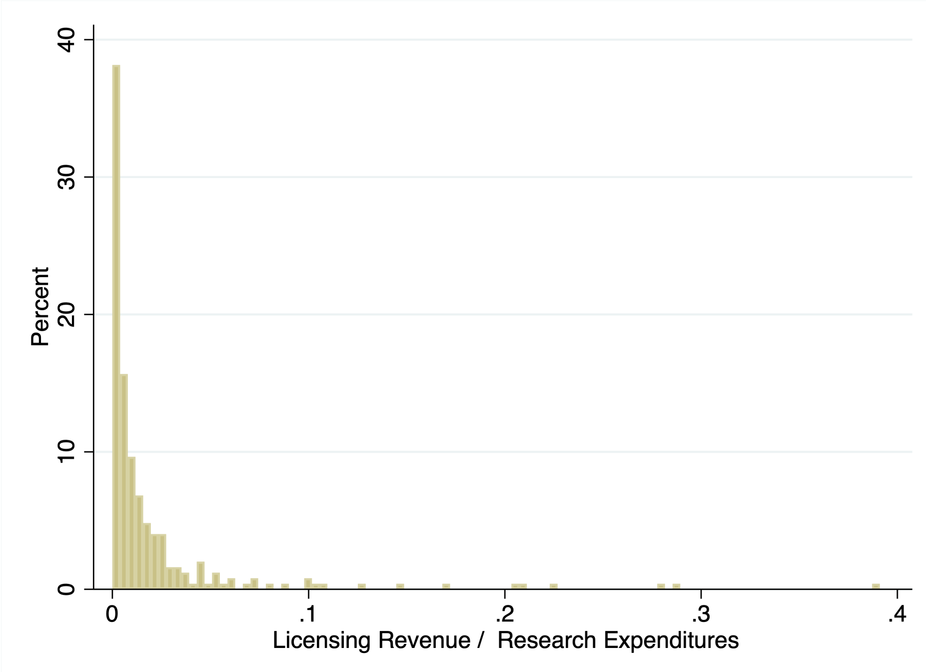
Notes: *The figure reports total university research expenditures by source of funds in 2012 U.S. dollars. Government sources include all research grants and contracts. Non-profit funding is included in the Other category. Underlying data are from the Higher Education Research and Development (HERD) survey, 1972-2018.*

fact that the federal government is the largest source of funds for university research, accounting for 52.97% of all funding in 2018, it also shows that internal institutional funds are the second largest source, accounting for 25.54% of total funding.[8] The remaining funds for university research come from state and local governments, which account for 5.46%, and sponsored research by non-profit organizations and private businesses, accounting for 6.89% and 5.97%, respectively. The final 3.17% includes non-categorized sources, such as funds from foreign governments, international institutions, and private gifts designated by the donor for research activities.

The internal funding of university research relies heavily on tuition, which constitutes the majority of unrestricted university revenue at both public and private institutions (see figure A2). The cross-subsidization of research with tuition is also evident in the details of university research accounting. As defined by OMB Circular A-21, R&D expenditures

[8]Moreover, the share of all university R&D that is internally funded has grown substantially in recent decades; in the 1970s, the federal government accounted for 75% of university research funding, while internal university funds only accounted for 10%. Similar patterns appear at both public and private institutions, and are driven both by cuts to federal funding as well as increases in internal research expenditure.

Figure 2: University Patent License Revenue relative to Research expenditures



Notes: *The figure reports the distribution of gross licensing revenue divided by total research expenditure. The underlying data source is the AUTM Licensing Activity Survey, and the sample includes all responding US universities from 1991-2018. Licensing revenue includes cumulative reported gross license income and research expenditure reports cumulative non-federal, non-industrial institutional research spending. All values are converted to cumulative real 2015 dollars using the GDP price deflator.*

from internal funds include both direct and indirect expenditures. Direct expenditures represented roughly $13 billion in 2018 and use funds from institutional accounts restricted for research purposes only, but which may have originally come from previously unrestricted revenue. Indirect expenditures totalled roughly $7 billion, and includes $1.5 billion in cost-sharing commitments and $5.5 billion of unrecovered indirect costs.[9] The latter are generally recurring facilities and administration (F&A) costs that are paid for with unrestricted operating revenue, which is composed almost entirely of tuition. The resulting link between tuition and research F&A costs lies at the center of ongoing, contentious debates between university administrators and policymakers who oversee the public funding of academic research (Council on Government Relations 2019).

In the model developed below, the cross-subsidization of university research with tuition occurs because R&D spending generates spillovers within the university which improve the quality of education it can offer its students. As a result, even if universities do not

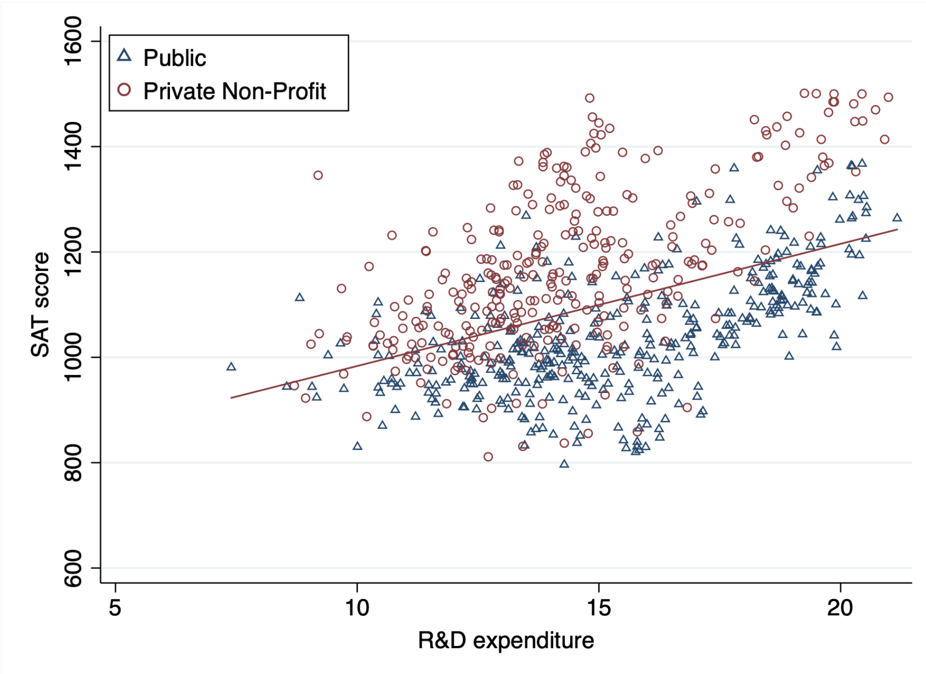
[9]Unrecovered indirect costs represents the portion of facilities and administration (F&A) costs associated with sponsored projects that is not reimbursed by the project's negotiated indirect cost rate.

care about research per se, they may still have a private incentive to spend on R&D stemming from their education mission. A cutting edge research laboratory built to conduct a sponsored experiment, can be maintained with tuition revenue so that students can learn frontier methods and train on the most technologically advanced vintages of capital.

Another motivation for university research spending could be the revenue associated with patenting scientific discoveries. This perspective is consistent with paradigmatic models of R&D at private firms and comports with the literature showing that faculty research output increases with additional ownership over their scientific discoveries (Lach and Schankerman 2008; Hvide and Jones 2018). However, despite many high profile anecdotes to the contrary, the data show that university revenue from patent licensing is a very small share of their total revenue, and is small even relative to research expenditure. Figure 2 displays the distribution of gross patent licensing revenue over total research expenditure at the university level. Between 1991-2018, the median university earned combined licensing fees totaling less than 2% of their expenditure on R&D. The size of these income streams mean that patenting cannot account for the majority of university spending on research. The finding is consistent with the literature showing that past expansions in university patenting rights (e.g. the Bayh-Dole Act) have had only negligible effects on aggregate university R&D (Mowery et al. 2015). Of course, while these data appear to preclude patents as the primary driver of university R&D spending, they do not rule out the possibility that patenting is an important incentive for individual faculty researchers or certain sub-fields of R&D, such as biomedical life sciences.

Finally, we document that university R&D expenditure is associated with increases in education quality, better students, and higher tuition in the cross-section of American universities. These correlations are consistent with the idea that universities spend on research, in part, to improve their position in the market for higher education, and that prospective students recognize the value of this spending. While there are possibly other channels through which R&D spending at a university may benefit students, our model emphasizes the knowledge spillovers it creates through teaching. Universities which spend more on R&D generate more scientific output, as measured by publications or citations (Figure A3). Students can benefit from access to this frontier knowledge and the ability to train on the most technologically advanced vintages of capital used in research labs. The mechanism is consistent with the microeconometric evidence in Biasi and Ma (2021), who show that students who graduate from universities which teach more frontier knowledge will earn more after graduation, and are more likely to enroll in graduate school and file patents later in life.

Figure 3: University Research Spending, Student Ability, and Teacher Quality

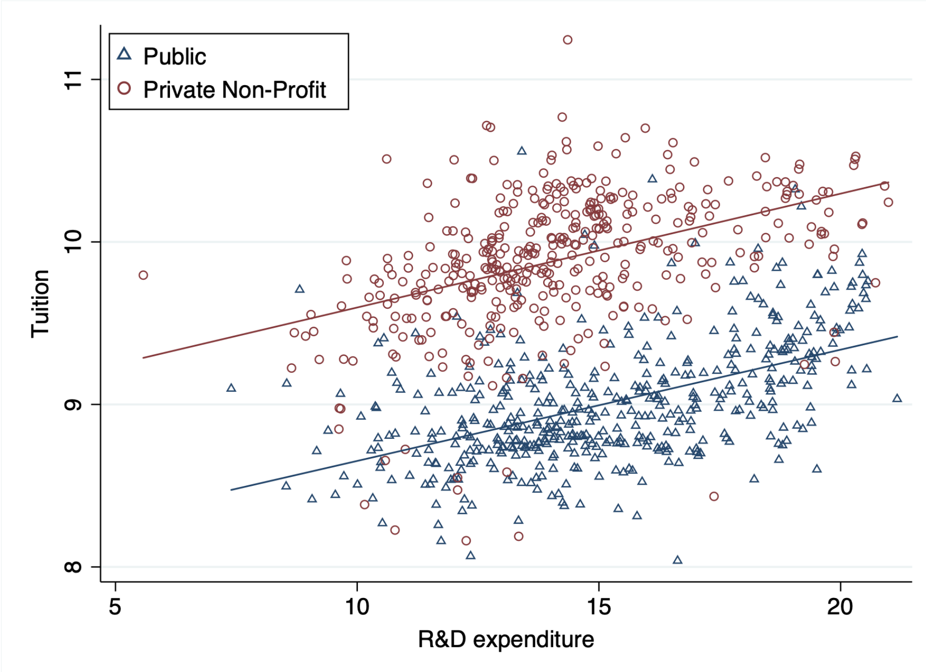


Note: *SAT scores are the sum of math and verbal scores calculated as the average of the university's reported* 25^{th} *and* 75^{th} *percentiles. Data comes from the Integrated Postsecondary Education Data System with points representing university level averages for 2012-2015.*

In the model, these quality improvements are amplified by the equilibrium sorting of wealthier and higher ability students into higher quality schools, consistent with evidence in Hoxby (2009), Lochner and Monge-Naranjo (2012), Hendricks, Herrington, and Schoellman (2021), and others. The microdata highlight the strong correlation between a university's research spending and its position in the hierarchical market for higher education. Universities which spend more on research are attended by higher ability students (Figure 3), employ better paid faculty (Figure A4), and consistently appear at the top of the rankings of best colleges (Table B1). That households recognize the value of these quality improvements is evident in the selectivity of these schools and the willingness of students to pay higher tuition to attend them. Figure 4 shows that universities which spend the most on R&D are also those who charge the highest tuition.[10] Through the lens of the model, the direct and indirect contributions of research to education quality

[10]The level difference in tuition between public and private non-profit universities can be largely attributed to state and local appropriations, much of which serve as tuition remission. Figure A5 plots the relationship between tuition and R&D when we include state and local appropriations in tuition and shows that cross-sectional elasticity remains largely unchanged. Figure A6 shows the patterns also do not appear to be driven by differences in student amenities, as proxied by student services expenditures.

Figure 4: University Research Expenditure and Tuition, by Sector



Notes: *Tuition is the average tuition revenue the university receives per full-time equivalent student, net of any university discounts or allowances. research expenditure is total university spending for activities specifically organized to produce research outcomes, including by institutes, research centers, and individuals. Data retrieved the Integrated Postsecondary education Data System (IPEDS). Points are 2012-2015 university averages in log-scale.*

can be disciplined by the cross-sectional correlations of tuition and student ability with university research expenditure (see Proposition 6).

2.3 The Impact of R&D on Tuition: Evidence from a Natural Experiment

A distinguishing feature of our framework is the presumption that by increasing its research output a university is able to charge higher tuition and attract better students in the future. The channel is important for explaining how universities are able to internally fund R&D despite negligible returns to patenting. The previous section showed evidence consistent with this view. This section provides estimates of the the causal effect of R&D on university tuition.

The quasi-natural experiment exploits a large increase in federal research funding through the National Institute of Health (NIH). Between 1998 and 2003, the federal government doubled the NIH budget for biological and life sciences research from $13.6 billion in 1998 to $27.1 billion in 2003 (Smith 2006). The NIH expansion accounts for the majority of the rise in federal funding for university research during this period, and its impact on total spending is clearly visible in the aggregate time series (see Figure 1).

To capture the exogenous variation in university R&D created by the policy change, we calculate the share of all federal research grants and contracts for life sciences research that each university won in the period preceding the doubling of the NIH budget. As an instrument, the shares are relevant since the NIH constituted the bulk of federal funding for university life sciences research even before the policy change, and it did not substantially change the award criteria after the budget increase. The shares are also exogenous provided universities did not systematically invest in life sciences research in the pre-period in anticipation of the NIH budget expansion. Our identification strategy is therefore similar to Azoulay et al. (2019), but at a coarser institutional level.

We estimate the causal impact of an increase in university R&D on its tuition growth by employing the following reduced-form econometric model of the mechanism we introduce in the next section. Denoting the tuition of university i at time t by y_{it}, we model its relationship to university R&D with the following set of simultaneous equations,

$$y_{it} = a_i + \mathbf{D}_g \times \mu_t + \boldsymbol{\gamma}_1 \, \mathbf{D}_s \times \mathrm{K}_{it} + \gamma_2 \, \mathbf{X}_{it} + \nu_{it} \tag{1}$$

$$\triangle \mathrm{K}_{it+1} = \delta_0 + \delta_1 \, \mathrm{RD}_{it} + \eta_{it} \tag{2}$$

where a_i are university fixed effects, μ_t are time fixed effects, D_s is a vector of sector (public and private) indicators, D_g is vector of group-specific indicators, and X_{it} includes covariates. The variable K_{it} represents the university-specific intangible capital which is associated with its past research and which improves its quality and desirability to students, and hence their willingness to pay. The second equation captures the fact that universities can increase this intangible capital ($\triangle \mathrm{K}_{it+1} \equiv \mathrm{K}_{i,t+1} - \mathrm{K}_{i,t}$) by spending on research, RD_{it}. These econometric specifications are analogous to equilibrium relationships in the model that we develop in section 3.

The D_g indicators allow the model to control for group-specific time trends in tuition that may be confounded with our instrument. In particular, the design allows for the possibility that there are different time trends in tuition growth for public versus private universities; for universities which receive federal grants versus those who did not; and for universities which were engaged in life sciences research (NIH funded or otherwise) before the policy intervention. The D_s indicator additionally allows the effect of R&D on tuition to differ at public and private universities.

To exploit the exogenous policy variation, we estimate the model in long-differences before and after the doubling of the NIH research budget. We instrument the increase in university R&D over this period using the shares of federal life sciences funding they were awarded before the policy change and measure its impact on the growth in tuition.[11] Specifically, taking long-differences of equation (1) and substituting in (2), we estimate

$$\triangle y_i = \bar{\mu} + \bar{\boldsymbol{\mu}}_g + \beta_1 \, \mathrm{RD}_i + \beta_s \, \mathrm{D}_s \times \mathrm{RD}_i + \beta_2 \, \triangle \mathbf{X}_i + \epsilon_{it} \tag{3}$$

where β_1 is the main parameter of interest, capturing the net impact of R&D on tuition through δ_1 and γ_1. The independent variable $\triangle y_i$ is the change in university i's net tuition from 1993-1997 to 2004-2008. The federal life sciences grant shares used as instruments are calculated with respect to the same pre-period, 1993-1997. The main dependent variable, RD_i, measures each university's cumulative R&D expenditure between 1998-2003, the period when the NIH research budget doubled. All regressions use robust standard errors, clustered at the state level, and allow for state-specific trends in tuition growth to control for changes in state level tuition policies.

The first column of Table 1 reports the estimation result of the specification (3) without the group-specific and sector-specific dummies. The results show that universities with the largest exogenous increase in R&D expenditure are also those with the largest increase in tuition. The impact is statistically significant and the magnitude of the effect is economically substantial. The result indicates that a one standard deviation in R&D expenditure at a representative college would increase tuition by $1,090.24 in the long-run. The estimated effect is economically large given that average weighted tuition growth during the sample period is $4,714, with a standard deviation of $3,417.

The results of the benchmark model provide evidence of a causal relationship between university R&D spending and tuition. However, limitations on the exogenous variation made available by the NIH natural experiment can raise concerns about the interpretation of the results. One issue is that it does not distinguish between public and private institutions. It is well known that average tuition growth differed across the two groups during the sample period. Moreover, many public universities face additional constraints in setting tuition and selecting students that may affect their ability to capture the returns to R&D compared with private universities. Another issue is that the NIH shock occurs at a common point in time for all universities, which means it can be correlated with other market changes unfolding at the same time. For instance, it is possible that initially

[11] Ideally, these exercises would include additional specifications to measure the exogenous impact of research on subsequent student ability as well. Unfortunately, this is not possible given that our data only reports proxies for student ability beginning after the NIH policy intervention.

better research universities experienced larger tuition increases because of a broad rise in inequality across schools that may be unrelated to R&D or the NIH shock.

To begin addressing these concerns, column (2) reports the results of the model that allows the effect of R&D and trend growth in tuition to differ at public and private universities. The statistically significant effect of research on tuition is preserved, though is moderately smaller; a one standard deviation increase in R&D leads to a $764.11 increase in tuition. The point estimates also provide evidence that the effect of R&D on tuition is $476.97 larger at private institutions which, though not statistically significant, is consistent with the idea that private universities respond more strongly to private incentives in the market for higher tuition.[12] The result also confirms the well known fact that private universities saw larger growth in tuition in recent decades, with their average tuition increasing by $2,744.88 more than at public universities during the sample period.

Column (3) reports results for the model which allows for the possibility that universities who receive federal grants, contracts, and appropriations (not including Pell grants) experience faster tuition growth than those who do not. It shows that universities which were already receiving federal funding in the 1993-1998 pre-period did indeed see their tuition grow by $1,813.88 more on average during the sample period. Column (4) allows for different trends in tuition growth at universities engaged in Life Sciences R&D before the policy expansion, but finds no statistically significant effect. Importantly, the impact of R&D spending on tuition remains largely unchanged and significant in both of the additional specifications.

Finally, column (5) attempts to control for the effect of growing inequality across colleges that may be driven by factors other than research spending. If the best universities were also the leading research universities in the pre-period, then our instrumental variation may be confounded with other macroeconomic factors that increased inequality across colleges during this period, like rising household income inequality (Capelle 2020; Hendricks, Herrington, and Schoellman 2021; Cai and Heathcote 2022). While it is not straightforward to control for distributional effects in reduced-form, we proxy them by including a university's lagged (1988-1992) tuition level as an independent variable to capture the idea that initially better universities may also be those experiencing larger tuition increases. Note that this is in addition to already controlling for persistent differences in the *level* of tuition across colleges, captured by the school specific fixed effect

[12] One reason that sectoral differences in the effect of R&D on tuition (β_S) may not be precisely estimated is that private non-profit universities often occupy both the top and bottom of the college hierarchy (see Cai and Heathcote 2022), suggesting non-linearities that may not be well captured by the empirical model.

Table 1: The Estimated Causal Impact of University R&D Expenditure on Tuition

	(1)	(2)	(3)	(4)	(5)	(6)
R&D Expenditure (β_1)	1,090.24	764.11	764.81	726.46	361.81	430.33
	(366.82)	(217.35)	(217.51)	(207.87)	(281.34)	(258.56)
R&D $\times$ Private Non-Profit (β_s)		476.97	465.72	481.00	-188.18	-203.45
		(533.22)	(532.76)	(536.53)	(537.31)	(520.53)
Private Non-Profit Trend (μ_g)		2,744.88	2,770.18	2,775.87	-47.22	-138.14
		(383.36)	(383.76)	(407.37)	(269.70)	(286.58)
Federally Funded Trend (μ_g)			1,813.88			25.54
			(394.74)			(461.83)
Life Science Trend (μ_g)				110.40		-213.68
				(156.69)		(131.88)
Lagged Pre-Period Tuition $y_{i,t-1}$					0.64	0.65
					(0.04)	(0.05)
R-squared	0.18	0.42	0.42	0.42	0.60	0.60
Observations	1,565	1,565	1,562	1,565	1,565	1,562
State FE	yes	yes	yes	yes	yes	yes

Notes: *The coefficients in the table report the estimated causal impact of a one-standard deviation increase in university R&D expenditure on its tuition using the model in (3). The independent variable is the change in university tuition between 1993–1997 and 2004–2008. The dependent variable is cumulative university R&D expenditure from 1998 to 2003, normalized by its cross-sectional standard deviation. The instrumental variable is the share of federal life sciences grants won by each university just before the doubling of the NIH grant budget (1993-1997). All regressions are weighted by a university's number of full-time equivalent (FTE) students over the sample period. All specifications include state-specific trends in tuition growth to control for changes in state level tuition policies. Parenthetical values report robust standard errors clustered at the state level.*

a_i in equation (1). The results in column (5) show that for every $1 higher a university's tuition was in the pre-period, its average tuition grew by $0.64 more during the 1993-1998 to 2003-2008 sample period. The inclusion of lagged tuition cuts the estimated effect of R&D on tuition in half and causes it to lose its statistical significance, while also flipping the point estimates of the sectoral effects in unintuitive ways. Column (6) estimates the complete robustness model. It shows a positive and statistically significant effect of R&D on tuition at the 10% level, though the point estimate is roughly half the benchmark effect, and lagged tuition continues to play a prominent role.

