University Research and the Market for Higher Education*

Titan Alon **Damien Capelle** **Kazushige Matsuda**

THIS DRAFT: September 2021

Abstract

American universities are an important source of domestic R&D, accounting for over 13% of aggregate R&D expenditure and more than half of spending on basic research. We develop a model which endogenizes university research expenditure and show how it depends on the market structure of the higher education sector. Spending on research improves the quality of education a university can offer by exposing its students to cutting edge ideas and techniques. Consequently, the competition between universities for tuition and talented students is an important determinant of R&D expenditure. The calibrated model successfully replicates key distributional characteristics of the U.S. higher education sector, including institution-level heterogeneity in revenue and research expenditure. We use the model to assess the impact on university R&D of implementing the Biden administration's proposed expansion of federal need-based student financial-aid. The model predicts a substantial 9.43% rise in university research spending in response to the policy. These gains materialize in the long-run, while simulations of the economy's transition path show that R&D declines in the short-run.

*We thank numerous seminar participants for useful comments which improved the paper. Alon: Department of Economics, University of California San Diego, 9500 Gilman Drive #0508, La Jolla, CA 92093 (email: talon[at]ucsd.edu), Capelle: International Monetary Fund, 700 19th St NW, Washington, DC 20431 (email: dcapelle[at]imf.org), Matsuda: Department of Economics, Kobe University, 1-1, Rokkodai-cho, Nada-ku, Kobe, Hyogo, 657-8501, Japan (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

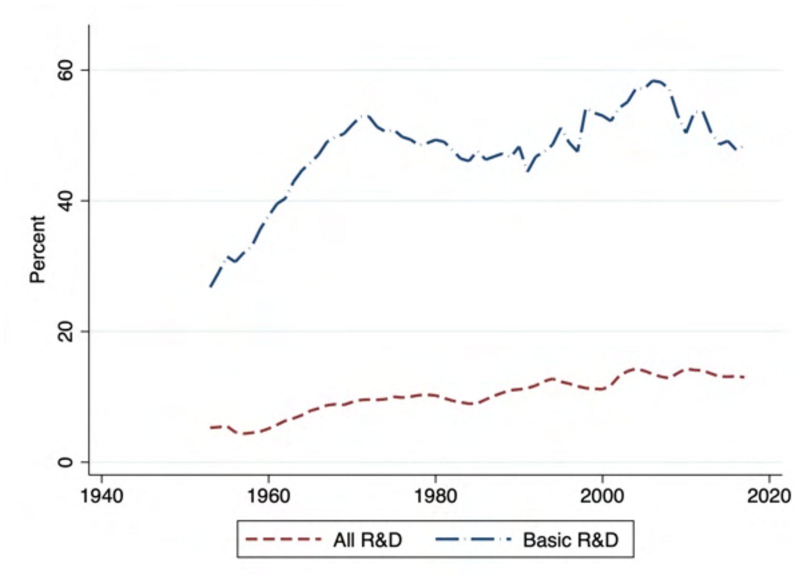
American universities are a major source of domestic technological innovation. They account for over 13% of aggregate spending on research and development (R&D) and over half of basic scientific research (see Figure 1). Discoveries stemming from university research underpin advances in many of the fastest growing sectors of the economy, such as communications, medicine, and aerospace (Rosenberg and Nelson 1994).

In this paper, we argue that competition between universities for tuition and talented students is an important driver of R&D expenditure in the higher education sector. In our model, spending on research is an investment which raises the quality of education a university can offer by exposing students to cutting edge ideas and techniques. Better quality institutions can charge higher tuition, win government grants, and attract more talented faculty and students who further augment education quality through peer-effects. Consequently, our framework does not rely on common mechanisms in the literature to rationalize university R&D expenditure, such as imperfect appropriability, patenting, or government funding. A novel implication of our framework is that policies which affect the demand for education, such as federal tuition assistance programs, can have large effects on university R&D.