Taken together, the results in Table 1 provide evidence of a significant causal relationship between a university's R&D spending and its tuition growth. The magnitude of the effect appears economically important. The empirical strategy accounts for sectoral effects, confounding trends in tuition growth across colleges, and attempts to control for other macroeconomic factors which may have widened institutional inequalities correlated with R&D at the same time. While more research is required for a definitive conclusion, the results above already provide important suggestive evidence and offer a compelling case for further study. The following sections develop a model consistent with this evidence that can also account for the equilibrium impact of outside factors–such as income inequality and sponsored research grants–on educational inequality and university incentives to spend on research.

3 A Model of University R&D in the Higher Education Sector

We develop a general equilibrium model of the higher education sector with universities that engage in teaching and research while competing for talented students and tuition revenue. The economy is populated by heterogeneous households and colleges. In each generation, parents imperfectly transmit human capital to their children and together decide which college their child will attend. Colleges choose the pool of students to admit and how to allocate resources between teaching and research activities.

3.1 Households

The economy is populated by overlapping generations of dynasties $i \in [0, 1]$. Each period corresponds to one generation. At time t, a household of dynasty i is characterized by the parents' human capital, h_{it}, and the human capital of their child at the end of high school, z_{it}. Households choose consumption, c_{it}, and a college quality for their child, q_{it}, which will determine their human capital as an adult. When no confusion results, we drop the subscripts and denote the state variables of the next generation with a prime "

′ ". Letting β represent the intergenerational discount factor, the household objective can be formulated recursively as

$$U(h,z) = \max_{c,q} \left\{ \ln(c) + \beta \mathbb{E}\left[U(h',z')\right] \right\} \tag{4}$$

subject to the household budget constraint

$$c + p(q,z) = wh \tag{5}$$

where w is the exogenous effective wage rate and $p(q,z)$ is the endogenous tuition schedule determining the cost of sending a student of ability z to a school of quality q. Upon graduating college, children become adults and enter the labor market with human capital h' that depends on the quality of college they attended and their pre-college ability,

$$h' = zq^{\alpha} \tag{6}$$

where α parameterizes the earnings elasticity with respect to college quality. We assume student ability z is known to both households and colleges and i.i.d. log-normally distributed $\ln z \sim \text{i.i.d.}\, \mathcal{N}\left(-\sigma_z^2/2, \sigma_z^2\right)$ where σ_z^2 is the population variance of student ability.

3.2 Universities

There is a mass one of heterogeneous colleges indexed by $j \in [0,1]$. We assume that all colleges are of the same size and each admits a continuum of heterogeneous students. The primary activity of a college is to educate its students. The quality of education, q, a college offers depends on its teaching expenditure per student (e_T), a student peer effect ($\bar{z}$), and the university's intangible capital (k). Formally,

$$q = k^{\omega_k} \bar{z}^{\omega_z} e_T^{\omega_e} \tag{7}$$

where ω_x parameterizes the elasticity of college quality with respect to input x. As in Cai and Heathcote (2022) we assume the technology exhibits constant returns to scale so that the size of a college is irrelevant.

The direct dependence of education quality on a university's intangible knowledge through ω_k represents the value of exposing students to frontier ideas and methodologies. Biasi and Ma (2021) provide micro-econometric evidence suggesting that $\omega_k > 0$, since universities which include more frontier knowledge in their academic curricula produce better educational outcomes, including higher earnings for their graduates.[13] In our model, such knowledge capital is a by-product of academic research. Universities spend on R&D in part to accumulate this intangible capital that improves the quality of education they can deliver to students. We posit a research technology which builds on a university's existing knowledge base through investment in R&D. Formally, the law of motion for a university's knowledge capital is given by,

$$k' = k^{\gamma_k} e_R^{\gamma_e} \tag{8}$$

where e_R is university expenditure on research and $\gamma_k < 1$ captures the notion that a university's knowledge capital is persistent, but depreciates over time.[14]

Finally, consistent with a wealth of empirical evidence, we allow the peer effect within a university, $\bar{z}$, to depend on student heterogeneity in both ability and socioeconomic background, so that: [15]

$$\ln \bar{z}(\phi; p) = \mathbb{E}_{\phi(.)}[\ln(z)] - \sigma_u^2(\phi; p) \tag{9}$$

The first term captures ability peer effects using a geometric average of student abilities within the college, where $\phi(.)$ denotes the endogenous distribution of abilities among the students admitted by the university. The second term, $\sigma_u(\phi; p)$, represents the indirect costs of socioeconomic heterogeneity in the student body. It captures the idea that the more heterogeneous the class in terms of student ability and economic background, the more difficult it is for a college to deliver a given education quality to its students. We model σ_u^2 as the within-college variance of a weighted average of (log) tuition $\sigma_u^2(\phi; p) = \frac{\Omega}{2} V_{\phi(.)}(\ln p(q, z))$ where Ω is an aggregate constant defined in appendix (45). Defining σ_u^2 in this manner ensures tractability of the equilibrium assignment rule which governs the sorting of students across colleges, by ability and family income.

[13] While we maintain the perspective that research improves education through the discovery of new knowledge, or building expertise among faculty, the model is consistent with alternative interpretations for how a university's spending on research augments its education quality, such as through reputation effects, network effects, or other forms of intangible capitals.

[14] We are not the first one to use a multiplicative law of motion for knowledge, see for example Hall and Hayashi (1989) and Klette (1996). The technology can be extended to allow for the slow diffusion of research discoveries across colleges. Higher levels of diffusion generally reduce the university's private incentive to invest in R&D. We omit detailing this process due to a lack of good data to discipline its importance. Instead, we capture these effects through the more general depreciation factor γ_k.

[15] See Epple and Romano (2010) and Sacerdote (2014) for a review of the empirical literature. The importance of these effects motivated early approaches to modelling universities as "club goods", see Epple and Romano (1998).

Following the literature , we assume colleges value the quality of education they deliver to their students as in Epple, Romano, and Sieg (2006) and behave competitively as in Cai and Heathcote (2022). We extend these static frameworks by including a research investment decision that makes the university problem dynamic. A natural extension is to assume that colleges value the discounted sum of the quality of education offered to current and future students. Letting the instantaneous flow payoff for a college delivering education of quality q be $\ln q$, the university problem can be formulated recursively as

$$V(k) = \max_{\phi,\, e_T,\, e_R} \ \ln q \,+\, \beta V(k') \tag{10}$$

subject to the education technology in (7), the peer effect (9), the research technology (8), and a flow budget constraint given by

$$\mathbb{E}_{\phi(.)}[p(q,z)] = e_R + e_T \tag{11}$$

where e_R and e_T are university research and teaching expenditures, respectively, and $\phi(z)$ represents the composition of the admitted student body as a density of student ability z. The university's tuition revenue, $\mathbb{E}_{\phi(.)}[p(q,z)]$, is determined by the quality of education it offers, the composition of its student body, and the equilibrium tuition schedule $p(q,z)$.

It is worth noting that while both teaching and research expenditure will increase education quality, university research plays an additional role. The intangible capital k' produced by research induces an ordering among institutions which shapes the hierarchy of colleges that prevails in the model's equilibrium. Universities which are higher on the ladder of colleges can charge more tuition and attract better students, further augmenting the quality of education they offer. An assumption of our model is that a university's intangible capital k', as its sole dynamic state variable, fully determines its position in next period's hierarchy of schools.

3.3 Market Clearing and Equilibrium

We focus a steady-state equilibrium in which all cross-sectional distributions are time invariant and prices are constant. An equilibrium consists of a household value function $U(h,z)$ and associated policy functions $\{c(h,z), q(h,z)\}$; the university value function $V(k)$ and associated university policy functions $\{\phi(z|k),\, e_R(k),\, e_T(k),\, q(k)\}$; and an equilibrium tuition price schedule $p(q,z)$ and measures $\mu_h(h,z)$ and $\mu_c(k)$ such that, given the tuition price schedule, $\{U(h,z), c(h,z), q(h,z)\}$ solves the household problem in (4),

$\{V(k), \phi(z|k), e_R(k), e_T(k)\}$ solves the university problem in (10), and markets clear for all combinations of quality $Q \subset \mathbb{R}^+$ and student ability $Z \subset \mathbb{R}^+$ so that

$$\int \mathbb{1}\left[q(h,z) \in Q\right] \mathbb{1}\left[z \in Z\right] d\mu_h = \int \left(\int_{z \in Z} d\phi(z|k) \right) \mathbb{1}\left[q(k) \in Q\right] d\mu_c \tag{12}$$

where $q(h,z)$ is the optimal education demand of a household (h,z) and policy $q(k)$ is the optimal education quality supplied by a university of type k.

4 Equilibrium Research and the Market for Higher Education

In this section, we highlight key equilibrium properties of the model which help elucidate the economic interdependencies between university R&D expenditure and the market for higher education. The results also motivate our calibration strategy and provide intuition for the results of our counterfactual simulations. All proofs can be found in appendix D.

4.1 Market Structure of the Higher Education Sector

The model's equilibrium features an endogenous hierarchy of colleges which differ in their education quality. Students are stratified across institutions by family income and student ability, with the highest ability and wealthiest students sorting predominantly into the highest quality colleges. This stratification is the result of optimal household decisions taking the tuition schedule as given. In turn, the equilibrium tuition schedule is determined in equilibrium by household choices and the admission policies chosen by colleges.

A college which decides to offer education of quality q must choose the optimal mix of inputs, in terms of expenditures and the type of students admitted, to best deliver its targeted quality q. Given that teaching expenditures and student ability are substitutable, a college trades off lower student ability for higher tuition. This trade-off is reflected in the first-order conditions of the college problem with respect to the density over student types and to expenditures. The unique tuition schedule consistent with colleges being at an interior solution is given by the following proposition.

Proposition 1. *The equilibrium tuition schedule clearing education markets is given by,*

$$p(q,z) = \bar{p}\, q^{\frac{1}{\epsilon_1}} z^{-\frac{\epsilon_2}{\epsilon_1}} \tag{13}$$

where $\bar{p}$, ϵ_1 and ϵ_2 are endogenous, non-negative elasticities, all defined in Appendix D.2. $\bar{p}$ is the endogenous intercept of the tuition schedule and determined to balance the mean of the supply and demand of quality of universities, which is discussed in more detail in section D.6. ϵ_1 and ϵ_2 are determined to balance the slopes of the supply and demand, which is discussed later. Consistent with the data and existing literature, university tuition increases with quality q and pricing exhibits third-order price discrimination by student ability. The extent to which higher ability students pay lower out-of-pocket tuition to attend institutions of a given quality q is given by the ratio ϵ_2/ϵ_1 which reflects the value of student ability relative to tuition from the perspective of colleges. As can be seen from equation (19), colleges value tuition not only because they fund current teaching expenditures, but also because they can be used for research expenditures.

Given the tuition schedule (13), households choose where to send their child. This is equivalent to choosing how much to spend on colleges because the tuition schedule is strictly increasing in quality. Since the elasticity of intergenerational substitution is unitary, and technologies are log-linear, and so households spend a constant share of their income on tuition.

Proposition 2. *Households spend a constant share, s, of their income on tuition with*

$$s = \beta\alpha\epsilon_1 \tag{14}$$

Households spend more on education when they are more altruistic (β), when a college education bring higher monetary returns (α), or when the supply of education quality is more elastic (ϵ_1).

Given the equilibrium tuition schedule and the optimal spending of households, one can obtain the equilibrium sorting rule, which gives the quality of school attended by each student, depending on their family background and their own ability (h, z).

Proposition 3. *The equilibrium student sorting across colleges is given by,*

$$q(h, z) = \left(\frac{swh}{\bar{p}}\right)^{\epsilon_1} z^{\epsilon_2} \tag{15}$$

This two-dimensional sorting rule of students is fully characterized by several endogenous variables common to all households: $sw/\bar{p}$, ϵ_1 and ϵ_2.[16] The first term, $swh/\bar{p}$, captures the real education spending of a household with parental income wh. Parameters

[16]We analyze the model's transition dynamics in section 7.3. One particularly tractable feature of our model is that along the transition path, the sorting rule and price schedule in Proposition **??** maintain their functional forms, with aggregate parameters $\bar{p}_t$, ϵ_{1t}, ϵ_{2t} varying over-time to ensure market clearing.

ϵ_1 and ϵ_2 capture the elasticity of the sorting rule with respect to family income and student ability, respectively. As discussed above, they reflect the relative valuation of tuition and student ability by colleges. As colleges' valuation for tuition increases because either teaching or research expenditures become more valuable, student sorting by family background strengthens.

4.2 University R&D and the Cross-Subsidization of Research

University research expenditure depends on the equilibrium on the market for higher education. In the model, university research is funded through cross-subsidization using tuition revenue derived from teaching activities (recall the university budget constraint in (11)). This form of funding corresponds to the internally-funded university R&D in the data (see Figure 1). The novelty of this mechanism is that universities spend on R&D in equilibrium, even in the absence of patents or government research grants.

Proposition 4 characterizes the share of internal revenue that universities allocate to research e_R in equilibrium. We will refer to it simply as the research share or as the cross-subsidization rate. The universities incentives to spend on research are shaped by equilibrium features of the market for higher education. Importantly, the research share depends on the *dispersion* in tuition and student ability on the one hand and on the dispersion of universities' intangible capital on the other.

Proposition 4. *The share of tuition revenue that universities spend on research is given by*

$$s_R = \frac{\beta\gamma_e(\Sigma_q/\Sigma_k)}{(1-\beta)\omega_e + \beta\gamma_e(\Sigma_q/\Sigma_k)} \tag{16}$$

where Σ_x is the standard deviation of $\ln x$ and

$$\frac{\Sigma_q}{\Sigma_k} = \omega_k + \omega_e\frac{\Sigma_R}{\Sigma_k} + \omega_z\frac{\Sigma_{\bar{z}}}{\Sigma_k} \tag{17}$$

where $R \equiv \mathbb{E}_{\phi(\cdot)}\left[p(q,z)\right]$ is the average net tuition a university receives from its students.

Equation (16) shows that university spending on research is increasing in the steepness of the college quality-ladder, measured by Σ_q/Σ_k. The ratio is a sufficient statistic for the endogenous universities incentives to invest in research. It summarizes the extent to which a university with more intangible capital k can deliver a better education quality q. Indeed, given the log-normality of the model, it corresponds to the equilibrium cross-sectional elasticity of q with respect to k.[17]

To better understand the forces which determine the steepness of the college quality-ladder, equation (17) shows how Σ_q/Σ_k can be further decomposed into three components. The first, ω_k, captures the direct contribution of research to teaching quality in the education technology (7). The latter two terms capture incentives which flow from university competition for tuition and talented students. They represent the fact that leading research universities attract better students, an effect summarized by $\omega_z \times \Sigma_{\bar{z}}/\Sigma_k$, and can charge higher tuition, captured by $\omega_e \times \Sigma_R/\Sigma_k$, both of which further improve education quality.

An important implication of equation (17) is that the relationship between education quality and university research capital is more than a feature of the education technology (ω_k); it also depends on how students endogenously sort across schools by ability and family background. When students are highly stratified in the higher education market, university incentives to spend on research may be much larger than what is implied by the direct contribution of research to teaching quality alone. The model can even rationalize university R&D even as the direct contribution of research to teaching quality becomes vanishingly small (e.g. $\omega_k \to 0$) by serving as a coordination device for high ability and wealthy students to congregate at the same colleges.

Another important implication of equation (17) is that the strength of the incentives to spend on research depend inversely on the dispersion of knowledge intangible capital across colleges. Intuitively, when institutions are highly unequal in their intangible capital, top schools are protected from competition and lower ranked schools are discouraged from competing. As a result, all types of universities invest less in research. Because knowledge capital changes slowly over time, this mechanism will play out only in the long-run.

Equation (17) also helps understand how we quantify in section 6 the empirical importance of these incentives on university research. While a university's knowledge capital k may be difficult to directly measure, the model allows us to link the sufficient statistic Σ_q/Σ_k to other characteristics of the university which are observable. From the research technology in equation (8), a university's intangible capital will be closely related to its research spending. The component $\Sigma_{\bar{z}}/\Sigma_k$ in equation (17) is therefore closely associated with the cross-sectional elasticity of student ability with respect to university research,

[17] In more technical terms, Σ_q/Σ_k measures the percent increase in education quality which results from an investment in R&D that increases university intangible capital by 1%.

displayed in Figure 3. Similarly, the term Σ_R/Σ_k is proportional to the cross-sectional elasticity of tuition with respect to research expenditure displayed in Figure 4.

While Proposition 4 provides conceptual insight and helps link the model to the data, one drawback of the characterization is that most of its elements are themselves highly endogenous objects. To provide a more fundamental characterization of how the model works, Proposition 5 expresses the college quality-ladder determining university research in terms of the model's state variables: the dispersion of z, h and k. The alternative characterization, while adding insight, still depends on the endogenous parameters governing the market structure of the higher education system, ϵ_1 and ϵ_2, whose determination we discuss in greater detail below.

Proposition 5. *The equilibrium college quality-ladder depends on the market structure of the higher education sector and can be expressed as*

$$\frac{\Sigma_q}{\Sigma_k} = \sqrt{\epsilon_1^2 \left(\frac{\Sigma_h}{\Sigma_k}\right)^2 + \epsilon_2^2 \left(\frac{\sigma_z}{\Sigma_k}\right)^2} \tag{18}$$

The first term in equation (18) captures the dispersion in household expenditures on education and, through ϵ_1, the extent to which they accrue to colleges of different qualities. The second term measures the variation in student abilities and, through ϵ_2, the extent to high ability students congregate at high quality colleges. University research spending is higher when household education expenditures are more unequal (e.g. high Σ_h) or if there is large variation in the abilities of students (e.g. high σ_z). Conversely, it is low when colleges' intangible capital are ex-ante highly unequal (e.g. high Σ_k), since universities are farther from their competing institutions and so overtaking them in the hierarchy of colleges would require larger and more costly investments in research. Proposition 5 also shows that universities spend more on research when the elasticities of education quality to tuition revenue and ability are high (e.g. high ϵ_1 and ϵ_2).

Propositions 4 and 5 also show how the supply and demand for education shape the equilibrium quality-ladder which incentivizes university research. On the demand side, heterogeneous households (h, z) demand heterogeneous education qualities. From the sorting rule in Proposition 3, a household of type (h, z) will demand education quality $\ln q = const. + \epsilon_1 \ln h + \epsilon_2 \ln z$. Due to the log-normality of the model, we can summarize the distribution of education quality demand using the second moments. Letting Σ_q^D

denote the distribution of household demand for education quality,

$$\Sigma_q^D = \sqrt{\epsilon_1^2 \Sigma_h^2 + \epsilon_2^2 \sigma_z^2}$$

Similarly, Proposition 4 shows that the distribution of education quality supplied is,

$$\Sigma_q^S = \omega_k \Sigma_k + \omega_e \Sigma_R + \omega_z \Sigma_{\bar{z}}$$

Market clearing in the higher education sector (equation (12)) requires that the distribution of education demand, Σ_q^D, equals the distribution of education supplied, Σ_q^S, so that

$$\sqrt{\epsilon_1^2 \frac{\Sigma_h^2}{\Sigma_k^2} + \epsilon_2^2 \frac{\sigma_z^2}{\Sigma_h^2}} = \omega_k + \omega_e \frac{\Sigma_R}{\Sigma_k} + \omega_z \frac{\Sigma_{\bar{z}}}{\Sigma_k}$$

which elucidates the two equivalent characterizations of the college quality-ladder in Propositions 5 and 6. The parameters ϵ_1 and ϵ_2, which shape the sorting rule and tuition schedule, are determined in equilibrium to balance supply and demand and clear education markets. More specifically, parameter ϵ_1 adjusts to clear the supply and demand for education in the above equation, while ϵ_2 is simultaneously determined by the marginal rate of substitution between student ability z and monetary inputs, given by equation 3d

$$\frac{\epsilon_1}{\epsilon_2} = \frac{\omega_e + \beta(1-\beta)^{-1}\gamma_e(\Sigma_q/\Sigma_k)}{\omega_z} \tag{19}$$

Although the solution to the model can be fully characterized analytically, ϵ_1 and ϵ_2 remain only implicitly defined since they depend on the (endogenous) steepness of the quality-ladder. The main difficulty is that the sorting of students is two-dimensional. To provide additional insight into the model's economic mechanisms, the next section considers a special case without peer effects ($\omega_z = 0$) in which the model's equilibrium can be fully characterized in terms of exogenous variables.

4.3 An Equilibrium without Peer Effects

Proposition 6 characterizes the model's equilibrium in the absence of peer effects.

Proposition 6. *When $\omega_z = 0$, there are no peer effects in education, so $\epsilon_2 = 0$ and*

$$\epsilon_1 = \omega_e + \omega_k \frac{\Sigma_k}{\Sigma_h} \tag{20}$$

where the steady state dispersion of college intangible capital is given by,

$$\Sigma_k = \frac{\gamma_e}{1-\gamma_k}\Sigma_h \tag{21}$$

and the steady state distribution of household human capital is given by,

$$\Sigma_h^2 = \frac{\sigma_z^2}{1-\left(\alpha\left(\omega_e+\omega_k\frac{\gamma_e}{1-\gamma_k}\right)\right)^2} \tag{22}$$

In the case without peer-effect, knowledge intangible capital and final goods are the only two inputs to produce education quality. Universities have no incentives to attract talented students and as a result there is perfect sorting across colleges based on family income ($\epsilon_2 = 0$).[18]

These expressions show important equilibrium interactions between the market for higher education and university research. For instance, consider a change in the research technology, such as a rise in γ_k or γ_e, which increases R&D expenditure and leads to an increase in the dispersion of college intangible capital Σ_k in equation (21). From equation (20), we see that the increased dispersion in college intangible capital will decrease the price elasticity of demand, leading to an increase in household education expenditures and greater stratification of students by family background in higher education. In the long-run, the increase in university research spending leads to greater inequality in educational outcomes and more earnings inequality, steepening the college quality-ladder further and reinforcing university incentives to spend on research.

This example illustrates the role of universities as engines of human capital accumulation and innovation, but also of inequality. Equation (22) shows show the education system amplifies initial differences between students, σ_z^2, leading to larger variation in adult human capital. The equation also demonstrates how this amplification depends on the returns to education through α, differences in educational expenditures per students through ω_e, and the contribution of research through $\omega_k\frac{\gamma_e}{1-\gamma_k}$. The final term captures an important equilibrium feedback in the model whereby university research today also influences the future demand-side of the education market.

This section has abstracted from government research and tuition policies. In the next section we formally introduce both policies into a quantitative version of our model, that

[18] Plugging the results of Proposition 6 into the formulas contained in Propositions 1 through 5 provides expressions for the equilibrium tuition schedule, household education expenditure, university research spending, and the sorting rule determining the stratification of students across the college quality-ladder.

we use to investigate the empirical importance of our model's mechanism.

5 A Quantitative Model with Government Policies

A novel implication of the model is that university research and teaching outcomes are co-determined in equilibrium. To understand the quantitative and practical importance of these interdependencies, we use the model to assess how the mechanism influences the impact of government research and tuition policies. We begin by introducing government policies and a number of additional quantitative extensions before presenting the calibration strategy.

5.1 Government Research and Tuition Policies

The government implements two types of policies in the higher education sector: merit-based research grants and need-based student financial aid. Government policies are funded by progressive income taxes, as in Heathcote, Storesletten, and Violante (2017), with excess revenues rebated to households through a linear non-distortive consumption rebate.[19] We take the prevailing tax schemes as given and do not consider the optimal design of government R&D taxation policies within our counterfactuals, as in Akcigit, Hanley, and Stantcheva (2022).

Federal tuition policies consist of progressive need-based financial aid. We augment the household budget constraint in (5) so that the out-of-pocket college expense for a household with income y is given by $\psi(y) \times p(q, z)$, where $\psi(y)$ represents the government tuition subsidy. While such aid is, in practice, distributed through a variety of policy instruments, we follow the parsimonious approach of Benabou (2002) and model the net effect of these policies in reduced-form using the two-parameter policy schedule

$$\psi(y) = \frac{y^{\tau_n}}{1 + a_n} \tag{23}$$

where τ_n is the rate of progressivity of the need-based subsidy and $1 + a_n$ is the intercept determining the overall level of support.

As with tuition, government subsidization of university research is administered through several different programs and agencies, including the National Institute of Health, the Department of Defense, NASA, the National Science Foundation, and others.[20] As above,

[19]Specifically, $\{a_y, \tau_y\}$ parameterizes the tax-system such that after-tax income is $(1 - a_y) \times (wh)^{1-\tau_y}$.
[20]See National Science Board (2018), *Expenditures and Funding for Academic R&D*.

we model these programs parsimoniously through a reduced-form allocation rule for government grants that captures both the level of subsidy and its distribution across institutions. Specifically, we augment the university budget constraint (11) so that government grants cover a fraction $1 - G(k)$ of a university's research (but not teaching) expenditures. The dependence of research subsidies on k reflects the meritocratic nature of government grant making and allows us to match the distribution of federal research funds observed in Figure A7. We parameterize the government's grant policy schedule using the two-parameter family

$$G(k) = \bar{G}k^{-\tau_G} \tag{24}$$

where $\bar{G}$ and τ_G capture the average subsidy and its distribution across universities.

5.2 Quantitative Extensions

Before simulating the model, we introduce a number of additional quantitative features to make full use of the detailed institution-level microdata available to us and account for other important economic forces pertaining to university research that have been emphasized in the literature.

Faculty in Teaching and Research. We generalize the university's research and teaching technologies to account for the contribution of faculty. Specifically, the university teaching technology (7) becomes $q = k^{\omega_k}\bar{z}^{\omega_z}e_T^{\omega_e}\bar{h}_T^{\omega_h}$ and the research technology (8) becomes $k' = k^{\gamma_k}e_R^{\gamma_e}\bar{h}_R^{\gamma_h}$, where the contribution of faculty depends on their average human capital, $\bar{h}_x = \mathbb{E}_{\mu_x(\cdot)}[\,h\,]$, where $\mu_x(\cdot)$ is the endogenous distribution of faculty chosen by the university to perform task $x \in \{R, T\}$. Allowing the university to choose different compositions of research faculty, $\mu_R(\cdot)$, and teaching faculty, $\mu_T(\cdot)$, captures the fact that they can partially specialize these tasks internally by hiring dedicated teaching faculty or by increasing teaching loads for research faculty. The inclusion of faculty requires an additional labor market clearing condition for labor across the education, research, and production sectors.[21]

Explicitly including faculty in the university production technologies allows us to make full use of the microdata which separately reports university expenditures on faculty and equipment, both within teaching and research activities. It also allows the model to generate the sorting of faculty across the college quality-ladder, as summarized in figure A4.

[21] In particular, letting H_F denote the measure of effective labor in the production sector, the labor market clearing condition is akin to $H_F + \int h\,\mu_R(h|k)d\mu_c + \int h\,\mu_T(h|k)d\mu_c = \int h\,d\mu_h$

Moreover, since faculty are drawn from the adult population, the distribution of faculty human capital is endogenous to the model and depends on Σ_h^2.

Supply-Side Spillovers from University Research. Much of the literature on academic research emphasizes the productivity spillovers it generates for the production sector. A wealth of empirical studies have found these externalities to be quantitatively significant (Jaffe 1989; Mansfield 1995; Cohen, Nelson, and Walsh 2002; Ponds, Oort, and Frenken 2009; Schoellman and Smirnyagin 2021; Andrews 2022).

To account for these effects, we assume that the technology to produce the final goods, which are used for consumption and as inputs in the education and research, is subject to productivity spillovers from the knowledge created by academic research. Formally, firms operate a constant returns to scale production technology $F(H_F) = A \cdot H_F$, where H_F is aggregate effective labor in the production sector and A is total factor productivity (TFP). To incorporate spillovers from academic research, we assume aggregate TFP is a function of the stock of knowledge created by the higher education sector, so that $A = \bar{A}K^{\iota_k}$ where $K = \mathbb{E}[k]$. While the productivity spillovers from academic research are not crucial for our mechanism or conclusions, they help quantitatively account for an important general equilibrium channel whereby changes in university research output can effect household demand for education through wages.

Household Demographics and Intergenerational Dynamics. We introduce a more general process to determine a child's ability at the start of college, z. Following Capelle (2020), we model z as the result of an intergenerational process given by,

$$z = \xi h^{\varphi} \tag{25}$$

where $\ln \xi \sim \text{i.i.d.}\mathcal{N}\left(-\sigma_z^2/2, \sigma_z^2\right)$ is a random birth shock and φ captures the intergenerational transmission of skill from parent to child. The more general process allows the model to match microdata on the intergenerational correlation in human capital that is not accounted for by education investments. In the absence of these effects, the model may overstate the productivity of financial investments in education quality. The intergenerational transmission process also increases the persistence and aggregate impact of shocks to the higher education system.

Finally, we adopt a more flexible parameterization of time preference and intergenerational benevolence. We allow households to borrow and save at exogenous interest rate r. Each period corresponds to 4 years and individual lifecycles evolve deterministically. Each person lives for five periods as a child, then attends college for one period, and finally works as an adult for ten periods. Each household has one child midway through their adult life and sends them to college before retiring when their children enter the the labor market. See appendix D for details.

These changes allows us to differentiate between a household's time preference (δ), its intergenerational benevolence (β), and the university discount factor (β_c). This allows us to calibrate the time-scale of the quantitative model (through δ) separately from the benevolence factors β and β_c, which determine the weight that households and universities assign to future generations.

6 Model Calibration

This section explains the calibration strategy, reports the model's fit of important targeted moments, and validates the calibrated model on the empirical distribution of university R&D spending.

6.1 Data Sources

Our main sample includes all accredited public and private non-profit colleges in the United States that offer at least a four-year bachelor's degree. Unless otherwise stated, cross-sectional calibration targets are calculated using institution-level averages derived from 2012 - 2018 data. Our calibration draws on all the same administrative microdata sources that were used in section 2. Additional data on students' ability and parental earnings is taken from the National Longitudinal Survey of Youth 1997 (NLSY97). Data on government tuition subsidies comes from the National Postsecondary Student Aid Study (NPSAS). Aggregate statistics on income inequality and aggregate household spending on education are from the Congressional Budget Office (CBO) and the OECD's *Education at a Glance*. Appendix C provides additional detail on the sample and sources.

6.2 Externally Calibrated Parameters

We begin by calibrating a number of parameters that can be set externally. These parameters are listed in Table 2. The time discount factor (δ) is calibrated to a 0.96 annual rate, a standard value used in the literature. We calibrate the strength of spillovers from university research, ι_k, and the persistence of intangible capital, γ_K, to match estimates in the literature reviewed by Hall, Mairesse, and Mohnen (2010). Specifically, we set $\iota_K = .1$, consistent with the median estimate in the literature. Similarly, we calibrate the persistence of university knowledge to generate a 15% annual depreciation rate, consistent with the literature, adjusted to a four-year frequency so that $\gamma_k = (1 - .15)^4 \simeq .52$.

Table 2: Externally Calibrated Parameters

Parameter	Description	Value	Source
δ	Time discount factor	0.85	Standard (annually 0.96)
γ_k	Research Productivity Spillover	0.52	Hall et al. (2010)
ι_k	Output Productivity Spillover	0.10	Hall et al. (2010)
τ_y	Income Tax Progressivity	0.15	Heathcote et al. (2017)
τ_n	Tuition Subsidy Progressivity	0.18	NPSAS
$\bar{a}_y$	Average household income tax	0.20	CBO
$\bar{a}_n$	Average student education subsidy	0.53	OECD

To calibrate the progressivity of income taxes τ_y, we take estimates from Heathcote, Storesletten, and Violante (2017) derived from CPS data and the NBER's TAXSIM program. Consistent with estimates from the CBO, we set the average income tax rate to 20%. We estimate parameters of the government's tuition subsidy schedule using micro data from the National Postsecondary Student Aid Study (NPSAS) on student financial assistance, tuition, and parental incomes. Specifically, we estimate the student-aid progressivity parameter τ_n in (23) using the regression,

$$\log\left(\text{net tuition}\right) = \tau_n \cdot \log(\text{household income}) + \mathbf{X}'\boldsymbol{\beta} + \epsilon \tag{26}$$

where $\mathbf{X}$ includes the log of ACT scores and a constant.[22] We set the level of the subsidy schedule, $\bar{a}_n$, to match the average public subsidy to higher education. From the OECD's Education at a Glance (2020), total, private and public spending in higher education amounts to 2.6%, 1.7% and 0.9% of GDP respectively. We obtain $\bar{a}_n = 2.6/1.7 - 1 = 0.53$.

6.3 Internally Calibrated Parameters

The model's remaining parameters are jointly calibrated internally to match equilibrium characteristics of the household sector and of the market for higher education. Table 3 reports the calibrated values and Table 4 summarizes how well the model fits the data. Although no moment uniquely identifies individual parameters, we provide intuition for which moments are most informative for each parameter.

The first four parameters $(\beta, \sigma_z, \alpha, \varphi)$ govern the process of human capital accumulation, the degree of heterogeneity in ability, and extent of intergenerational altruism. These

[22]Because some students have zero net tuition in our sample, we use a pseudo-poisson-maximum-likelihood estimator (PPML) introduced by Silva and Tenreyro (2006).

Table 3: Internally Calibrated Parameters

Parameter	Description	Value
β	Inter-generational household preference	0.20
σ_z	Children ability shock	0.74
α	Elasticity of human capital w.r.t. college quality	0.15
φ	Elasticity of children ability w.r.t. parents' human capital	0.28
β_c	College time preference	0.10
ω_z	Elasticity of school quality w.r.t peer effects	0.51
ω_k	Elasticity of school quality w.r.t knowledge	0.25
ω_e	Elasticity of school quality w.r.t equipment	0.09
γ_e	Elasticity of knowledge w.r.t equipment	0.26
(a_G, τ_G)	External research grant award schedule	(0.02, 0.85)

Notes: *Additional details are contained in appendix C.*

parameters are most closely associated with the four household sector data targets. The intergenerational altruism parameter β is most closely associated with the share of household income spent on tuition, as can be seen in equation (14).

To discipline the parameters governing the intergenerational transmission of ability (σ_z, φ), we use the NLSY97 micro data and regress children's ASVAB score on thier parents earnings.[23] The slope coefficient of the regression is closely related to the elasticity of intergenerational transmission of human capital, φ, while the share of total variance explained by variation in parental income, the R^2, is inversely related to the standard error of the ability shock, σ_z. Finally the intergenerational elasticity (IGE) is informative about α, which connects the quality of education to future income of children.

The remaining set of parameters govern the technologies used by the higher education sector and the research grants award schedule. The elasticities of school quality and research output with respect to its inputs (e.g. ω_e, γ_e for equipment, ω_h, γ_h for faculty human capital) govern the share of revenues spent on each input. Two of these four parameters are therefore identified by the equipment expenditure share in teaching and in research. Imposing constant returns to scale on the university's education technology (e.g. $\omega_k + \omega_z + \omega_e + \omega_h = 1$), we can identify ω_h and ω_e separately for a given value of $\omega_k + \omega_z$. Similarly, with constant returns in the research technology (e.g. $\gamma_k + \gamma_e + \gamma_h = 1$), we can identify γ_h and γ_e separately given a value of γ_k.

[23] The Armed Services Vocational Aptitude Battery (ASVAB) consists of a battery of ten tests that measure knowledge and skill in several areas from maths to sentence comprehension.

Table 4: Jointly fit data targets for internal calibration

Description	Source	Data	Model
Households Sector			
Reg. test-scores on parent's earning (slope)	NLSY	0.12	0.12
Reg. test-scores on parent's earning (R^2)	NLSY	0.11	0.10
Share of household income spent on tuition	OECD	1.6%	1.6%
Inter-generational elasticity (IGE)	Davis and Mazumder (2017)	0.40	0.39
Higher Education Sector			
Research share in university expenditure	IPEDS & HERD	0.24	0.24
Std (log) university revenues	IPEDS	0.63	0.59
Innovation-Education Gap	Biasi and Ma (2021)	0.013	0.013
Grants share in total university revenue	IPEDS	0.17	0.17
Elasticity of tuition w.r.t. research expenditure	IPEDS	0.14	0.15
Elasticity of student ability w.r.t. research expenditure	IPEDS	0.34	0.34
Equipment expenditure share in teaching	IPEDS	0.40	0.40
Equipment expenditure share in research	IPEDS & HERD	0.54	0.54

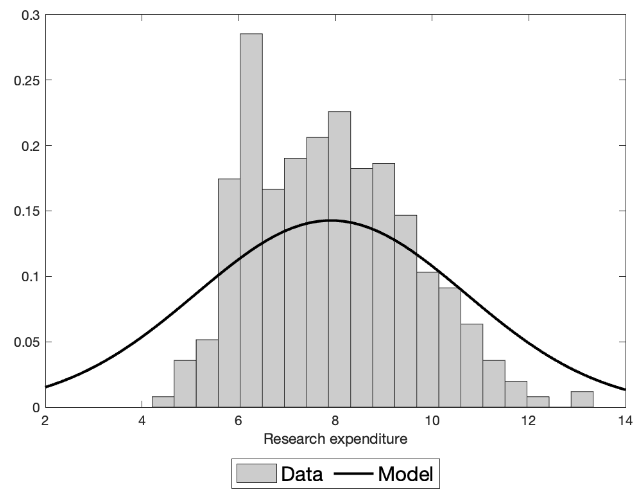
Notes: *Additional details are contained in appendix C.*

We use the research share of total university expenditure to identify β_c. As can be seen from equation (16), the research share increases in β_c since the more forward-looking colleges are, the more they invest in research. The dispersion of university revenues and the elasticity of mean ability with regard to research expenditure identify ω_z; the higher ω_z, the more colleges subsidize student ability and the lower the dispersion in revenues per student across colleges. An increase in ω_z also leads high-ranked colleges to to recruit more high ability students, increasing the cross-sectional correlation between average student ability and research expenditures.

To discipline ω_k, we use micro-estimates of the relation between frontier-focused education in universities and the earnings of its graduates. More specifically, we follow Biasi and Ma (2021) who estimate that a one unit decrease in the education-innovation gap is associated with a 0.011% increase in student income, after controlling for other college characteristics. Through the lens of our model, we interpret their measure of the innovation gap as $\log(k)$. Because the innovation gap is measureless, we normalize the point estimates by the standard deviation. In particular, the authors report that the standard deviation at the school level of the innovation gap is 0.85.[24] After normalization, we equate the marginal impact of knowledge on wages with the marginal impact of the innovation gap on wages, so that $\omega_k \alpha = .011/.85 \times \Sigma_k = .0129 \times \Sigma_k$.

[24]See Table 6, column 4 and section 6 in Biasi and Ma (2021).

Figure 5: A Validation Exercise: The Distribution of University Research Expenditures.



Notes: *Histogram plots the (log) in sample empirical distribution of university R&D expenditure per student. The solid line is the calibrated model's (non-targeted) prediction of the distribution of university R&D expenditure. Source: IPEDS.*

Since the only sources of university revenue in the model are tuition and government grants, the cross-sectional elasticity of tuition revenue with respect to research expenditure identifies the concentration of research grants across school, as parameterized by τ_G. For a value of τ_G, we can retrieve the intercept of the grant schedule, a_G, using the tuition share in total university revenue, which is the complement to the share of grants given the university budget constraint.

6.4 External Validation

As evident from Table 4, the model performs well in matching both the aggregate and distributional characteristics of the market for higher education. Given our aim of explaining the determinants of university R&D spending, an important additional question is the extent to which our calibrated model endogenously generates realistic variation of university research expenditures, which our calibration strategy does not directly target. Figure 5 compares the variation in university R&D generated by the calibrated model to what is observed in the data.[25] The figure validates that the model is able to mostly replicate the observed, but untargetted, cross-university variation in R&D expenditures, albeit with more mass in the tails than what is observed in the data. One potential explanation

[25]To ease the comparison, we overlap the two distributions by choosing units in the model to match the mean *level* of R&D expenditure in the data

for the discrepancy in the left tail could be the fact that the NSB surveys do not cover universities with less than $150,000 in research expenditures in the previous year.

7 Quantitative Implications of Federal Research and Tuition Policies

To evaluate the quantitative significance of the interdependencies between the market for higher education and research spending, we conduct several counterfactual exercises to analyze the impact of federal tuition and research policies on educational and research outcomes. In particular, we show how university research expenditures, educational outcomes, earnings inequality, and output would change if we removed government tuition ($\{a_n, \tau_n\} = \{0, 0\}$) or research policies ($\{a_g, \tau_g\} = \{0, 0\}$). Given that the universities incentives to spend on research depend on the dispersion of tuition and student, we also show how these outcomes would change if we removed the distributional aspects of these policies (i.e. $\tau_n = 0$, $\tau_g = 0$) while keeping their level unchanged.

7.1 The Impact of Progressive Federal Tuition Policies

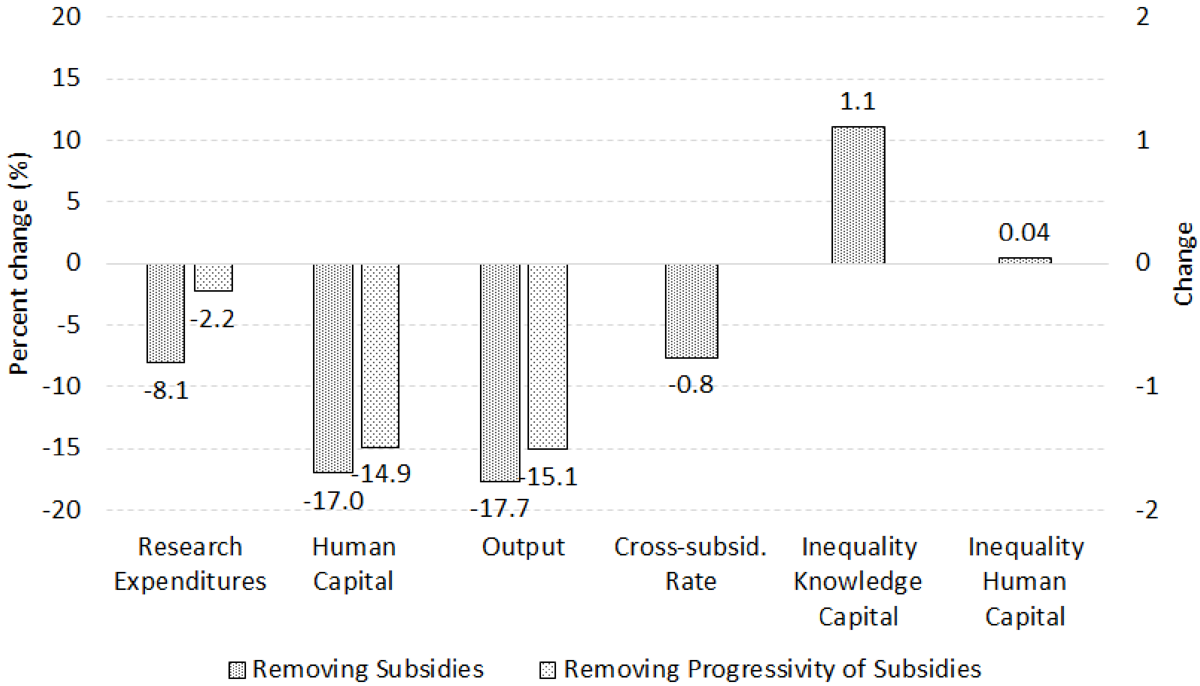
The first exercise computes the long-run economic impact of removing federal tuition policies, ($\{a_n, \tau_n\} = \{0, 0\}$). In historical perspective, the counterfactual corresponds to the period before the federal government substantialy increased its subsidies for student tuition through programs like the G.I. Bill of 1944, the National Defense Education Act of 1958, the Higher Education Act of 1965, and subsequent legislation.

Figure 6 displays the results of the model simulation. Removing current federal tuition policies causes university research expenditure to fall by 8.1%, household human capital to decline by 17.0%, and output to fall by 17.7%. From the last two columns, we also see that household income inequality rises (Σ_h) and inequality of research capital across colleges increases (Σ_k).

The lightly shaded bars report the impact of replacing progressive federal tuition policies with a flat tuition subsidy (i.e. $\tau_n = 0$) chosen so that overall government spending on subsidies remains unchanged. Comparing the bars, we see that the impact on human capital and output arises nearly entirely from the progressivity of government tuition policies.[26] Moving to a flat subsidy would also reduce university research spending by 2.2%, accounting for just over one-quarter of the total effect. Most of the rise in university research induced by government tuition policies arises from the overall increase in

[26]Note that we do not report additional columns for the effect on the cross-subsidization rate s_R, the dispersion in college intangible capital Σ_k, and the dispersion in household types Σ_h, since all three depend solely on the distributional aspects of policy, not on the level of support, and so remain unchanged.