Combining several sources of institution-level administrative micro-data, we document facts consistent with the idea that university research spending is influenced by the nature of competition it faces in the market for higher education. First, we show that patent licensing revenue is too small to be the primary driver of university research. Between 1991-2018, the median research university earned combined licensing fees totaling less than 1% of their expenditures on research. Second, we show that while the role of the federal government is large, roughly one-third of university research is paid for with internal funds, the majority of which come from tuition revenue. This *cross-subsidization* of research represents an internal incentive for universities to spend on R&D stemming from their educational activities. Finally, we show that university R&D expenditure has a strong institution-level correlation with other important determinants of education quality, including student ability, tuition, school rank, scientific output, and faculty pay.

To capture the interdependence of university teaching and research, we jointly model its instructional and R&D decisions in a general equilibrium model of the higher education sector. In the model, heterogeneous households imperfectly transmit human capital to their children and together they decide where to go to college subject to a

Figure 1: 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).*

credit constraint. A government provides need-based tuition subsidies and meritocratic research grants financed with taxes on income and consumption. Universities differ in the stock of knowledge they can impart to students. This knowledge, together with student peer-effects, teacher quality, and spending on instructional equipment will determine the quality of education a university offers. By purchasing research equipment and allocating faculty to R&D, a university can accumulate a greater stock of knowledge and improve its future education quality. Improvements in education quality bring in more tuition revenue and attract better students and faculty who further augment quality improvements through peer-effects. The model's equilibrium features a hierarchy of universities differing in education quality with two-dimensional student sorting by ability and family income. Tuition are market-clearing equilibrium prices that depend on college quality, government policy, and student characteristics.

Using the model, we assess the impact on university R&D of implementing the Biden administration's proposed reforms to federal student-aid programs. We focus in particular on the expansion of tuition assistance and the increase in need-based targeting of government funds. The calibrated model successfully matches key characteristics of the U.S. higher education system, including institution level heterogeneity in revenues and research expenditure. We model the policy as a one-time, unanticipated and permanent shift in the level and progressivity of government tuition-assistance and simulate the economy's transition to a new long-run equilibrium. The model predicts a rise in university R&D expenditure of 9.43%, a substantial increase in economic terms. The expansion is driven in almost equal parts by an increase in university revenues and a rise in the cross-subsidization rate of research. The economy's transition path indicates that these gains materialize in the long-run, while in the short-run the policy reforms trigger a decline in university research spending and a fall in the cross-subsidization rate.

Related Literature Our paper contributes to the literature investigating the economic incentives behind basic research. The traditional view is that the type of knowledge created by basic research cannot be fully appropriated by its developers due to imperfect intellectual property rights, and so government subsidies are required to facilitate basic R&D investment (Nelson 1959; Arrow 1962). Building on this view, much of the subsequent literature on basic research proceeded broadly along two tracks. The first focused on measuring the extent of spillovers from basic research, predominantly academic, and often found them to be substantial (Jaffe 1989; Adams 1990; Mansfield 1995). The second investigated how strengthening intellectual property rights impacts the provision of basic research, especially at universities. The major focus was on the impact of the Bayh-Dole of 1980 which gave academic institutions the right to patent and commercialize discoveries supported by government funding. Ultimately the findings were mixed, with many studies finding small or moderate roles for the act relative to longer-term trends in university R&D and patenting (Henderson, Jaffe, and Trajtenberg 1998; Jensen and Thursby 2001; Mowery et al. 2001; Colyvas et al. 2002; Sampat 2006; Lach and Schankerman 2008).