Figure 6: Long-Run Impact of Removing Current Tuition Subsidies



Notes: *Outcomes report changes relative to the calibrated benchmark steady state. Research expenditure, human capital, and output are plotted on the left y-axis. Changes in the cross-subsidization rate and in inequality (standard deviations) in the final three columns are reported on the right y-axis.*

university revenue; a point also evident in the small -0.8 percentage point change the university research expenditure share.

The propositions in section 4 provide the economic intuition behind the results in Figure 6. By design, progressive tuition subsidies compress the dispersion in household education expenditure by redistributing spending from wealthy households to poor households. Since lower quality schools are disproportionately attended by children from poor families, and children from rich families disproportionately enroll in high quality colleges, the compression in household education expenditure also reduces the dispersion in college revenues, Σ_R and therefore in college quality, Σ_q . In the long run, the equalization of college revenues also leads to more similar research expenditures causing Σ_k to decline. The dispersion in college quality Σ_q falls by less than Σ_k because education quality also depends on non-monetary inputs, like student ability z, which do not readily change with market conditions. Consequently, the college ladder Σ_q/Σ_k steepens with progressive tuition policies, leading to an increase in the intensity of university research spending, s_R.

The above highlights how tuition policies can impact university research by altering the degree of stratification in higher education. Alongside these effects, progressive tuition subsidies also increase research by boosting the *level* of college revenues. The subsidy does so not only directly, by decreasing the cost of education, but also indirectly through a long-run general equilibrium increase in productivity and income that raises the demand for education.

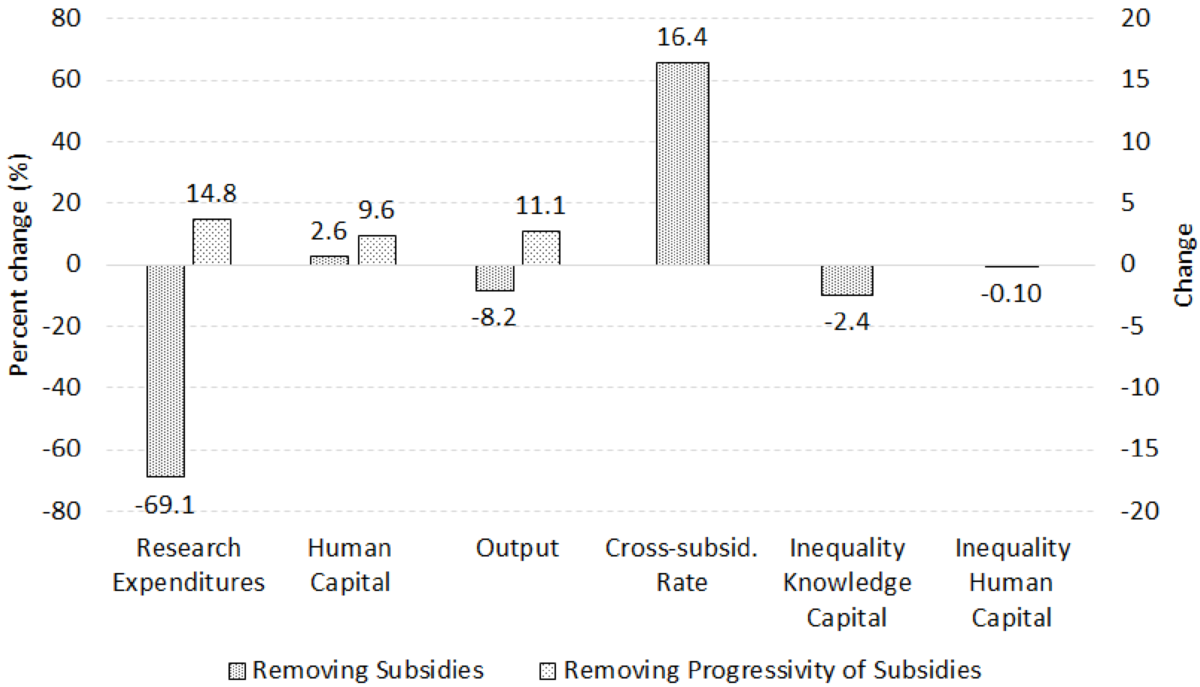
The results in Figure 6 show that in addition to reducing inequality and boosting human capital accumulation–as policy makers likely intend–progressive federal tuition subsidies can also substantially increase university R&D expenditure. The rise is supported both by an increase in the level of university revenues and by a change in the dispersion of research capital across colleges which increases competition between universities to spend more on research. In other words, the model suggests a novel complementarity between equity and innovation in higher education; policies which promote more equitable educational outcomes also incentivize more spending on innovative research.

7.2 The Impact of Meritocratic Federal Research Grants

Figure 7 displays the long-run impact of removing federal research grants $\{a_g, \tau_g\} = \{0, 0\}$. In historical perspective, the counterfactual corresponds roughly to the period just before World War II, when the federal government rarely funded university research. During the war, federal funding increased dramatically as faculty researchers were mobilized for wartime R&D projects, such as the Manhattan Project. Federal support for university research continued growing after the war, disbursed through an expanding range of competitive award programs at federal agencies like the National Institute of Health, the National Science Foundation, the Department of Defense, NASA, the Department of Energy, the Environmental Protection Agency, and others.

The results show that without federal grants, university research declines by 69.1%, leading to an 8.2% long-term decline in output as a consequence of lost spillovers from academic research. Interestingly, the decline in university research is 6.9 percentage points *less* than the government's share of total funding for university research, which stands at 76% in the data and calibrated model (see Table 4). The gap suggests a crowding out of university research spending by government grants. In other words, universities increase internal spending on research to partially offset declines in government funds. This behavior is noteworthy in that it is at odds with the traditional view of university R&D, while at the same time being consistent with data trends in recent years that suggest universities undertake just such offsetting behavior (see Figure 1).

Figure 7: Long-Run Impact of Removing Research Grants



Notes: *Outcomes report changes relative to the calibrated benchmark steady state. Research expenditure, human capital, and output are plotted on the left y-axis. Changes in the cross-subsidization rate and in inequality (standard deviations) in the final three columns are reported on the right y-axis.*

The reason that government grants partially crowd out private spending is that their current award structure concentrates grants at top universities which already have the most resources. This increases the dispersion in university research expenditures which, in the long-run, leads to a rise in Σ_k as the gaps between institutions widen. As a result, the dispersion in college quality Σ_q increases, but by less than Σ_k because education quality also depends on the distribution of student ability z. Consequently, the college quality ladder Σ_q/Σ_k flattens, and universities have less incentive to spend on research. Quantitatively, these effects appear to be economically significant. Figure 7 shows that the current government grant regime decreases the university research share by 16.4 percentage points. A change of this size is substantial and alone could account for a 20% decline in university R&D (see Figure A9). This induced decline in the research share is the main force behind the observed crowding-out effect of government research grants.

Government research grants also affect educational outcomes in higher education by altering the distribution of resources across universities. The resulting rise in the dispersion of college quality, Σ_q, leads to a modest increase in long-term income inequality Σ_h. The rise in inequality weakens the model's peer effect, leading to a 2.6% fall in human capital as colleges struggle to educate students from increasing disparate backgrounds.

The lightly shaded bars in Figure 7 display the long-run impact of adopting a flat research subsidy ($\tau_g = 0$) whose level is chosen so that government expenditure on research grants remains unchanged. The model predicts that moving to a flat research subsidy would increase university research expenditure by 14.8%, increase human capital by 9.6%, and boost output by 11.1%. The flat subsidies perform better because they continue to reduce the relative cost of research expenditure without distorting the market structure of the higher education sector. As a result, government grants no longer discourage internal university spending. Universities raise their internal spending on research by 16.4 percentage points, which further boosts human capital accumulation and aggregate output. By reducing the concentration of resources at top schools, the flat subsidy also results in a modest reduction of income inequality.

Taken together, the results in Figure 7 show that federal R&D grants boost university research, but also reduce human capital accumulation, exacerbate educational inequality, and discourage internal university spending on research. This is because the prevailing grant awarding system further concentrates resources at top programs, which already have the most resources. Holding constant government expenditures, moving to a flat research subsidy would boost research expenditure by removing distortions to the higher education sector which feed back into university research decisions. The flat research subsidy also increases output, human capital accumulation, and reduces inequality. Hence, the model suggests that flat research subsidies could be more effective while also eliminating the trade-off between equity and innovation present in the research grant system.

7.3 Transition Dynamics following Policy Reforms

Having analyzed the long-term effects, we now compute the transition dynamics following policy reforms in the higher education sector. We do so first in the context of simulating the implementation of President Biden's recently proposed "free tuition" policies which aim to substantially increase the generosity and progressivity of federal student financial-aid.[27] For research policies, we simulate the transition dynamics of adopting flat research subsidies. The model reveals that the adjustment processes are not monotone, and that the short-run dynamics of the economy can differ from the long-run impact in both size and sign.

[27] Our calibration is based on this version of the proposal available on the campaign website.

7.3.1 Calibrating Federal Tuition Policy Reforms

Under the free tuition program, households with annual income less than $125,000 would pay no tuition at public universities and the maximum Pell-grant would be doubled. To implement this policy, we simulate the counterfactual distribution of out-of-pocket tuition implied by the policy change using NPSAS micro data where we observe parental income, tuition paid and the breakdown of federal, state, institutional and private aid received by a representative sample of U.S. students in 2008. We then re-estimate the parameters of the student aid schedule $(\tau_n, \bar{a}_n)$ on the simulated data.

In the simulation, we assume that households with income less than $125,000 receive a federal grant equal to the full amount of the tuition if their child goes to a public college. To predict the Pell-grants received by each student after the increase of the maximum Pell-grant, we use the formula borrowed from Epple et al. (2017),

$$\text{Pell-grants} = \min\left(\max\left[0, p - EFC\right], \overline{\text{Pell-grant}}\right)$$

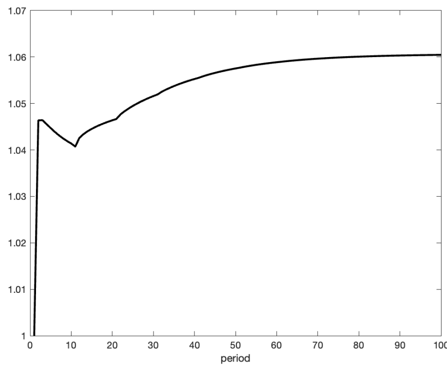
where we use for p the cost of attendance that includes the federal grant for families with income less than $125,000 and where EFC denotes the expected family contribution that the government computes and which is an increasing function of income. In the counterfactual, we multiply the maximum Pell-grant, $\overline{\text{Pell-grant}}$, by two. We assume that there is no change in other forms of aid.

We choose the degree of progressivity, τ_n', that best matches the simulated out-of-pocket tuition payments by re-running (26) on the simulated sample. The PPML estimate is $\tau_n' = 0.39$, which is substantially higher than the current level (.18). Under our assumption of no change in the other sources of aid, the free tuition policy would significantly increase the degree of progressivity of the need-based aid schedule. From the average next tuition payment after the policy $\mathbb{E}_{\text{After}}\left[\text{net tuition}\right]$ we can obtain the average rate of college subsidies $\bar{a}_n'$ using the formula $\bar{a}_n' = (1+\bar{a}_n)\frac{\mathbb{E}_{\text{Before}}[\text{net tuition}]}{\mathbb{E}_{\text{After}}[\text{net tuition}]} - 1 = 1.53 \times 1.21 - 1 = .85$.

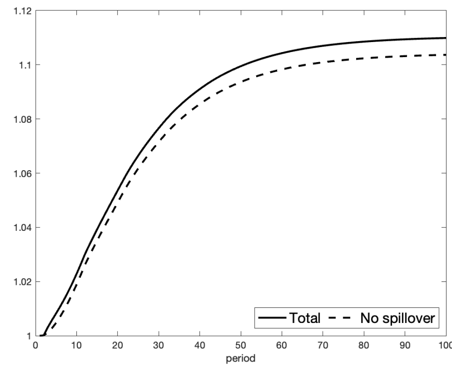
7.3.2 The Transition Path following an Expansion of Federal Tuition Subsidies

Figure 8 plots the transition path of several key economic variables following an unexpected one-time and permanent increase in the level and progressivity of federal tuition subsidies. The first point on each panel represents the pre-intervention calibrated steady state. The model's prediction of the policy's long-term effect are consistent with what policy makers presumably intended when enacting them: worker human capital increases,

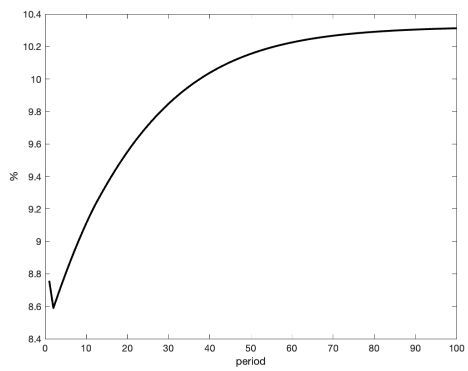
Figure 8: Transition Path following Student-Aid Policy Intervention



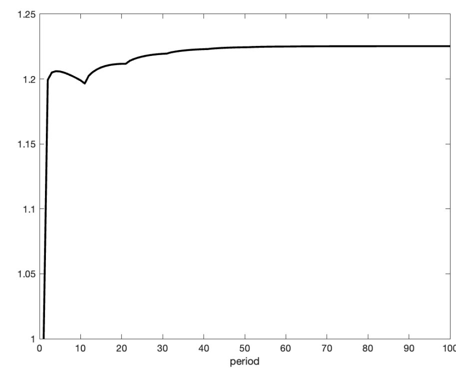
(a) Research Expenditure



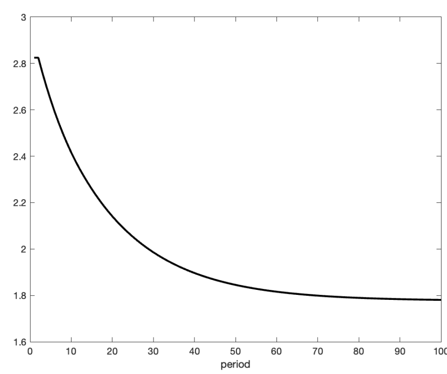
(b) Aggregate Output (w/ spillover)



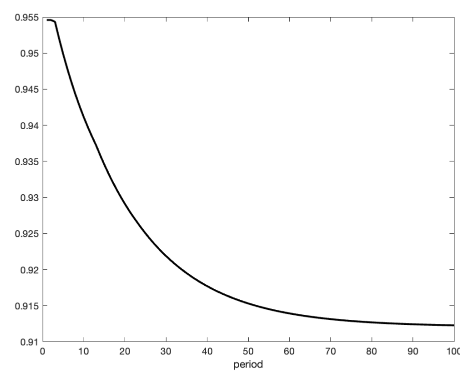
(c) Cross-Subsidization Rate (s_R)



(d) University Revenue



(e) Std. Knowledge Capital (Σ_K)



(f) Household Income Inequality (Σ_h)

Note: *Initial points corresponds to the pre-intervention calibrated steady state. Standard deviations for household income and knowledge capital correspond to logs of the underlying variables, as in the text. Student ability sorting reports the correlation between student ability and college quality. Level variables, such as output and research expenditure, report cumulative changes with respect to the initial steady state.*

university intangible capital becomes more equal (panel (e)), household income inequality falls (panel (f)), and output rises (panel (b)). In addition, the model predicts that university research expenditure increases by 5% (panel (a)), consistent with the mechanisms discussed above.

While the model's long-run predictions are consistent with the results of the previous section, examining the economy's transition path reveals that the adjustment process is not monotone and so the short-run dynamics of the economy can look quite different from long-term outcomes. Most notable is the behavior of the cross-subsidization rate (panel (c)), which falls on impact and then gradually rises above its initial level. This occurs because opposing economic forces which link the progressivity of tuition subsidies to university R&D push in different directions, and evolve differentially over time.

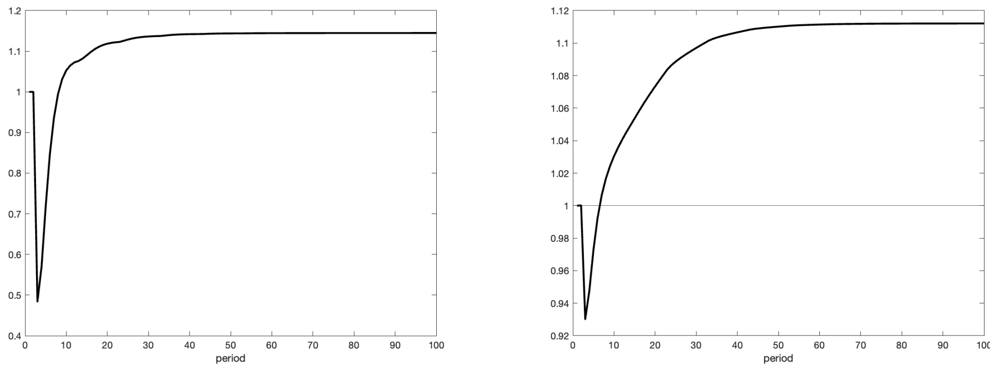
As discussed in section 7.1, progressive tuition subsidies ultimately reduce both household income inequality Σ_h and the dispersion of college intangible capital Σ_k. However in the short-run, the effects of Σ_h dominate, while in the long-run Σ_k dominates. This is because progressive tuition subsidies reduce the dispersion in household education expenditures immediately, while the convergence in university capital occurs gradually over time. The slow manifestation can be traced to the high persistence of university intangible capital.

Another insight from the transition path is that while the rise in university R&D is relatively stable, the economic forces driving the increase change over the course of the adjustment. Initially R&D spending rises despite a fall in the cross-subsidization because of a large jump in university revenue. The jump is driven by lower-income households who suddenly find themselves able to send their children to higher quality colleges with the help of more generous and progressive federal subsidies. However, over time household expenditure on tuition gradually decreases as the differences between institutions shrink. As colleges become more similar, competition increases and the quality-ladder steepens, leading to a sustained rise in the research share that supports university R&D even as revenues slow and stagnate.

7.3.3 The Dynamics of Research Policy Reforms

The transition dynamics following changes in government research policies exhibit similar non-monotonicity whereby the short-run policy impacts looks different from the long-run effects. The reason stems again from the fact that changes in the distribution of intangible capital across colleges, Σ_k, unfolds gradually over time. To illustrate the point,

Figure 9: Transition Path after moving to Flat Research Subsidy



(a) Knowledge Capital (b) Aggregate Output

Figure 9 plots the transition path for aggregate output and the aggregate stock of knowledge created by university research following a shift from the existing grant system to a counterfactual flat-subsidy using the same budget. In the short-run, moving to a flat research subsidy substantially *decreases* university research output by redistributing resources from leading research universities with high knowledge capital to lower ranked institutions with less knowledge capital. While inefficient in the short-run, this reduction in the concentration of resources allows lower-ranked institutions to compete more effectively with top programs by spending more on research. As a consequence, differences between universities narrow over time and the long-run equilibrium features more equal educational institutions and higher university spending on research.

An important implication of the transition dynamics is that high-frequency studies of policy reforms in education markets should be interpreted with caution since certain general equilibrium forces may only materialize in the long-run. In the model, these long-run general equilibrium effects depend largely on the distribution of college knowledge capital, which adjusts to market conditions only slowly over time.

8 Conclusion

This paper develops a model in which university research depends endogenously on the market for higher education. Universities invest in research to improve their education quality and better compete for tuition and talented students. The greater the positive assortative matching of students and schools, the stronger is the incentive for universities to spend on research. The model can match important new features of the microdata on university research and its core mechanism is consistent with causal evidence from a natural experiment exploiting large increases in the NIH grants. The model also rationalizes why universities cross-subsidize research with tuition revenue and why they continue to spend on R&D despite low returns to patenting. The mechanism also has important implications for public research and tuition policies. The computational exercises predict that current federal need-based student financial aid programs not only reduce inequality, they also lead universities to spend more on research. Federal research grants also boost research, but partially crowds out private spending and contributes to educational inequality by concentrating resources at the top.

In historical perspective, the model's predictions also demonstrate how its mechanism can help us understand important post-war trends in university research and the market for higher education. Our counterfactuals correspond roughly to the early post-war period, before the federal government began to intervene heavily through university R&D and tuition subsidies. The literature documents many other important changes in the market for higher education which would interact with our mechanism and so warrant further study. For instance, the mid-20th century national integration of the U.S. college market increase competition between universities and dramatically changed the sorting patterns of students across schools (Hoxby 2009; Hendricks, Herrington, and Schoellman 2021). The subsequent globalization of higher education internationalized university competition as cross-border flows of faculty, specialists, and students rose dramatically (Kerr et al. 2016; Bound et al. 2021). Through the lens of our model, such changes could have profound effects on university research and educational inequality which future work should aim to understand.

Finally, while we believe the model provides new insights relative to existing theories, much important work remains to be done in understanding the value and determinants of university R&D. Universities are exceptionally complex institutions and future work should further examine the substantial heterogeneity across universities in research intensity, academic specialization, funding sources, and education quality. Our results also point to the importance of jointly considering government research and tuition subsidies in the optimal design of policies, as the two interact in equilibrium and can have offsetting effects. Finally, more work on cross-country differences in university R&D and how they relate to the structure of education markets is needed. Such work would provide new and valuable insights into the determinants of aggregate innovation and their relation to the structure of education around the world.

References

Akcigit, Ufuk, Douglas Hanley, and Nicolas Serrano-Velarde. 2020. "Back to Basics: Basic Research Spillovers, Innovation Policy, and Growth." *The Review of Economic Studies* 88 (1): 1–43.

Akcigit, Ufuk, Douglas Hanley, and Stefanie Stantcheva. 2022. "Optimal Taxation and RD Policies." *Econometrica* 90:645–684.

Akcigit, Ufuk, Jeremy G Pearce, and Marta Prato. 2020, September. "Tapping into Talent: Coupling Education and Innovation Policies for Economic Growth." Working paper 27862, National Bureau of Economic Research.

Andrews, Michael. 2022. "How Do Institutions of Higher Education Affect Local Invention? Evidence from the Establishment of U.S. Colleges." *American Economic Journal: Economic Policy*.

Arrow, Kenneth. 1962. "Economic welfare and the allocation of resources for invention." *The Economics of Science and Innovation* 1 (01): 61–77.

Azoulay, Pierre, Joshua Graff Zivin, Danielle Li, and Bhaven N. Sampat. 2019. "Public R&D Investments and Private-sector Patenting: Evidence from NIH Funding Rules." *Review of Economic Studies* 86 (1): 117–152.

Azoulay, Pierre, Joshua S. Graff Zivin, and Gustavo Manso. 2011. "Incentives and creativity: evidence from the academic life sciences." *The RAND Journal of Economics* 42 (3): 527–554.

Benabou, Roland. 2002. "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?" *Econometrica* 70 (2): 481–517 (March).

Biasi, Barbara, David J. Deming, and Petra Moser. 2020, January. In *Education and Innovation*. University of Chicago Press.

Biasi, Barbara, and Song Ma. 2021. "The Education-Innovation Gap." Working paper.

Bound, John, Breno Braga, Gaurav Khanna, and Sarah Turner. 2021. "The Globalization of Postsecondary Education: The Role of International Students in the US Higher Education System." *Journal of Economic Perspectives* 35 (1): 163–184 (Winter).

Cai, Zhifeng, and Jonathan Heathcote. 2022. "College Tuition and Income Inequality." *American Economic Review* 112 (1): 81–121 (January).

Capelle, Damien. 2020. "The Great Gatsby Goes to College: Tuition, Inequality, and Intergenerational Mobility in the U.S." *Working Paper.*

Cohen, Wesley M., Richard R. Nelson, and John P. Walsh. 2002. "Links and Impacts: The Influence of Public Research on Industrial R&D." *Management Science* 48 (1): 1–23.

Council on Government Relations. 2019, April. "Excellence in Research: The Funding Model, FA Reimbursement, and Why the System Works." Technical Report.

Davis, Jonathan, and Bhashkar Mazumder. 2017, March. "The Decline in Intergenerational Mobility After 1980." Working paper series WP-2017-5, Federal Reserve Bank of Chicago.

Epple, Dennis, and Richard Romano. 2010. "Peer Effects in Education: A Survey of the Theory and Evidence." *Handbook of Social Economics* 1B (1): 885–928.

Epple, Dennis, Richard Romano, Sinan Sarpça, and Holger Sieg. 2017. "A general equilibrium analysis of state and private colleges and access to higher education in the U.S." *Journal of Public Economics* 155 (C): 164–178.

Epple, Dennis, Richard Romano, and Holger Sieg. 2006. "Admission, Tuition, and Financial Aid Policies in the Market for Higher Education." *Econometrica* 74 (4): 885–928.

Epple, Dennis, and Richard E. Romano. 1998. "Competition between Private and Public Schools, Vouchers, and Peer-Group Effects." *The American Economic Review* 88 (1): 33–62.

Gersbach, Hans, and Maik T. Schneider. 2015. "On the global supply of basic research." *Journal of Monetary Economics* 75:123–137.

Gersbach, Hans, Maik T Schneider, and Olivier Schneller. 2013. "Basic Research, Openness, and Convergence." *Journal of Economic Growth* 18:33–68.

Hall, Bronwyn, and Fumio Hayashi. 1989, May. "Research and Development As An Investment." Nber working papers 2973, National Bureau of Economic Research, Inc.

Hall, Bronwyn, Jacques Mairesse, and Pierre Mohnen. 2010. "Measuring the Returns to RD." Chapter 24 of , Volume 2, 1033–1082. Elsevier.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante. 2017. "Optimal Tax Progressivity: An Analytical Framework*." *The Quarterly Journal of Economics* 132 (4): 1693–1754 (06).

Hendricks, Lutz, Christopher Herrington, and Todd Schoellman. 2021. "College Quality and Attendance Patterns: A Long-Run View." *American Economic Journal: Macroeconomics* 13 (1): 184–215 (January).

Hendricks, Lutz, and Todd Schoellman. 2018. "Human Capital and Development Accounting: New Evidence from Wage Gains at Migration*." *The Quarterly Journal of Economics* 133 (2): 665–700 (12).

Hoxby, Caroline M. 2009. "The Changing Selectivity of American Colleges." *Journal of Economic Perspectives* 23 (4): 95–118 (December).

Hvide, Hans K., and Benjamin F. Jones. 2018. "University Innovation and the Professor's Privilege." *American Economic Review* 108 (7): 1860–98 (July).

Jaffe, Adam B. 1989. "Real Effects of Academic Research." *The American Economic Review* 79 (5): 957–970.

Kerr, Sari Pekkala, William Kerr, Çağlar Özden, and Christopher Parsons. 2016. "Global Talent Flows." *Journal of Economic Perspectives* 30 (4): 83–106 (November).

Klette, Tor Jakob. 1996. "RD, Scope Economies, and Plant Performance." *The RAND Journal of Economics* 27 (3): 502–522.

Lach, Saul, and Mark Schankerman. 2008. "Incentives and invention in universities." *The RAND Journal of Economics* 39 (2): 403–433.

Lochner, Lance, and Alexander Monge-Naranjo. 2012. "Credit constraints in education." *Annu. Rev. Econ.* 4 (1): 225–256.

Mansfield, Edwin. 1995. "Academic Research Underlying Industrial Innovations." *The Review of Economics and Statistics* 77 (1): 55–65.

Martellini, Paolo, Todd Schoellman, and Jason A. Sockin. 2022, March. "The Global Distribution of College Graduate Quality." Working papers 791, Federal Reserve Bank of Minneapolis.

Merton, Robert K. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press.

Mowery, David C., Richard R. Nelson, Bhaven N. Sampat, and Arvids A. Ziedonis. 2015. *Ivory Tower and Industrial Innovation: University-Industry Technology Transfer Before and After the Bayh-Dole Act*. Stanford University Press.

National Science Board. 2018. "The State of U.S. Science and Engineering." Technical Report NSB-2018-1.

Nelson, Richard R. 1959. "The Simple Economics of Basic Scientific Research." *Journal of Political Economy* 67 (3): 297–306.

Ponds, Roderik, Frank van Oort, and Koen Frenken. 2009. "Innovation, spillovers and university–industry collaboration: an extended knowledge production function approach." *Journal of Economic Geography* 10 (2): 231–255.

Rosenberg, Nathan. 1990. "Why Do Firms Do Basic Research (with Their Own Money)?" *Research Policy* 19 (02): 165–174.

Rosenberg, Nathan, and Richard R. Nelson. 1994. "American universities and technical advance in industry." *Research Policy* 23 (3): 323–348.

Rothaermel, Frank T., Shanti D. Agung, and Lin Jiang. 2007. "University entrepreneurship: a taxonomy of the literature." *Industrial and Corporate Change* 16 (4): 691–791 (07).

Sacerdote, Bruce. 2014. "Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward?" *Annual Review of Economics* 6 (1): 253–272.

Schoellman, Todd, and Vladimir Smirnyagin. 2021. "The Growing Importance of Universities for Patenting and Innovation." *Working Paper.*, 08.

Silva, J. M. C. Santos, and Silvana Tenreyro. 2006. "The Log of Gravity." *The Review of Economics and Statistics* 88 (4): 641–658 (November).

Smith, Pamela W. 2006, October. "The National Institutes of Health (NIH): Organization, Funding, and Congressional Issues." Nber working papers ADA458502, Library of Congress.

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

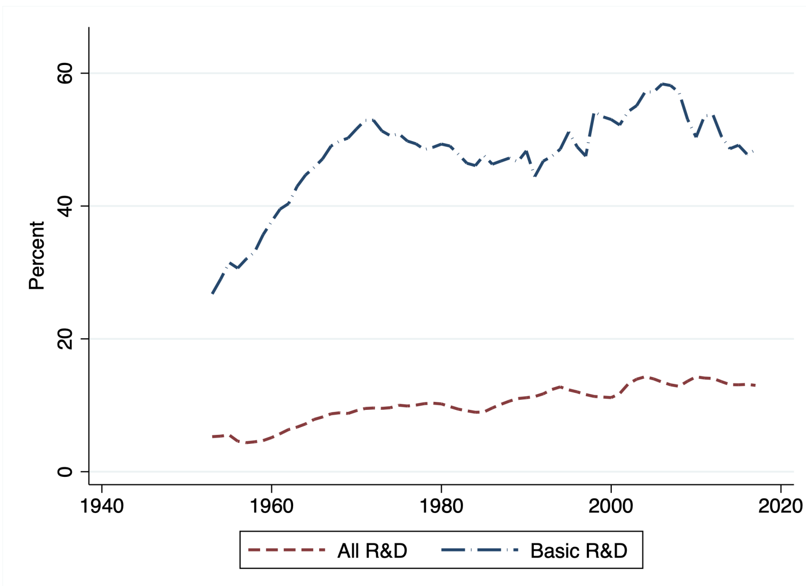
Urquiola, Miguel. 2020. *Markets, Minds, and Money*. London: Harvard University Press.

Willetts, David. 2018. *A University Education*. London: Oxford University Press.

Appendix

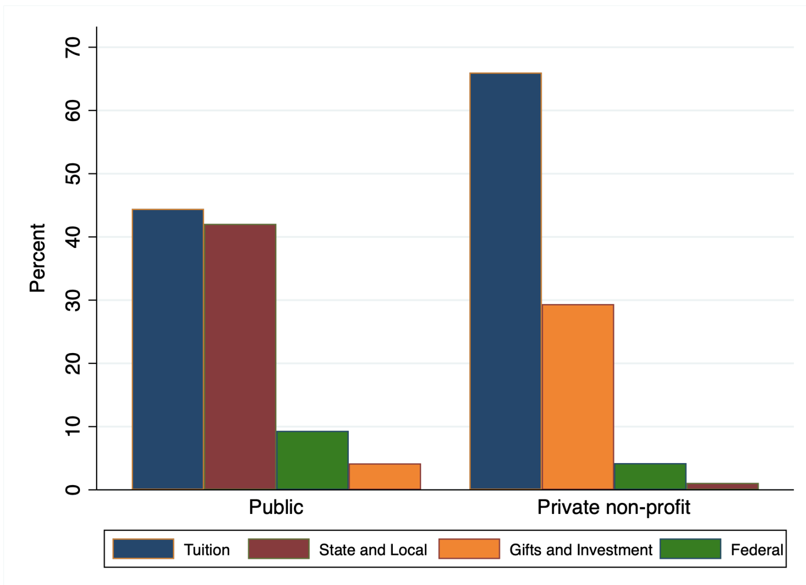
A Additional Figures

Figure A1: Higher education research expenditure as share of national total



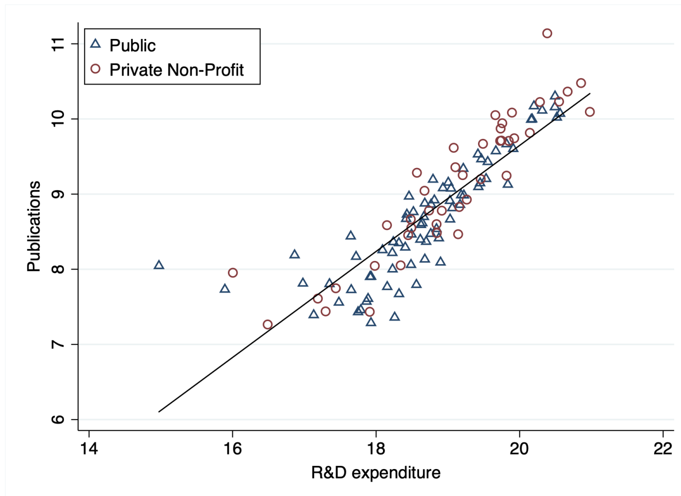
Notes: *Y-axis represents the higher education sector's share of total domestic research and development expenditures, by type. Underlying data come from National Science Board (2018).*

Figure A2: University revenue sources, by sector

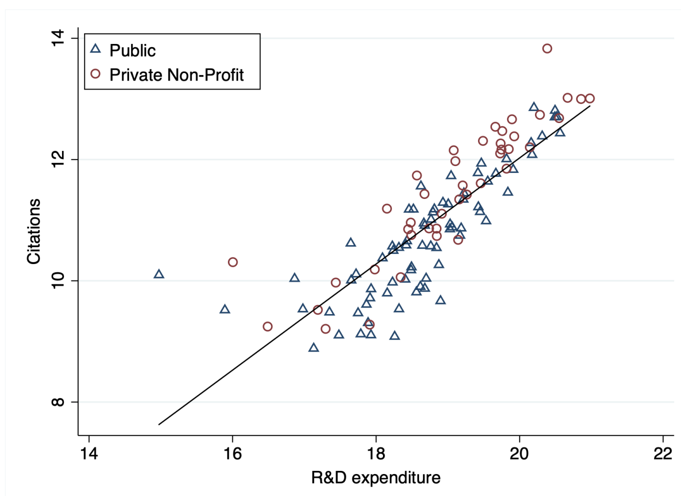


Notes: *Underlying data from the NCES's Integrated Postsecondary Education Data System (IPEDS). Y-axis reports share of total university revenue. University revenue includes tuition; federal, state, and local appropriations, grants, and contracts; affiliated entities, private gifts, grants, and contracts; investment return; and endowment earnings. Revenue from auxiliary, hospitals, and other independent operations are excluded.*

Figure A3: University Research Spending and Knowledge Creation



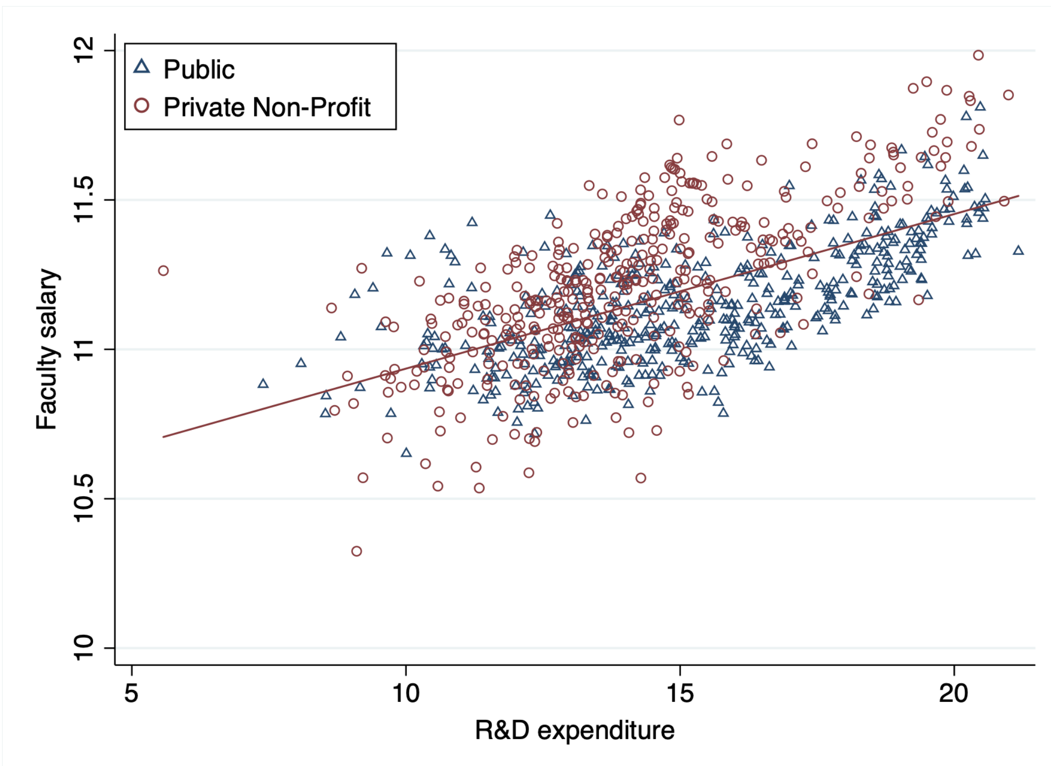
(a) Publications



(b) Citations

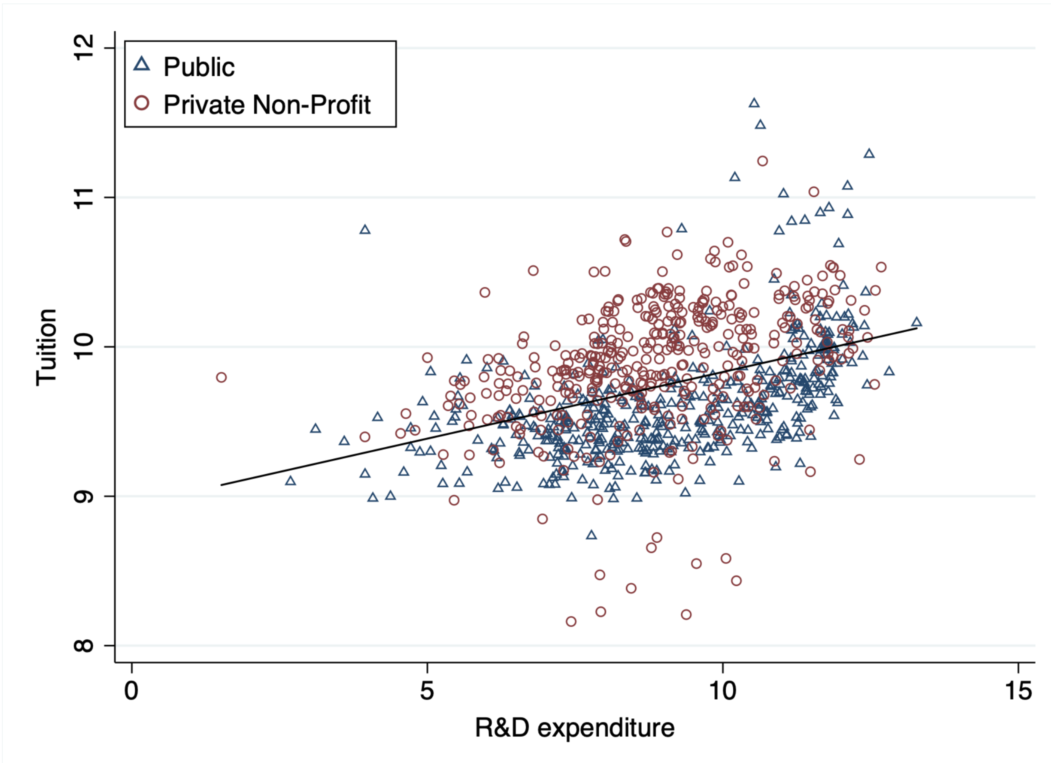
Note: *Publication and citation data come from the CWTS Leiden Ranking derived from the core collection of the Web of Science (WoS) for the years 2015-2018. research expenditure is total university spending for activities specifically organized to produce research outcomes, including by institutes, research centers, and individuals. Research expenditure data come from IPEDS with points representing university averages for 2012-2015.*

Figure A4: University Research Spending and Faculty Compensation



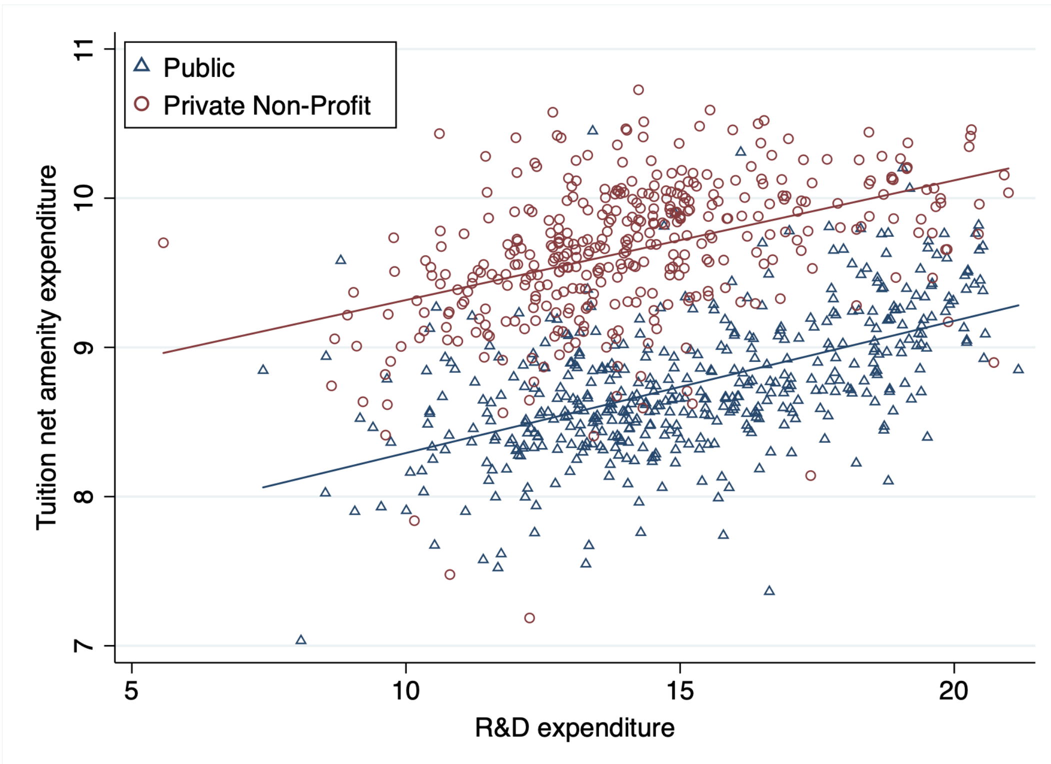
Note: *Faculty salary is the average salary for full-time faculty members on 9-month equated contracts. Research expenditure is total university spending for activities specifically organized to produce research outcomes, including by institutes, research centers, and individuals. Data comes from the Integrated Postsecondary Education Data System with points representing university level averages for 2012-2015.*

Figure A5: University research spending and tuition with state and local appropriations