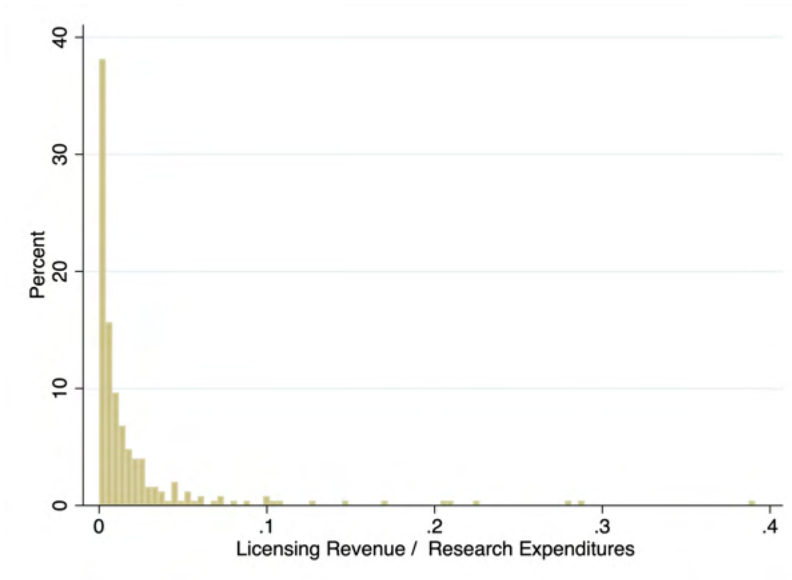
Most of this early literature was largely empirical. More recently, researchers have begun integrating basic R&D into models of long-run growth to study the extent of its under-provision due to imperfect property rights and the optimality of government subsidization policies (Gersbach, Schneider, and Schneller 2013; Cozzi and Galli 2014; Gersbach and Schneider 2015; Akcigit, Hanley, and Serrano-Velarde 2020). While all of these papers incorporate unique properties of basic research which differentiate it from applied R&D, none emphasize the special role of universities in producing it. Moreover, basic R&D in these models outside of private enterprises generally relies wholly on government funding. Consequently, while these models offer important insights into the efficient provision of basic research, they cannot explain why so much basic R&D

and government subsidies are taken up by universities, rather than the business sector, nor why universities spend a considerable amount of their own internal funds on basic research (see section 2). One exception is Aghion, Dewatripont, and Stein (2008), who provide a model rationalizing the university's out-sized role in basic research. In their framework, researchers pursuing basic science value creative control and academia has a comparative advantage over the business sector in guaranteeing them the freedom to pursue their own research agenda without interference from commercial considerations.

Similar to Aghion, Dewatripont, and Stein (2008), our paper develops a model that rationalizes the higher education sector's out-sized role in basic research, though our economic mechanism is completely different. We conceptualize university R&D as an input into the overall quality of education it offers in a competitive market for higher education. Building a model which link university research spending to the structure of education markets it operates in requires taking a stand on what is the objective function that universities are maximizing. To do so, we draw on popular models which endogenize the market structure of the higher education sector in general equilibrium (Epple and Romano 1998; Epple, Romano, and Sieg 2006; Hendricks, Herrington, and Schoellman 2021). In particular, our modelling approach is most similar to the competitive market framework of Cai and Heathcote (2020), who study the equilibrium effects of tuition policies on the dynamics of income inequality. Similar to them, we build a heterogeneous-agent model of the college market with endogenously differentiated colleges maximizing education quality. Our main modelling departure is to differentiate between different types of university expenditure, most importantly between teaching and research, but also between faculty and equipment expenditures. Critical to our purposes, these extensions endogenize university R&D as a dynamic investment decision. They also allow us to quantify the importance of our mechanism by structurally estimating key parameters from detailed university-level administrative microdata.

The remainder of the paper is organized as follows. Section 2 combines several sources of microdata to establish empirical facts consistent with our arguments. Section 3 formalizes our theory into a tractable general equilibrium model of the higher education sector. Section 4 derives closed form solutions for the model and presents analytical properties of the equilibrium. Section 5 presents the calibration strategy and analyzes the model's fit. Section 6 uses the calibrated model to quantify the consequences of recently proposed tuition policies on fundamental research and aggregate productivity. Section 7 concludes with directions for future research.