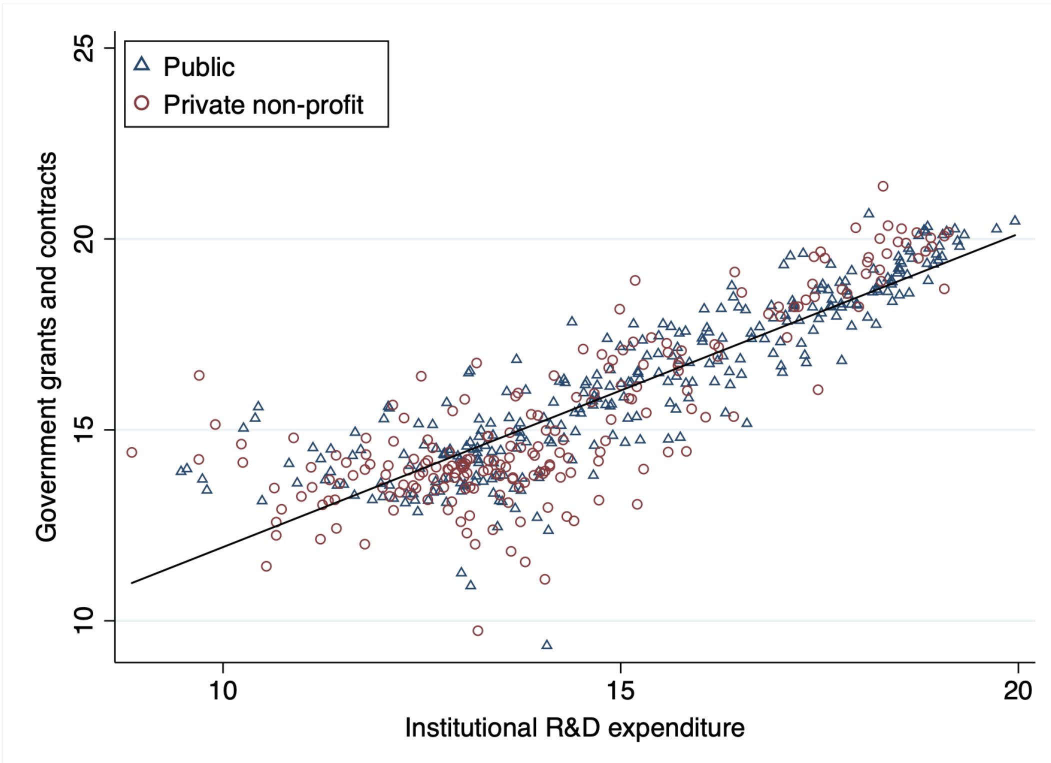
Notes: *Tuition is the average tuition revenue the university receives per full-time equivalent student, net of any university discounts or allowances, but including state and local appropriations per capita. research expenditure is total university spending for activities specifically organized to produce research outcomes, including by institutes, research centers, and individuals. Data retrieved the Integrated Postsecondary education Data System (IPEDS). Points represent (log) university level averages from 2012-2015.*

Figure A6: University R&D spending and tuition net of student services expenditures, by sector



Notes: *Tuition is the average tuition revenue the university receives per full-time equivalent student, net of any university discounts or allowances, and net of expenditures on student services per capita. Student services includes spending on activities whose primary purpose is to contribute to students emotional and physical well-being and to their intellectual, cultural, and social development outside the context of the formal instructional program. Registrar and admissions expenses are also included. research expenditure is total university spending for activities specifically organized to produce research outcomes, including by institutes, research centers, and individuals. Data retrieved the Integrated Postsecondary education Data System (IPEDS). Points represent (log) university level averages from 2012-2015.*

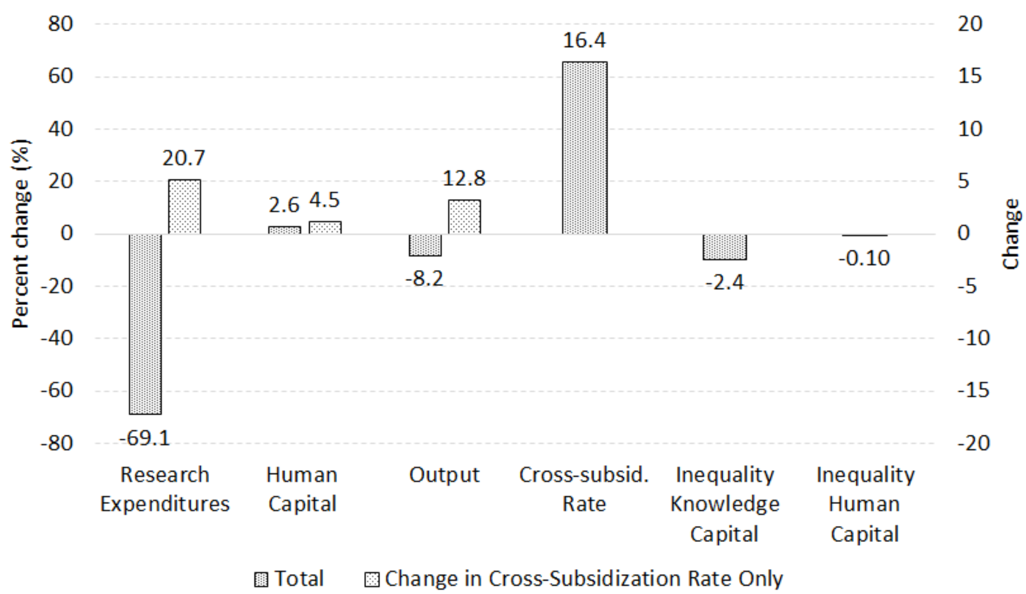
Figure A7: University Research Spending and Government Grants



Notes: *Institutional research expenditures correspond to internal university funds that are separately budgeted for individual research projects. Government grants and contracts include funds received from the federal, state, or local government for research, training, or other public service. Points correspond to 2012-2015 university averages in log-scale. Data on institutional research is from the Higher Education Research and Development Survey (HERD). Data on grants and contracts is from the Integrated Postsecondary Education Data System (IPEDS).*

Figure A8: Long-Run Impact of Removing the current US Subsidies

Figure A9: Long-Run Impact of Removing Research Grants



B Additional Tables

Table B1: Characteristics of top 25 research universities, by total research spending

Institution	Total research	Type of research			Source of research funding					
	(millions USD)	Fundamental	Applied	Development	Federal Gov	State Gov	Intsitutional	Business	Nonprofit	Other
Johns Hopkins University	2206	64%	27%	9%	87%	0%	4%	2%	6%	0%
University of Michigan-Ann Arbor	1354	59%	40%	1%	57%	0%	34%	4%	4%	1%
University of Washington-Seattle Campus	1165	65%	23%	12%	78%	2%	6%	3%	8%	2%
University of Wisconsin-Madison	1118	93%	6%	1%	50%	7%	33%	2%	7%	2%
University of California-San Diego	1080	80%	6%	13%	58%	4%	14%	7%	9%	8%
University of California-San Francisco	1072	86%	0%	14%	51%	4%	18%	6%	12%	9%
Duke University	1019	37%	16%	47%	56%	0%	13%	22%	7%	2%
University of California-Los Angeles	985	65%	24%	11%	51%	4%	19%	5%	14%	7%
Stanford University	958	63%	27%	10%	68%	4%	10%	8%	9%	1%
University of North Carolina at Chapel Hill	954	63%	27%	10%	64%	2%	23%	3%	7%	1%
Harvard University	940	70%	26%	4%	61%	0%	21%	4%	11%	2%
Massachusetts Institute of Technology	891	63%	29%	9%	55%	0%	10%	15%	11%	10%
Columbia University in the City of New York	884	67%	25%	8%	70%	2%	13%	4%	9%	3%
Cornell University	871	35%	49%	17%	52%	8%	22%	4%	11%	3%
University of Pittsburgh-Pittsburgh Campus	864	64%	27%	9%	68%	1%	17%	2%	4%	8%
University of Minnesota-Twin Cities	861	67%	29%	4%	57%	6%	29%	3%	2%	4%
University of Pennsylvania	842	92%	1%	7%	75%	2%	7%	8%	7%	0%
Texas A & M University-College Station	809	78%	20%	2%	37%	20%	28%	8%	6%	2%
Pennsylvania State University-Main Campus	807	28%	50%	23%	66%	6%	19%	4%	5%	0%
Yale University	755	94%	4%	2%	66%	1%	21%	4%	7%	1%
University of California-Berkeley	748	91%	9%	0%	44%	7%	21%	11%	12%	5%
Georgia Institute of Technology-Main Campus	728	63%	22%	15%	71%	2%	19%	6%	1%	1%
University of California-Davis	718	66%	22%	12%	47%	8%	25%	6%	8%	6%
University of Florida	710	86%	10%	3%	41%	15%	33%	4%	5%	2%
Washington University in St Louis	688	48%	26%	26%	62%	1%	17%	7%	12%	1%

Notes: *Top 25 research universities, ranked by average annual research between 2012-2018. Research expenditures reported in millions of 2015 US dollars. Columns provide breakdown by type of research and source of funding. Underlying data is from the NSF Higher Education Research and Development (HERD) survey.*

C Data Sources

Below we provide further details on the main micro data sources employed in the paper.

C.1 Integrated Postsecondary Education Data System (IPEDS)

The Integrated Postsecondary Education Data System is managed by the National Center for Education Statistics and brings together interrelated annual surveys. The completion of all IPEDS surveys is required by law for any institution participating in federal student financial aid programs (such as Pell grants or federal student loans). The data system provides a wealth of university level longitudinal data on institutional characteristics, prices, admissions, enrollment, student financial aid, degrees conferred, and detailed revenue and expenditure summaries.

The main variables we take from IPEDS are university research expenditure, tuition, government grants, student SAT scores, faculty salaries.

C.2 Higher Education Research and Development Survey (HERD)

The Higher Education Research and Development Survey (HERD) is administered by the National Science Foundation and gathers information on research expenditures at U.S. colleges and universities. The survey provides detailed breakdowns of university level research spending by type, source, and field as well as auxiliary institutional details. It is an annual census of all higher education institutions which separately accounted for at least $150,000 in research expenditure in the fiscal year.

We use the HERD survey primarily to disaggregate university research by source (i.e. government, internal, business) and by the type of expenditure (i.e. equipment or salaries).

We also use HERD to construct the instruments for our NIH regressions. In particular, we take microdata on university RD spending by source and field to calculate the university share of federal life science funding before the policy change.

C.3 Association of University Technology Managers (AUTM) Patent Licensing Survey

AUTM grew out of the Society of University Patent Administrators (SUPA) and is focused on developing and disseminating best practices for university technology transfer offices (TTO). Its annual Licensing Activity Survey has run for over twenty years and gathers self-reported data from member institutions on research funding, the impact of innovation, patenting activity, licensing activity, the number of campus start-ups, and other innovation related metrics.

We use the AUTM Licensing survey primarily for information on university patenting and the gross licensing revenue it takes in.

C.4 CWTS Leiden Rankings Bibliometric Micro Data

The Leiden Rankings are produced by the Center for Science and Technologies Studies (CWTS) at Leiden University. The rankings are based on bibliometric publication and citation data in the Web of Science (WoS) database produced by Clarivate Analytics. The data are processed with sophisticated bibliometric techniques to ensure comparable and consist of only high quality international scientific journals that are amenable to citation analysis.

We use the bibliometric micro data underlying the Leiden Rankings to measure university publications and citations.

C.5 National Postsecondary Student Aid Study (NPSAS)

The National Postsecondary Student Aid Study, conducted by the NCES, is a nationally representative cross-sectional survey of undergraduate and graduate students enrolled in postsecondary education. It provides individual level characteristics of postsecondary students with a special focus on how they finance their education.

We use the NPSAS to gather individual level data on tuition, education subsidies, and family income. We use these variables to estimate a reduced form schedule for higher education subsidies.

C.6 National Longitudinal Survey of Youth (NLSY)

The NLSY is a nationally representative longitudinal survey managed by the U.S. Bureau of Labor Statistics that follows a cohort of American youth born between 1980-1984. Respondents are between the ages of 12-17 when they first enter the interview rotation in 1997. The survey collects data on labor market activity, schooling, fertility, program participation, health, family background, beliefs, and much more.

We draw on the NLSY 1997 for data on student test scores and family background which informs parameters governing inter-generational dynamics.

D Proofs

D.1 University problem

We start by solving the university's problem. We guess that in equilibrium there exists a log-linear mapping from a university teaching quality, q, to its revenue per student R, knowledge capital k, research and teaching faculty average quality $\bar{h}_R, \bar{h}_T$, student peer effect $\bar{z}$, research and teaching intermediate goods e_R, e_T: namely that there exist a set of variables $m_R, m_k, m_{hR}, m_{hT}, m_z, m_{ER}, m_{eT}$ and $\chi_R, \chi_k, \chi_{hR}, \chi_{hT}, \chi_z, \chi_{eR}, \chi_{eT}$, such that

$$\log R(q) = m_R + \chi_R(\log q - m_q) \tag{27}$$

$$\log k(q) = m_k + \chi_k(\log q - m_q) \tag{28}$$

$$\log \bar{h}_R(q) = m_{hR} + \chi_{hR}(\log q - m_q) \tag{29}$$

$$\log \bar{h}_T(q) = m_{hT} + \chi_{hT}(\log q - m_q) \tag{30}$$

$$\log \bar{z}(q) = m_z + \chi_z(\log q - m_q) \tag{31}$$

$$\log \bar{e}_R(q) = m_{eR} + \chi_{eR}(\log q - m_q) \tag{32}$$

$$\log \bar{e}_T(q) = m_{eT} + \chi_{eT}(\log q - m_q) \tag{33}$$

All the ms and χs variables are functions of time, and we will omit the time subscript whenever no confusion results.Although we directly solve for the full quantitative model described in section 5, the simpler version analyzed in section 4 corresponds to the case $\beta_c = \beta$ and $\omega_h = 0$.

The universities' value function. The first step is to simplify the recursive formulation of the value function and reformulate the college problem as a maximization problem of a static objective with two components: teaching quality and research output. We guess that the value function is log-linear in knowledge capital

$$V_t(k) = \bar{v}_t + v_t \ln k \tag{34}$$

Replacing this guess into the expression for the value function and using our guesses (28)-(30) gives

$$\begin{aligned}
&(\bar{v}_t + v_t \ln k_t) = \ln q_t + \beta_c (\bar{v}_{t+1} + v_{t+1} \ln k_{t+1}) \\
&= \ln q_t + \beta_c v_{t+1} \left(\gamma_h \ln \bar{h}_R + \gamma_e \ln e_{Rt} + \gamma_k \ln k_t\right) + \beta_c \bar{v}_{t+1} \\
&= \ln q + \beta_c v_{t+1}(\gamma_e(m_{eRt} + \chi_{eRt}(\ln q_t - m_{qt})) + \gamma_h(m_{hRt} + \chi_{hRt}(\ln q_t - m_{qt})) + \gamma_k \ln k_t) + \beta_c \bar{v}_{t+1} \\
&= \ln q_t + \beta_c v_{t+1}((\gamma_e \chi_{eRt} + \gamma_h \chi_{hRt})(\ln q_t - m_{qt}) + \beta_c v_{t+1} \gamma_k \ln k_t + const. \\
&= (1 + \beta_c v_{t+1}(\gamma_e \chi_{eRt} + \gamma_h \chi_{hRt})) \frac{1}{\chi_{kt}}(\ln k_t - m_{kt}) + \beta_c v_{t+1} \gamma_k \ln k_t + const.
\end{aligned}$$

Equating the terms in $\ln k_t$ from the left and right hand side of this equation gives

$$v_t = \frac{1}{\chi_{kt}} + \beta_c \left(\gamma_k + \gamma_e \frac{\chi_{eRt}}{\chi_{kt}} + \gamma_h \frac{\chi_{hRt}}{\chi_{kt}} \right) v_{t+1} \tag{35}$$

When all elasticities χs are constant, for example in the steady-state of the model, it simplifies to

$$v = \frac{1}{\chi_k} \frac{1}{1 - \beta_c \left(\gamma_k + \gamma_e \frac{\chi_{eR}}{\chi_k} + \gamma_h \frac{\chi_{hR}}{\chi_k} \right)} \tag{36}$$

An equivalent static problem. We now use these guesses to reformulate the college's problem as a maximization problem of a static objective with two components: teaching quality and research output. Given that the implied elasticity of the value function to knowledge capital v_t is independent on a college's own choices, the solutions to the original problem coincides with the solution to the following static problem

$$\max_{q,\phi(\cdot),e_R,e_T,\mu_R(\cdot),\mu_T(\cdot)} \ln q + \beta_c v_{t+1} \ln k' \tag{37}$$

$$\text{subject to} \quad \ln q = \ln \bar{h}_T^{\omega_h} e_T^{\omega_e} \bar{z}^{\omega_z} k^{\omega_k} \tag{38}$$

$$k' = k^{\gamma_k} e_R^{\gamma_e} \bar{h}_R^{\gamma_h} \tag{39}$$

$$\mathop{\mathbb{E}}_{\phi(.)} [p(q,z)] = G(k) \left[e_R + \int wh\mu_T(h) \right] + e_T + \int whd\mu_R(h) \tag{40}$$

$$\bar{h}_R = \int wh\mu_R(h) \tag{41}$$

$$\bar{h}_T = \int wh\mu_T(h) \tag{42}$$

$$\ln \bar{z}(\phi; p) = \mathbb{E}_{\phi(.)}[\ln(z)] - \sigma_u^2(\phi; p). \tag{43}$$

This problem has two appealing characteristics. First the weight on research is endogenous to v_{t+1} which captures the future discounted payoffs of knowledge production. Second this weight is common across all colleges.

Let $s_{eR}, s_{eT}, s_{hR}, s_{hT}$ denote the share of tuition revenues, $R = \mathbb{E}_{\phi(.)}[p(k,z)]$, spent on research and teaching equipment, $s_{eR} = G(k)e_R/R$, $s_{eT} = e_T/R$ and on research and teaching faculty wages, $s_{hR} = G(k)w\bar{h}_R/R$, $s_{hT} = w\bar{h}_T/R$. Using the definition of expenditure shares, teaching quality, q, becomes

$$\ln q = \ln \left(\mathop{\mathbb{E}}_{\phi(.)}[p(q,z)] s_{hT}/w \right)^{\omega_h} \left(\mathop{\mathbb{E}}_{\phi(.)}[p(q,z)] s_{eT} \right)^{\omega_e} \bar{z}^{\omega_z} k^{\omega_k}.$$

We now guess that tuition are log-normally distributed within a college. Denoting $\ln \tilde{R} = \mathbb{E}_{\phi(.)}[\ln p(k,z)]$ the arithmetic mean of the associated normal distribution of log tuition fees within a college, this guess implies the following equality between average tuition, the variance and the mean of log-tuitions

$$\ln R - \frac{1}{2} V_{\phi(.)}(\ln p(q,z)) = \ln \mathop{\mathbb{E}}_{\phi(.)}[p(q,z)] - \frac{1}{2} V_{\phi(.)}(\ln p(q,z)) = \mathop{\mathbb{E}}_{\phi(.)}[\ln p(q,z)] = \ln \tilde{R}.$$

We verify later that the guess that tuition are log-normally distributed within a college is true (see equation (74).

The last step before taking the first order conditions is to substitute the peer-effect (43) into the expression for teaching quality and to specify the expression for the cost of heterogeneity σ_u^2.

Assumption 1. *The cost of heterogeneity across students σ_u^2 is assumed to have the following form:*

$$\sigma_u^2(\phi; p) = \frac{\Omega_t}{2} V_{\phi(.)}(\ln p(q,z)) \tag{44}$$

$$\textit{with} \quad \Omega_t = \frac{\omega_e + \omega_h + \beta v_{t+1} (\gamma_e + \gamma_h)}{\omega_z} \tag{45}$$

This choice for Ω ensures the tractability of the college problem.

Using the expression for the cost of heterogeneity (45), the definition of the peer-effect (43) and the guess that tuition are log-normally distributed, the college problem becomes fully log-linear in tuition and student' ability

$$\max_{q,\phi(z),s_{hR},s_{hT},s_{eR},s_{eT}} \ln \left(\frac{s_{hT}\tilde{R}}{w} \right)^{\omega_h} \left(\tilde{R} s_{eT} \right)^{\omega_e} \tilde{z}^{\omega_z} k^{\omega_k} + \beta_c v_{t+1} \ln \left(\frac{\tilde{R} s_{eR}}{G(k)} \right)^{\gamma_e} \left(\frac{s_{hR}\tilde{R}}{wG(k)} \right)^{\gamma_h} k^{\gamma_k}$$

$$\text{with} \quad \ln \tilde{R} = \mathop{\mathbb{E}}_{\phi(.)}[\ln p(q,z)] \text{ and } \ln \tilde{z} = \mathop{\mathbb{E}}_{\phi(.)}[\ln z].$$

Optimal policy functions. We first derive the FOC with respect to the density of students, $\phi()$. An equilibrium where colleges are indifferent across students requires that colleges are at an interior point for all students

$$0 = (\omega_e + \omega_h + \beta v_{t+1} (\gamma_e + \gamma_h)) \ln \frac{p(q,z)}{\tilde{R}} + \omega_z \ln \frac{z}{\tilde{\bar{z}}}$$

and hence that tuition be equal to

$$p(q,z) = \tilde{R} \left(\frac{z}{\tilde{\bar{z}}}\right)^{-\frac{\omega_z}{\omega_e + \omega_h + \beta v_{t+1}(\gamma_e + \gamma_h)}} = R \left(\frac{z}{\bar{z}}\right)^{-\frac{\omega_z}{\omega_e + \omega_h + \beta v_{t+1}(\gamma_e + \gamma_h)}}$$

where we use $\ln \tilde{R} = \ln R - \frac{1}{2} V_\phi(\ln p(k,z)) = \ln R$ and $\ln \tilde{z} = \ln \bar{z} + \sigma_u^2(\phi; p)$ for the second equality. The elasticity of tuition to ability $\frac{\omega_z}{\omega_e + \omega_h + \beta v_{t+1}(\gamma_e + \gamma_h)}$ increases in absolue terms with the elasticity of teaching quality to student peer effects ω_z and decreases with the elasticity of teaching quality to equipment ω_e. Importantly, it is also lower when research and knowledge production are highly valued by colleges, v_{t+1}. In other words, research increases the valuation of financial resources relative to student ability.

We now consider the optimal choice of spending on research and teaching equipment and faculty wages. Taking tuition revenues R as given, we take the F.O.C. w.r.t. s_{eR}, s_{eT}, s_{hT} and s_{hR}:[28]

$$s_{eR} = \frac{\beta v_{t+1} \gamma_e}{\beta v_{t+1}(\gamma_e + \gamma_h) + \omega_e + \omega_h} \tag{46}$$

$$s_{eT} = \frac{\omega_e}{\beta v_{t+1}(\gamma_e + \gamma_h) + \omega_e + \omega_h} \tag{47}$$

$$s_{hR} = \frac{\beta v_{t+1} \gamma_h}{\beta v_{t+1}(\gamma_e + \gamma_h) + \omega_e + \omega_h} \tag{48}$$

$$s_{hT} = \frac{\omega_h}{\beta v_{t+1}(\gamma_e + \gamma_h) + \omega_e + \omega_h} \tag{49}$$

D.2 Equilibrium Tuition Schedule (Proof of Proposition 1)

Starting from the expression for tuition we just derived and using the definition of teaching quality (38), together with our guesses (27)-(33), one gets

$$p(q,z) = R \left(\frac{z}{\bar{z}}\right)^{-\frac{\omega_z}{\omega_e + \omega_h + \beta_c v(\gamma_e + \gamma_h)}}$$

[28] The assumption that tuition revenues are not affected by the choice of spending relies on the notion that once students are sorted through colleges and paid for their tuition, they cannot leave the college even if the latter were to deviate from its equilibrium choices of s_{eR}, s_{eT} and s_{hR} and therefore deviate from its promised quality of education. We believe that this assumption is realistic, since applying and moving for colleges is very costly. Note that the notion that students will not leave their college doesn't mean that students are deceived, since they form rational expectations *ex ante* about the equilibrium.

$$
\begin{aligned}
&= R\left[\bar{h}_T^{\omega_h} e_T^{\omega_e} \bar{z}^{\omega_z} k^{\omega_k}\right]^{\frac{1}{\omega_e+\omega_h+\beta_c v(\gamma_e+\gamma_h)}} \left(\bar{h}_T^{\omega_h} e_T^{\omega_e} k^{\omega_k}\right)^{\frac{-1}{\omega_e+\omega_h+\beta_c v(\gamma_e+\gamma_h)}} z^{-\frac{\omega_z}{\omega_e+\omega_h+\beta_c v(\gamma_e+\gamma_h)}} \\
&= \bar{p} q^{\frac{1}{\omega_e+\omega_h+\beta_c v(\gamma_e+\rho_h)}+\chi_R-\frac{(\omega_e+\omega_h)\chi_R}{\omega_e+\omega_h+\beta_c v(\gamma_e+\gamma_h)}-\chi_k\frac{\omega_k}{\omega_e+\omega_h+\beta_c v(\gamma_e+\gamma_h)}} z^{-\frac{\omega_z}{\omega_e+\omega_h+\beta_c v(\gamma_e+\gamma_h)}} \\
&= \bar{p} q^{\frac{1}{\epsilon_1}} z^{-\frac{\epsilon_2}{\epsilon_1}} \qquad (50)
\end{aligned}
$$

where we have defined the following aggregate endogenous variables:

$$
\log \bar{p} = \frac{\left(\beta_c v(\gamma_e+\gamma_h) m_R - \omega_k m_k - \left(\frac{\omega_e+\omega_h+\beta_c v(\gamma_e+\gamma_h)}{\epsilon_1} - 1\right) m_q - \log\left(s_{eT}\right)^{\omega_e}\left(\frac{s_{hT}}{w}\right)^{\omega_h}\right)}{\omega_e+\omega_h+\beta_c v\left(\gamma_e+\gamma_h\right)} \qquad (51)
$$

$$
\epsilon_1 = \frac{\omega_e+\omega_h+\beta_c v\left(\gamma_e+\gamma_h\right)}{1+\chi_R\beta_c v\left(\gamma_e+\gamma_h\right)-\chi_k\omega_k} \qquad (52)
$$

$$
\epsilon_2 = \frac{\omega_z}{1+\chi_R\beta_c v\left(\gamma_e+\gamma_h\right)-\chi_k\omega_k} \qquad (53)
$$

where we use $w\bar{h}_T = s_{hT}R$, $e_T = s_{eT}R$, (27), and (28).

D.3 Household problem in section 4 (Proof of Propositions 2 and 3)

In this section, we derive the expression for the optimal spending rate of household on tuition shown in proposition 2. The derivation of the spending rate in the full quantitative model is given in the next section.

Using equation (13), the household problem is given by

$$
\begin{aligned}
U(h,z) &= \max_{c,q,h'} \ln c + \beta U(h',z') \\
\text{s.t.} \quad & c + \bar{p} q^{\frac{1}{\epsilon_1}} z^{-\frac{\epsilon_2}{\epsilon_1}} = wh \\
\text{and} \quad & \ln h' = \ln z + \alpha \ln q
\end{aligned}
$$

We guess that $U(h,z) = \bar{u} + u\ln h + u_z \ln z$ and we denote s the fraction of their income households spend on tuition $\bar{p} q^{\frac{1}{\epsilon_1}} z^{-\frac{\epsilon_2}{\epsilon_1}} = swh$. It follows from the latter that

$$
\ln q = \epsilon_1 \ln(swh) - \epsilon_1 \ln \bar{p} + \epsilon_2 \ln z \qquad (54)
$$

and combining with the guess on the value functions gives

$$
\bar{u}+u\ln h+u_z\ln z = \ln((1-s)wh)+\beta u(\bar{u}+\ln z+\alpha(\epsilon_1\ln s+\epsilon_1\ln(wh)-\epsilon_1\ln\bar{p}+\epsilon_2\ln z))-\beta u_z\frac{\sigma_z^2}{2}.
$$

Comparing the coefficients on $\ln h$ we obtain the following condition on u

$$u = 1 + \beta u \alpha \epsilon_1.$$

The next step is to take the first order with respect to s. It is given by

$$\frac{1}{1-s} = \beta u \alpha \epsilon_1 \frac{1}{s}$$

and thus

$$s = \frac{\beta u \alpha \epsilon_1}{1 + \beta u \alpha \epsilon_1}.$$

Since s is independent of h and z, the guess on the value function is correct.

Finally, using the expression for u, we obtain

$$s = \frac{\beta u \alpha \epsilon_1}{1 + \beta u \alpha \epsilon_1} = \beta \alpha \epsilon_1. \tag{55}$$

D.4 Household problem in the quantitative model

We now solve for the optimal spending of households on tuition in the full quantitative model. We start by giving the timing of a lifetime in detail. A period corresponds to four years. Each individual lives for five periods as a child, then attends college for one period, and finally works as an adult for ten periods, which we index by $a \in \{1, ..10\}$. Each adult has one child when they are in their fifth period of adult life, $a = 5$, and sends them to college before retiring at $a = 10$. Parents retire after their children graduate from college, so that parents and children do not overlap in the labor market. For simplicity, we assume that $r = 1/\delta - 1$.

The household problem can be formulated recursively as

$$U(h, z) = \max_{c_a, q} \left\{ \left[\sum_{a=1}^{10} \delta^{a-1} \ln c_a \right] + \beta \mathbb{E} \left[U(h', z') \right] \right\} \tag{56}$$

where δ and β are the time and intergenerational discount factor, respectively.

Denoting $c_{t,a}$ is the consumption of an individual whose age is a at time t, the household's life-time income and budget constraint are given by:

$$\sum_{a=1}^{10} \delta^{a-1}(1 + a_{ct+a-1}) c_{t+a-1,a} + \delta^{10-1} \frac{y^{\tau_n}}{1 + a_n} p_{t+10-1}(q, z) := (1 - a_y) h^{1-\tau_y} \sum_{a=1}^{10} \delta^{a-1} w_{t+a-1}$$

where $(1-a_y)\left(w_t h\right)^{1-\tau_y}$ is the after tax-and-transfers labor earnings in a given period.

We define the lifetime after-tax income y_t and lifetime before-tax wage per unit of human capital $\tilde{w}_t$ as

$$y_t = (1-a_y)(\tilde{w}_t h)^{1-\tau_y}$$
$$\text{with} \quad \tilde{w}_t = \left(\delta^{-9}\sum_{a=1}^{10}\delta^{a-1}w_{t+a-1}^{1-\tau_y}\right)^{\frac{1}{1-\tau_y}}$$

We first solve for the optimal allocation of consumption across periods of an adult's lifetime. From the assumption that $r = 1/\delta - 1$, after-tax consumption is constant within one's lifetime:

$$(1+a_{c,t+a-1})c_{t+a-1,a} = \delta^{a-1}\Pi_{\tau=1}^{a-1}(1+r_{t+\tau})(1+a_{ct})c_{t1} = (1+a_{ct})c_{t1}$$

Hence the problem of the household can be written more simply as

$$U(h,z) = \max_{c,q}\left\{\ln c\sum_{a=1}^{10}\delta^{a-1} + \beta\mathbb{E}\left[U(h',z')\right]\right\} \tag{57}$$
$$\text{s.t.} \quad y = \delta^{-9}(1+a_{ct})c\sum_{a=1}^{10}\delta^{a-1} + \frac{y^{\tau_n}}{1+a_n}p_{t+10}(q,z) \tag{58}$$

For the next step it is useful to first derive the sorting rule (15). We combine the expression for tuition (50) and financial aid (23) and we denote s the share of net income that households spend on tuition $\bar{p}q^{\frac{1}{\epsilon_1}}z^{-\frac{\epsilon_2}{\epsilon_1}} = s(1+a_n)y^{1-\tau_n}$. Hence, the sorting rule is given by

$$q = \left(\frac{s(1+a_n)y^{1-\tau_n}}{\bar{p}}\right)^{\epsilon_1} z^{\epsilon_2} \tag{59}$$

We guess that the value function is a log-linear function of h and z: $U(h,z) = \tilde{U}(h,\xi) = u_t \ln h_t + u_{zt}\ln\xi + \bar{u}_t$ with $u_t, u_{zt}, \bar{u}_t$ three endogenous and aggregate variables. Recall that ξ denotes the shock to the transmission of human capital from parents to child given in equation (25). Like in section (D.3), we use this guess to substitute for the value function in the current and future period in equation (57) and we find that a necessary condition is that u_t obeys the following forward difference equation:

$$u_t = (1-\tau_y)\sum_{a=1}^{10}\delta^{a-1} + \beta u_{t+10}\rho_t$$
$$\text{with} \quad \rho_t = \varphi + \alpha\left[\epsilon_{2t}\varphi + \epsilon_{1t}(1-\tau_n)(1-\tau_y)\right]$$

where we have used the sorting rule (59).

Like in section (D.3), we then take the derivative of the right hand side of the value function with respect to s which gives

$$s_t = \frac{\beta\alpha\epsilon_{1t}(1-\tau_n)u_{t+10}}{\sum_{a=1}^{10}\delta^{a-1} + \beta\alpha\epsilon_{1t}(1-\tau_n)u_{t+10}} \tag{60}$$

and s_t is independent of h and z and the guess on the policy function is correct.

One can easily check that the expression for s obtained in equation (55) corresponds to the special case with $\phi = \tau_n = \tau_y = 0$.

Law of accumulation of human capital. We start by defining the elasticity of college quality to income and ability. They capture the strength of the income-sorting and ability-sorting channel taking into account the progressivity of taxes and financial aid:

$$\epsilon_I = \epsilon_1(1-\tau_n)(1-\tau_y) \tag{61}$$
$$\epsilon_A = \varphi\epsilon_2 \tag{62}$$

Notice that when there is no income tax and financial aid, the income-sorting coefficient simplifies and becomes $\epsilon_I = \epsilon_1$.

Taking the log of the sorting rule (59) and using the previous definitions (61), (62) and ias well as the transmission of human capital over generation $z = \xi h^{\varphi}$, the optimal college is given by

$$\ln q = \epsilon_1(C_h - \ln\bar{p}) + (\epsilon_I + \epsilon_A)\ln h + \epsilon_2\ln\xi \tag{63}$$
$$\text{with} \quad C_h \equiv \ln\left(s(1+a_n)\tilde{w}^{(1-\tau_y)(1-\tau_n)}(1-a_y)^{(1-\tau_n)}\right). \tag{64}$$

Note that, in the simpler model without policies and life-cycle, $C_h = \ln s + \ln h$.

Replacing this expression in the law of accumulation of human capital given by (6) we obtain

$$\ln h' = \alpha(\epsilon_1(C_h - \ln \bar{p}) + (\epsilon_I + \epsilon_A)\ln h + \epsilon_2 \ln \xi) + \varphi \ln h + \ln \xi. \tag{65}$$

D.5 Equilibrium distributions of human capital and universities intangible capital

Law of motion of the distribution of human capital We denote m_{hta} and Σ_{hta} the mean and standard deviation of the log of human capital of adults with age a at time t. We denote m_{ht}, Σ_{ht} the mean and standard deviation of the log of human capital of the parents who are currently sending their child to college, namely those with $a = 10$:

$$m_{ht} = m_{ht10}$$
$$\Sigma_{ht} = \Sigma_{ht10}$$

From these definitions and the law of motion of human capital at the individual level (65) we have that the law of motion of the distribution of human capital across households is given by

$$\ln h_{t+1,a} \sim \mathcal{N}\left(m_{h,t+1,a+1}, \Sigma^2_{ht+1,a+1}\right) \tag{66}$$

$$m_{h,t+1,a+1} = m_{h,t,a} \qquad \text{for} \quad a < 10 \tag{67}$$

$$\Sigma_{h,t+1,a+1} = \Sigma_{h,t,a} \qquad \text{for} \quad a < 10 \tag{68}$$

$$m_{h,t+1,1} = \rho_t m_{ht} - (\alpha\epsilon_{2t} + 1)\frac{\sigma_z^2}{2} + \alpha\epsilon_{1t}(C_{ht} - \ln \bar{p}_t) \qquad \text{for} \quad a = 10 \tag{69}$$

$$\Sigma^2_{h,t+1,1} = \rho_t^2 \Sigma^2_{ht} + \sigma_y^2 + \left(\varphi\left(\alpha\epsilon_{2t} + 1\right)\right)^2 \sigma_z^2 \qquad \text{for} \quad a = 10 \tag{70}$$

$$\text{where } \rho_t = \varphi + \alpha\left[\epsilon_{2t}\varphi + \epsilon_{1t}(1 - \tau_n)(1 - \tau_y)\right]$$

It is intuitive that the shifter C_{ht} in the law of motion of the mean of the distribution (69) is increasing in the saving rate s_t, in the average education subsidies a_n but decreasing in the intercept of the tuition schedule $\bar{p}_t$. From the expression given by (51), the latter is increasing in the share of resources devoted to research. Finally, the persistence coefficient ρ_t is decreasing in the progressivity of financial aid τ_n.

This also confirms that human capital is log-normally distributed on the transition path as well as in steady state as long as the initial distribution is log-normal.

Distribution of college quality. From (63), it follows that the distribution of college quality is given by

$$\ln q \sim \mathcal{N}\left((\epsilon_I + \epsilon_A)m_h + \epsilon_1(C_h - \log \bar{p}) - \epsilon_2 \frac{\sigma_z^2}{2}, \epsilon_2^2\sigma_z^2 + (\epsilon_I + \epsilon_A)^2\Sigma_h^2\right) \tag{71}$$

Thus

$$m_q = (\epsilon_I + \epsilon_A)m_h + \epsilon_1(C_h - \log \bar{p}) - \epsilon_2 \frac{\sigma_z^2}{2}. \tag{72}$$

College quality is log-normally distributed on the transition path as well as in steady state.

Within college parental income distribution. Using (63), we now solve for the conditional distribution of parental human capital within a college, which is given by

$$\ln h|q \sim \mathcal{N}\left(m_{h|q}, \sigma_{h|q}^2\right)$$

where

$$\begin{aligned} m_{h|q} &= s_z m_h + (1 - s_z)\frac{\left(\ln q - \epsilon_1(C_h - \log \bar{p}) + \epsilon_2 \frac{\sigma_z^2}{2}\right)}{\epsilon_I + \epsilon_A} \\ \sigma_{h|q}^2 &= s_z \Sigma_h^2 \end{aligned}$$

$$\text{with} \quad s_z = \frac{\epsilon_2^2\sigma_z^2}{(\epsilon_I + \epsilon_A)^2\Sigma_h^2 + \epsilon_2^2\sigma_z^2} \tag{73}$$

where s_z is the share of the variance not explained by parent's human capital.

Parental income within a college is log-normally distributed in transition as well as in steady state. The distribution function is defined as $\phi_h(h|q)$.

Notice that

$$E(m_{h|q}) = m_h.$$

Within college student ability distribution. From the definition of abilities $\ln z = \varphi \ln h + \ln \xi$ and the sorting rule used above $\ln q = (\epsilon_I + \epsilon_A)\ln h + \epsilon_2 \ln \xi + \epsilon_1(C_h - \log \bar{p})$, one gets

$$\ln z = \frac{1}{\epsilon_2}\left(\ln q - \epsilon_I \ln h - \epsilon_1(C_h - \log \bar{p})\right) \tag{74}$$

$$\Rightarrow \ln z|q \sim \mathcal{N}\left(\frac{1}{\epsilon_2}\left(\ln q - \epsilon_I m_{h|q} - \epsilon_1(C_h - \log\bar{p})\right), \left(\frac{\epsilon_I}{\epsilon_2}\right)^2 \sigma^2_{h|q}\right) \tag{75}$$

Since $\ln p(q,z) = \ln\bar{p} + \frac{1}{\epsilon_1}\ln q - \frac{\epsilon_2}{\epsilon_1}\ln z$, tuition are log-normally distributed within a college and its variance is common across universities.

Distribution of tuition revenues. The average tuition revenue of a college is the mean of tuition paid by households in this college. All households pay the same share of their income, hence using the distribution of income within a college, one gets that the mean tuition is

$$\begin{aligned}
R(q) &= \int p(q,z)\phi(z|q)dz \\
&= \int s(1+a_n)(1-a_y)^{1-\tau_n}\left(\tilde{w}h\right)^{(1-\tau_y)(1-\tau_n)}\phi_h(h|q)dh \\
\ln R(q) &= C_h + (1-\tau_y)(1-\tau_n)m_{h|q} + \frac{((1-\tau_y)(1-\tau_n))^2\sigma^2_{h|q}}{2} \\
&= C_h + (1-\tau_y)(1-\tau_n)\left[s_z m_h + (1-s_z)\frac{\ln q - \epsilon_1(C_h - \log\bar{p}) + \epsilon_2\frac{\sigma_z^2}{2}}{\epsilon_I + \epsilon_A}\right] \\
&+ \frac{((1-\tau_y)(1-\tau_n))^2\sigma^2_{h|q}}{2}
\end{aligned}$$

This also verifies our guess (27). Finally these results enables us to get an expression for the distribution of revenue per student across colleges:

$$\ln R \sim \mathcal{N}\left(m_R, \Sigma_R^2\right) \tag{76}$$

$$m_R = C_h + \frac{((1-\tau_y)(1-\tau_n))^2\sigma^2_{h|q}}{2} + (1-\tau_y)(1-\tau_n)m_h \tag{77}$$

$$\Sigma_R^2 = \chi_R^2\left[\epsilon_2^2\sigma_z^2 + (\epsilon_I + \epsilon_A)^2\Sigma_h^2\right] \tag{78}$$

College tuition revenue is log-normally distributed on the transition path as well as in steady state.

Identifying coefficients with the guess, one gets:

$$\chi_R = (1-\tau_y)(1-\tau_n)(1-s_z)\frac{1}{\epsilon_I + \epsilon_A}$$

Combining this with equations (73) and (78),

$$\Sigma_R = (1 - \tau_y)(1 - \tau_n)\sqrt{1 - s_z}\Sigma_h. \tag{79}$$

If there is no peer effect ($\omega_z = 0$) and $\epsilon_A = 0$, then $s_z = 0$ and Σ_R is proportional to Σ_h. With the peer effect, as Σ_h decreases, s_z decreases. Intuitively, there is more mix of students from different family backgrounds within colleges as the peer effect becomes stronger and total revenue of colleges get less dispersed.

Verification of guesses. From the definition of $\ln \bar{z}$ and equation (75),

$$\ln \bar{z} = \frac{1}{\epsilon_2}\left(\ln q - \epsilon_I m_{h|q} - \epsilon_1(C_h - \log \bar{p})\right) - \frac{\Omega}{2}\left(\frac{\epsilon_I}{\epsilon_1}\right)^2 \sigma^2_{h|q}.$$

Thus (31) is verified. From the technology for college quality $\ln q = \ln \left(R s_{hT}/w\right)^{\omega_h} \left(R s_{eT}\right)^{\omega_e} \bar{z}^{\omega_z} k^{\omega_k}$, equation (31) and (27), our guess (28) is also verified. From $s_{eR} = G(k)e_R/R$, $s_{eT} = e_T/R$, $s_{hR} = G(k)w\bar{h}_R/R$, $s_{hT} = w\bar{h}_T/R$, equations (29), (30), (32), (33) are also verified. This verifies that all college-specific variables are log-linear functions of quality q and log-normally distributed.

Law of motion of knowledge capital From the law of motion of capital, we obtain

$$\ln k_{t+1} = \ln e^{\gamma_e}_{Rt}\bar{h}^{\gamma_h}_{Rt}k^{\gamma_k} = \ln\left(\frac{s_{eRt}}{G_t}\right)^{\gamma_e}\left(\frac{s_{hRt}}{w_t G_t}\right)^{\gamma_h} + (\gamma_e + \gamma_h)\ln R_t + (\gamma_k + \tau_G(\gamma_e + \gamma_h))\ln k_t$$

We denote m_R and Σ_R (m_k and Σ_k) the mean and standard deviation of log-income (log-knowledge capital) across colleges. Taking the mean and variance of the law of motion above gives that if knowledge capital is log-normally distributed, then it remains log-normally distributed and the law of motion of the mean and the variance of the associated normal distribution are given by the following proposition.

Proposition 7. *Assume knowledge capital is log-normally distributed in the first period. It is log-normally distributed along the transition path. The law of motion of the distribution of knowledge capital across colleges is given by*

$$\ln k_{it} \sim \mathcal{N}\left(m_{kt}, \Sigma^2_{kt}\right) \tag{80}$$

$$m_{kt+1} = \ln\left(\frac{s_{eRt}}{G_t}\right)^{\gamma_e}\left(\frac{s_{hRt}}{w_t G_t}\right)^{\gamma_h} + (\gamma_e + \gamma_h)m_{Rt} + (\gamma_k + \tau_G(\gamma_e + \gamma_h))m_{kt} \tag{81}$$

$$\Sigma_{kt+1} = (\gamma_k + \tau_G(\gamma_e + \gamma_h))\Sigma_{kt} + (\gamma_e + \gamma_h)\Sigma_{Rt} \tag{82}$$

where we used the fact, that in the equilibrium we look at, (log) income and (log) knowledge are perfectly correlated $cov\,(\ln R, \ln k) = \sqrt{\Sigma_R^2 \Sigma_k^2} = \Sigma_R \Sigma_k$.