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.*

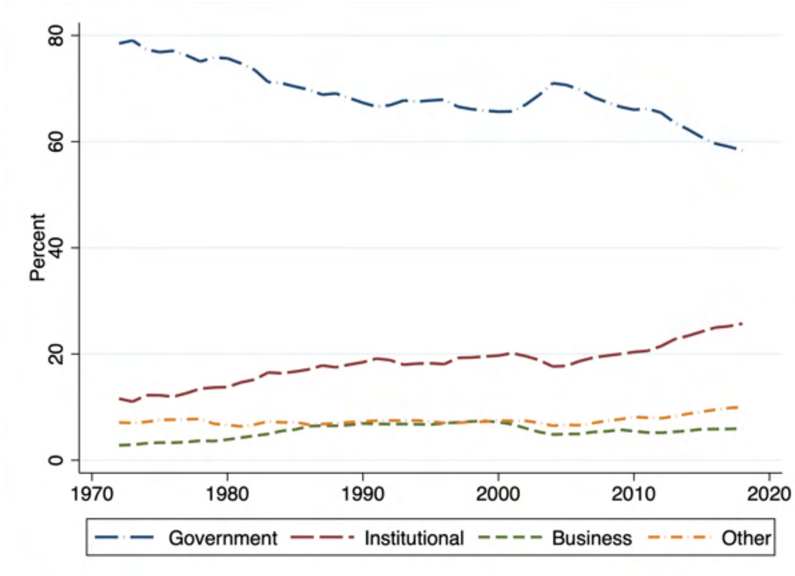
2 Research in the Higher Education Sector: Stylized Facts

In this section we document several important economic characteristics of the market for higher education and how they relate to university research expenditure. In part, these facts highlight important institution-level features of the data which are consistent with our hypothesis. These facts guide our modelling choices and we employ them in section 5 to discipline key parameters in the model's calibration.

Our sample includes all accredited 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 accredited domestic institutions. We merge in additional data from the National Science Foundation's Higher Education Research and Development Survey (HERD), 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.

The first important fact we document is that university revenue from patent licensing is

Figure 3: University research spending by source of funds



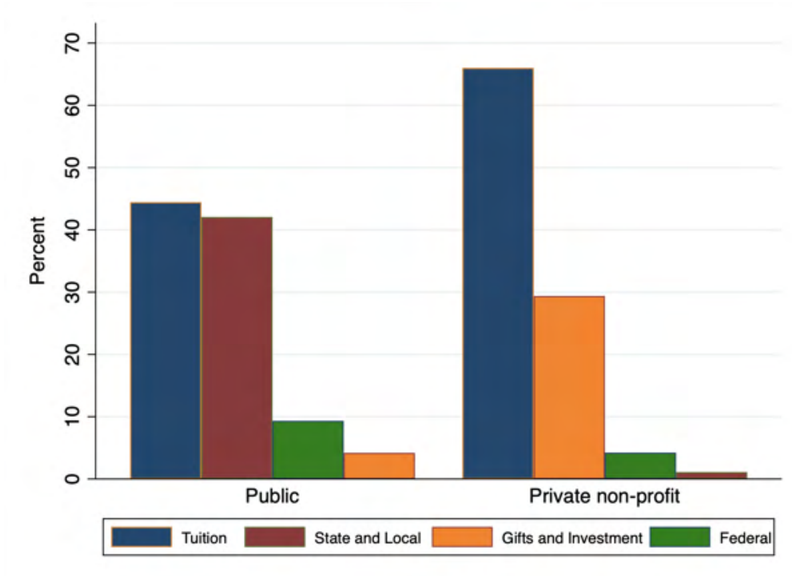
Notes: *Y-axis reports total university research spending by source of funds. Government sources includes federal, state, and local grants and contracts, though the federal represents the majority. Non-profit funding included in Other. All values converted to real 2012 dollars using the GDP deflator. Data from National Science Board (2018).*

a very small share of total revenue and small even relative to its expenditure on research. Figure 2 shows 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 1% of their expenditure on research. The size of these income streams means that patenting cannot be a major motivation for university research. This conclusion is reinforced by the fact that much of the rise in university research predates the recent increase in patenting.[1] The data indicate that higher education research is not well described by models of patent-driven innovation.