D.6 Law of motion for m_k, m_h - Simplification

We now express the laws of motion of the distribution of knowledge capitals and human capital in a compact format.[29]

$$\begin{aligned}
m_{kt+1} &= \gamma_{kkt} m_{kt} + \gamma_{kht} m_{ht} + \gamma_{kt} && (83)\\
m_{h,t+1,1} &= \gamma_{hht} m_{ht} + \gamma_{hkt} m_{kt} + \gamma_{ht} \quad \text{for} \quad a = 10 && (84)\\
m_{h,t+1,a+1} &= m_{h,t,a} \quad \text{for} \quad a < 10 && (85)
\end{aligned}$$

with

$$\begin{aligned}
\gamma_{kkt} &= \gamma_K + (\gamma_e + \gamma_h)\tau_G && (86)\\
\gamma_{kht} &= (\gamma_e + \gamma_h)(1 - \tau_y)(1 - \tau_n) && (87)\\
\gamma_{kt} &= \ln\left(\frac{s_{eR}}{G}\right)^{\gamma_e}\left(\frac{s_{hR}}{wG}\right)^{\gamma_h} + (\gamma_e + \gamma_h)\left[C_h + \frac{((1-\tau_y)(1-\tau_n))^2 \sigma_{h|q}^2}{2}\right] && (88)\\
\gamma_{hht} &= (1 - \tau_y)(1 - \tau_n)\alpha(\omega_e + \omega_h) + \varphi(1 + \alpha\omega_z) && (89)\\
\gamma_{hkt} &= \alpha\omega_k && (90)\\
\gamma_{ht} &= -(1 + \alpha\omega_z)\frac{\sigma_z^2}{2} + \alpha(\omega_e + \omega_h)C_h\\
&\quad - \alpha\beta_c v(\gamma_e + \gamma_h)\frac{((1-\tau_y)(1-\tau_n))^2 \sigma_{h|q}^2}{2} + \alpha\left[\ln\left(s_{eT}\right)^{\omega_e}\left(\frac{s_{hT}}{w}\right)^{\omega_h}\right] && (91)
\end{aligned}$$

We now briefly give an intuition for each term from (86)-(91). Looking at equation (86), current average knowledge capital has a strong effect on future knowledge capital when knowledge depreciates slowly (low γ_k). Looking at equation (87), current average human capital has a strong effect on future average knowledge capital when fundamental research is intensive in equipment and faculty ($\gamma_e + \gamma_h$). Looking at equation (88), the growth of fundamental knowledge is high when the rate of cross-subsidization is high (s_{eR}, s_{hR}) or when households spend a large share of their income on tuition, s.

[29] Recall that

$$m_{kt+1} = \ln\left(\frac{s_{eR}}{G}\right)^{\gamma_e}\left(\frac{s_{hR}}{wG}\right)^{\gamma_h} + (\gamma_e + \gamma_h)m_{Rt} + (\gamma_k + \tau_G(\gamma_e + \gamma_h))m_{kt}$$

where m_R in the first line is given by equation (77).

Looking at equation (89), current average human capital has a strong effect on future human capital when the transmission of of abilities from parents to children is strong (φ), the peer effect and the effect of teaching equipment and faculty is high ($\omega_z, \omega_e, \omega_h$). Looking at equation (90), current average knowledge capital has a strong effect on future human capital when knowledge capital matters a lot for teaching equality ω_q. Finally, looking at equation (91), the growth of human capital is high when household spend a significant share of their income on tuition s and universities spend a lot on teaching equality s_{eT}, s_{hT}.

Expressing $\bar{p}$ as a function of m_h and m_k. From the earlier expression for $\bar{p}$ given by equation (51), and using the expressions for the mean of college quality (72), and college revenues (77), we obtain

$$\epsilon_1 \log \bar{p} = (\epsilon_1 - (\omega_e + \omega_h))(C_h + (1-\tau_y)(1-\tau_n)m_h) + \beta_c v(\gamma_e + \gamma_h)\left(\frac{((1-\tau_y)(1-\tau_n))^2\sigma^2_{h|q}}{2}\right)$$
$$- \omega_k m_k - \left(\frac{\omega_e + \omega_h + \beta_c v\,(\gamma_e + \gamma_h)}{\epsilon_1} - 1\right)\left(\epsilon_A m_h - \epsilon_2 \frac{\sigma^2_z}{2}\right) - \log\left(s_e\right)^{\omega_e}\left(\frac{s_h}{w}\right)^{\omega_h} \quad (92)$$

Now we discuss the intuition of this equation. For exposition, consider the simpler model in section 3. Then this equation becomes

$$\epsilon_1 \log \bar{p} = (\epsilon_1 - \omega_e)(C_h + m_h) + \beta v \gamma_e \left(\frac{\sigma^2_{h|q}}{2}\right) - \omega_k m_k - \left(\frac{\omega_e + \beta v \gamma_e}{\epsilon_1} - 1\right)\left(-\epsilon_2 \frac{\sigma^2_z}{2}\right) - \omega_e \log s_e$$

Organizing this equation,

$$\omega_k m_k + \omega_e(\ln s_e + C_h + m_h + \frac{\sigma^2_{h|q}}{2}) - (\omega_e + \beta v \gamma_e)\left(\frac{\sigma^2_{h|q}}{2}\right) - \epsilon_2\left(\frac{\omega_e + \beta v \gamma_e}{\epsilon_1}\right)\frac{\sigma^2_z}{2}$$
$$= \epsilon_1(C_h + m_h) - \epsilon_1 \log \bar{p} - \epsilon_2 \frac{\sigma^2_z}{2}$$

From (52), (53), and (77),

$$\omega_k m_k + \omega_e(\ln s_e + m_R) - (\omega_e + \beta v \gamma_e)\left(\frac{\sigma^2_{h|q}}{2}\right) - \omega_z \frac{\sigma^2_z}{2} = \epsilon_1(\ln s + m_h) - \epsilon_1 \log \bar{p} - \epsilon_2 \frac{\sigma^2_z}{2} \quad (93)$$

Since $m_{\bar{z}} = E[\ln \bar{z}] = E[E_{\phi(\cdot)}[\ln z]] - \sigma_u^2(\phi; p)$ and the mean of $\ln z$ and from(44),

$$m_{\bar{z}} = -\frac{\sigma_z^2}{2} - \frac{1}{2}\frac{\omega_e + \beta v \gamma_e}{\omega_z} V_{\phi(\cdot)}(\ln p(q,z)) = -\frac{\sigma_z^2}{2} - \frac{1}{2}\frac{\omega_e + \beta v \gamma_e}{\omega_z} V_{\phi(\cdot)}\left(\frac{\epsilon_2}{\epsilon_1}\ln z\right)$$

From (75),

$$m_{\bar{z}} = -\frac{\sigma_z^2}{2} - \frac{\omega_e + \beta v \gamma_e}{\omega_z}\frac{\sigma_{h|q}^2}{2}$$

Plugging this into (93) and from (63),

$$\omega_k m_k + \omega_e(\ln s_e + m_R) + \omega_z m_{\bar{z}} = \epsilon_1(\ln s + \ln w + m_h) - \epsilon_1 \log \bar{p} - \epsilon_2 \frac{\sigma_z^2}{2} \tag{94}$$

The left hand side is the mean supply of log quality of colleges m_q, where the first, second, and third terms are the mean contributions of log intangible capital, teaching expenditure, and the peer effect. The right hand side is the mean demand of log quality from taking log and the mean of (15). $\log \bar{p}$ is determined to balance the supply and demand of mean quality of colleges.

D.7 Proof of Proposition 4

Following our guess for the value function (34), in steady-state it is given by $V(k) = \bar{v} + v \ln k$. Substituting this guess into the recursive formulation of the value function $V(k) = \ln q + \beta V(k')$ and differentiating the latter with respect to $\ln k$ gives

$$v = \frac{d \ln q}{d \ln k} + \beta v \frac{d \ln k'}{d \ln k} = \frac{\Sigma_q}{\Sigma_k} + \beta v$$

From our guess (28) it follows that

$$\frac{d \ln q}{d \ln k} = \frac{1}{\chi_k} = \frac{\Sigma_q}{\Sigma_k}.$$

Combining both equations gives

$$v = \frac{\Sigma_q / \Sigma_k}{1 - \beta}$$

Substituting this expression of v in equation (46) (and also substituting $\omega_h = 0$ and $\gamma_h = 0$) gives

$$s_R = \frac{\beta v \gamma_e}{\beta v \gamma_e + \omega_e} = \frac{\beta \gamma_e (\Sigma_q / \Sigma_k)}{(1-\beta)\omega_e + \beta \gamma_e (\Sigma_q / \Sigma_k)}$$

We now find an expression for $\frac{\Sigma_q}{\Sigma_k}$ as a function of $\frac{\Sigma_R}{\Sigma_k}$ and $\frac{\Sigma_{\bar{z}}}{\Sigma_k}$. We start from the production function of quality of education $\ln q = \omega \ln k + \omega_e \ln e_T + \omega_z \ln \bar{z}$. Using $e_T = (1 - s_R)R$, and taking the variances of both sides, we get

$$\Sigma_q = \omega_k \Sigma_k + \omega_e \Sigma_R + \omega_z \Sigma_{\bar{z}}.$$

where we used the fact that $\ln k$, $\ln e_T$, and $\ln \bar{z}$ are perfectly correlated. Dividing both sides by Σ_k gives

$$\frac{\Sigma_q}{\Sigma_k} = \omega_k + \omega_e \frac{\Sigma_R}{\Sigma_k} + \omega_z \frac{\Sigma_{\bar{z}}}{\Sigma_k}.$$

D.8 Proof of Proposition 6: the case of $\omega_z = 0$.

We begin with solving for the standard deviation of university intangible capital and household human capital. Let's start with the law of motion of the variance of knowledge.

$$\Sigma_{kt+1} = \gamma_k \Sigma_k + \gamma_e \Sigma_R$$

We now turn to the standard deviation of human capital, Σ_h. When there is no peer effect, the variance of college income is equal to the variance of household income $\Sigma_R = \Sigma_h$. The solution is

$$\frac{\Sigma_k}{\Sigma_h} = \frac{\gamma_e}{1 - \gamma_k}.$$

From Propositions 2 and 3,

$$\sqrt{\epsilon_1^2 \frac{\Sigma_h^2}{\Sigma_k^2} + \epsilon_2^2 \frac{\sigma_z^2}{\Sigma_h^2}} = \omega_k + \omega_e \frac{\Sigma_R}{\Sigma_k} + \omega_z \frac{\Sigma_{\bar{z}}}{\Sigma_k},$$

$$\frac{\epsilon_2}{\epsilon_1} = \frac{\omega_z}{\omega_e + \beta(1-\beta)^{-1}\gamma_e(\Sigma_q/\Sigma_k)}.$$

Since $\omega_z = 0$, the latter becomes

$$\epsilon_2 = 0$$

and the former becomes

$$\epsilon_1 \frac{\Sigma_h}{\Sigma_k} = \omega_k + \omega_e \frac{\Sigma_R}{\Sigma_k} = \omega_k + \omega_e \frac{\Sigma_h}{\Sigma_k}.$$

In addition, from equation (15), we have

$$\Sigma_q = \epsilon_1 \Sigma_h$$

Combining the last two equations gives

$$\epsilon_1 = \omega_k \frac{\Sigma_k}{\Sigma_h} + \omega_e = \omega_k \frac{\gamma_e}{1-\gamma_k} + \omega_e. \tag{95}$$

Since $h' = zq^\alpha$ and $\Sigma_q = \epsilon_1 \Sigma_h$ and z and h are independent, we get

$$\Sigma_h^2 = \sigma_z^2 + \alpha^2 \Sigma_q^2 = \sigma_z^2 + \alpha^2 \epsilon_1^2 \Sigma_h^2$$

which gives

$$\Sigma_h^2 = \frac{\sigma_z^2}{1 - \alpha^2 \left(\omega_k \frac{\gamma_e}{1-\gamma_k} + \omega_e\right)^2}.$$

We now solve for the mean of the distribution of human capital m_h and of university knowledge capital m_k. First note that the average tuition revenue is a simple function of the spending rate on tuition s and the average household human capital:

$$m_R = \ln(sw) + m_h.$$

Taking the mean of the log of equation (13) and using $swh = p(q,z)$, we get

$$m_R = E[\ln p(q,z)] = \ln \bar{p} + \frac{1}{\epsilon_1} m_q - \frac{\epsilon_2}{\epsilon_1} m_z = \ln \bar{p} + \frac{1}{\epsilon_1}(\omega_k m_k + \omega_e \ln(1-s_R) + \omega_e m_R + \omega_z m_{\bar{z}}) - \frac{\epsilon_2}{\epsilon_1} m_z$$

Since $\omega_z = 0$ and $\epsilon_2 = 0$, one gets

$$m_R = \ln \bar{p} + \frac{1}{\epsilon_1}(\omega_k m_k + \omega_e m_R + \omega_e \ln(1-s_R))$$

which gives

$$\ln \bar{p} = \left(1 - \frac{\omega_e}{\epsilon_1}\right) m_R - \frac{\omega_k}{\epsilon_1} m_k - \frac{\omega_e}{\epsilon_1} \ln(1-s_R).$$

Using the law of accumulation of human capital $h' = zq^\alpha$

$$\ln h' = \ln z + \alpha \ln q = \ln z + \alpha \epsilon_1 (\ln(sw) + \ln h - \ln \bar{p})$$

and taking the expectation of both sides gives

$$m_h = -\frac{\sigma_z^2}{2} + \alpha\epsilon_1(\ln(sw) + m_h - \ln\bar{p}) = -\frac{\sigma_z^2}{2} + \alpha\epsilon_1(m_R - \ln\bar{p})$$

Using the law of accumulation of university capital,

$$m_k = \gamma_e \ln s_R + \gamma_e m_R + \gamma_k m_k = \gamma_e \ln s_R + \gamma_e(\ln(sw) + m_h) + \gamma_k m_k,$$

it follows that

$$m_k = \frac{\gamma_e \ln s_R + \gamma_e m_R}{1-\gamma_k}.$$

Substituting these expressions into our earlier expression for $\bar{p}$ gives

$$\begin{aligned}
\ln\bar{p} &= \left(1 - \frac{\omega_e}{\epsilon_1}\right) m_R - \frac{\omega_k}{\epsilon_1}\frac{\gamma_e \ln s_R + \gamma_e m_R}{1-\gamma_k} - \frac{\omega_e}{\epsilon_1}\ln(1-s_R) \\
&= \left(1 - \frac{\omega_e}{\epsilon_1} - \frac{\omega_k}{\epsilon_1}\frac{\gamma_e}{1-\gamma_k}\right) m_R - \frac{\omega_k}{\epsilon_1}\frac{\gamma_e}{1-\gamma_k}\ln s_R - \frac{\omega_e}{\epsilon_1}\ln(1-s_R) \\
&= -\frac{\omega_k \frac{\gamma_e}{1-\gamma_k}\ln s_R + \omega_e \ln(1-s_R)}{\omega_k \frac{\gamma_e}{1-\gamma_k} + \omega_e}
\end{aligned}$$

where the last line uses equation (95).

D.9 Government budget constraints

We denote $\bar{G}$ the average spending on research grants (per student), $\bar{a}_y$ and $\bar{a}_n$ the average tax rate on income and the average rate of tuition subsidy. These are three parameters that we calibrate. The following equations pin down the endogenous value of the intercepts of the research grant schedule G, of the income tax schedule a_y and of the financial aid schedule a_n respectively:

$$\bar{G} = \mathbb{E}_j\left[\left[1 - Gk_j^{-\tau_G}\right]\left(e_{Rj} + \mathbb{E}_{\mu_{Rj}(.)}[wh]\right)\right] \tag{96}$$

$$(1-\bar{a}_y)\int wh_{Ri}di = \int (1-a_y)(wh_{Ri})^{1-\tau_y}di \tag{97}$$

$$\int \frac{y_i^{\tau_n}}{(1+a_n)}e_i di = \int \frac{e_i}{(1+\bar{a}_n)}di. \tag{98}$$

The government balances its budget every period:

$$\bar{a}_c\int c_i di + \bar{a}_y\int wh_{Ri}di = \frac{\bar{a}_n}{(1+\bar{a}_n)}\int p_i di + \bar{G}. \tag{99}$$

Solving for G. We start from the budget constraint of the agency that distributes research grants and use the guesses (27), (28) and $s_{eR} = G(k)e_R/R$:

$$\int G(k)\left[e_R + \int \mu_k(h)w_f(h)\right]dj + \bar{G} = \int \left[e_R + \int \mu_k(h)w_f(h)\right]dj$$
$$\iff \int (s_{eR} + s_{hR})R(q)dj + \bar{G} = \int (s_{eR} + s_{hR})R(q)G^{-1}k^{\tau_G}dj$$
$$\iff \int (s_{eR} + s_{hR})e^{m_R - \frac{\chi_R}{\chi_k}m_k}k^{\chi_R/\chi_k}dj + \bar{G} = \int (s_{eR} + s_{hR})e^{m_R - \frac{\chi_R}{\chi_k}m_k}k^{\chi_R/\chi_k}G^{-1}k^{\tau_G}dj$$
$$\iff (s_{eR} + s_{hR})e^{m_R - \frac{\chi_R}{\chi_k}m_k}\int k^{\chi_R/\chi_k}dj + \bar{G} = (s_{eR} + s_{hR})e^{m_R - \frac{\chi_R}{\chi_k}m_k}G^{-1}\int k^{\chi_R/\chi_k + \tau_G}dj$$
$$\iff (s_{eR} + s_{hR})e^{m_R - \frac{\chi_R}{\chi_k}m_k}e^{\left(\frac{\chi_R}{\chi_k}m_k + \left(\frac{\chi_R}{\chi_k}\right)^2\frac{\Sigma_k^2}{2}\right)} + \bar{G} = (s_{eR} + s_{hR})e^{m_R - \frac{\chi_R}{\chi_k}m_k}G^{-1}e^{\left(\left(\frac{\chi_R}{\chi_k} + \tau_G\right)m_k + \left(\frac{\chi_R}{\chi_k} + \tau_G\right)^2\frac{\Sigma_k^2}{2}\right)}$$

$$G = \frac{(s_{eR} + s_{hR})e^{m_R - \frac{\chi_R}{\chi_k}m_k}e^{\left(\left(\frac{\chi_R}{\chi_k} + \tau_G\right)m_k + \left(\frac{\chi_R}{\chi_k} + \tau_G\right)^2\frac{\Sigma_k^2}{2}\right)}}{(s_{eR} + s_{hR})e^{m_R - \frac{\chi_R}{\chi_k}m_k}e^{\left(\frac{\chi_R}{\chi_k}m_k + \left(\frac{\chi_R}{\chi_k}\right)^2\frac{\Sigma_k^2}{2}\right)} + \bar{G}} = \frac{(s_{eR} + s_{hR})e^{\tau_G m_k + \tau_G\left(2\frac{\chi_R}{\chi_k} + \tau_G\right)\frac{\Sigma_k^2}{2}}}{(s_{eR} + s_{hR}) + e^{-m_R - \left(\frac{\chi_R}{\chi_k}\right)^2\frac{\Sigma_k^2}{2}}\bar{G}}$$

where m_R in the last line is given by equation (77).

Finally, given that we target the ratio of research grants $\bar{G}$ over GDP $\bar{g} = \bar{G}/Y$, we get that G is given by

$$G = \frac{(s_{eR} + s_{hR})e^{\tau_G m_k + \tau_G\left(2\frac{\chi_R}{\chi_k} + \tau_G\right)\frac{\Sigma_k^2}{2}}}{(s_{eR} + s_{hR}) + e^{-m_R - \left(\frac{\chi_R}{\chi_k}\right)^2\frac{\Sigma_k^2}{2}}\bar{g}Y}$$
$$\text{where} \quad Y_t = w_t\sum_{a=1}^{10}\exp\left(m_{h,t,a} + \frac{1}{2}\Sigma_{h,t,a}^2\right)$$

Average income tax rate and intercept of income tax schedule If we target average income tax rate $\bar{a}_y$ then the following should be true

$$1 - \bar{a}_y = \frac{\int (1 - a_y)(wh_i)^{1-\tau_y}di}{\int wh_i di}$$
$$= \frac{(1 - a_y)(w)^{1-\tau_y}\sum_{a=1}^{10}\exp\left((1 - \tau_y)m_{h,t,a} + ((1 - \tau_y))^2\frac{\Sigma_{h,t,a,}^2}{2}\right)}{(w)\sum_{a=1}^{10}\exp\left(m_{h,t,a} + \frac{\Sigma_{h,t,a}^2}{2}\right)} \qquad (100)$$

Average tuition subsidy and intercept of tuition subsidy schedule If we target the average subsidy to higher education $\bar{a}_n$, then the following should be true

$$
\begin{aligned}
1+\bar{a}_n &= \frac{\int (1+a_n) e_i y_i^{-\tau_n} di}{\int e_i di} = \frac{\int (1+a_n) s y_i^{1-\tau_n} di}{\int s y_i di} = (1+a_n)(1-a_y)^{-\tau_n} \frac{\int \left(\tilde{w} h_i\right)^{(1-\tau_n)(1-\tau_y)}}{\int \left(\tilde{w} h_i\right)^{1-\tau_y}} \\
\iff 1+\bar{a}_n &= \frac{(1+a_n)}{(1-a_y)^{\tau_n}} \frac{\left(\tilde{w}\right)^{(1-\tau_y)(1-\tau_n)} \exp\left((1-\tau_y)(1-\tau_n) m_h + ((1-\tau_y)(1-\tau_n))^2 \frac{\Sigma_h^2}{2}\right)}{\left(\tilde{w}\right)^{1-\tau_y} \exp\left((1-\tau_y) m_h + ((1-\tau_y))^2 \frac{\Sigma_h^2}{2}\right)} \\
\iff 1+\bar{a}_n &= \frac{(1+a_n)}{(1-a_y)^{\tau_n}} \left(\tilde{w}\right)^{-(1-\tau_y)\tau_n} \exp\left(-(1-\tau_y)\tau_n m_h + \tau_n(\tau_n - 2)\left((1-\tau_y)\right)^2 \frac{\Sigma_h^2}{2}\right) \\
\iff a_n &= (1-a_y)^{\tau_n} \left(\tilde{w}\right)^{\tau_n(1-\tau_y)} \exp\left(\tau_n(1-\tau_y) m_h + \tau_n(2-\tau_n)((1-\tau_y))^2 \frac{\Sigma_h^2}{2}\right)(1+\bar{a}_n) - 1
\end{aligned}
\tag{101}
$$