The second important fact is that a large share of research is conducted using internal funds, indicative of a substantial degree of cross-subsidization of university research. It is widely known that the government, primarily federal, is the largest source of direct funds for university research, but this is only part of the story. As Figure 3 shows, almost one-third of university research is funded using internal institutional funds, and that share has been consistently rising for the nearly five decades for which we have data. These trends are largely similar at private and public institutions, and are driven

[1] While these observations are true for academic research in general, there are sub-fields of research in which industry partnerships and patenting appear to be important, such as pharmaceuticals. See Sampat (2006) for additional evidence and discussion on the minor role played by patenting in academic research.

Figure 4: University revenue sources, by sector



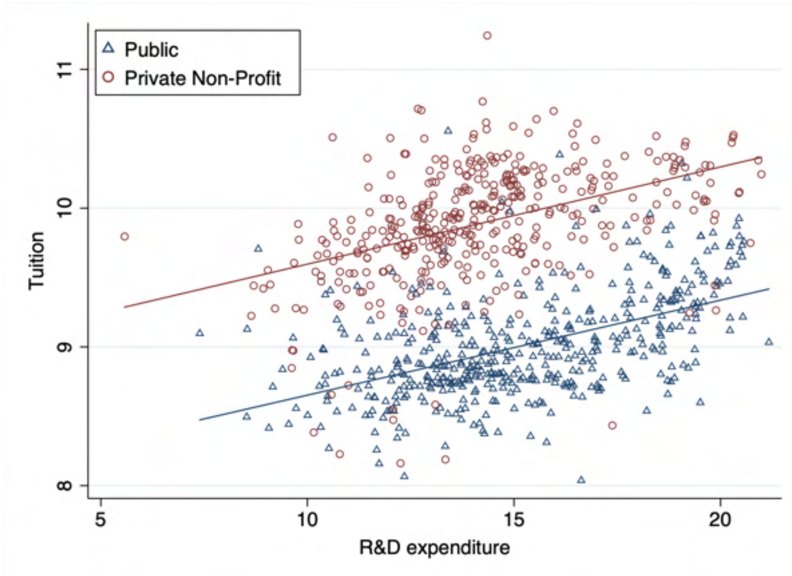
Notes: *Underlying data from the NCES's Integrated Postsecondary education Data System (IPEDS). Y-axis reports share of total university revenue. University revenue includes tuition; federal, state, and local appropriations, grants, and contracts; affiliated entities, private gifts, grants, and contracts;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

by both slower growth in government support and accelerating growth in institutional spending (see Figure A1 in the Appendix). Therefore, while the literature has correctly emphasized the key role played by government grants, we show that there still exists large private incentives which remain to be explained.

To identify what incentives may be important, we examine the major sources of income which are used to cross subsidize university research. Figure 4 shows that the most important of these is tuition revenue, which composes the lion's share of university income. This is true for both public and private institutions and may even understate its importance given that state appropriations and private gifts are often important indirect forms of tuition. The cross subsidization of research with tuition is also evident in the cross-sectional data, which shows that tuition is highly unequal across institutions and exhibits a strong positive correlation with university research (see Figure 5).[2] A key implication of our framework is that high levels of cross subsidization will lead competition between universities for tuition and talented students to influence the provision

[2]The size of the correlation is economically meaningful: moving from the 10th to the 90th percentile of research expenditures is associated with a 60% percent increase in tuition revenue per-student.

Figure 5: University research expenditure and tuition, by sector

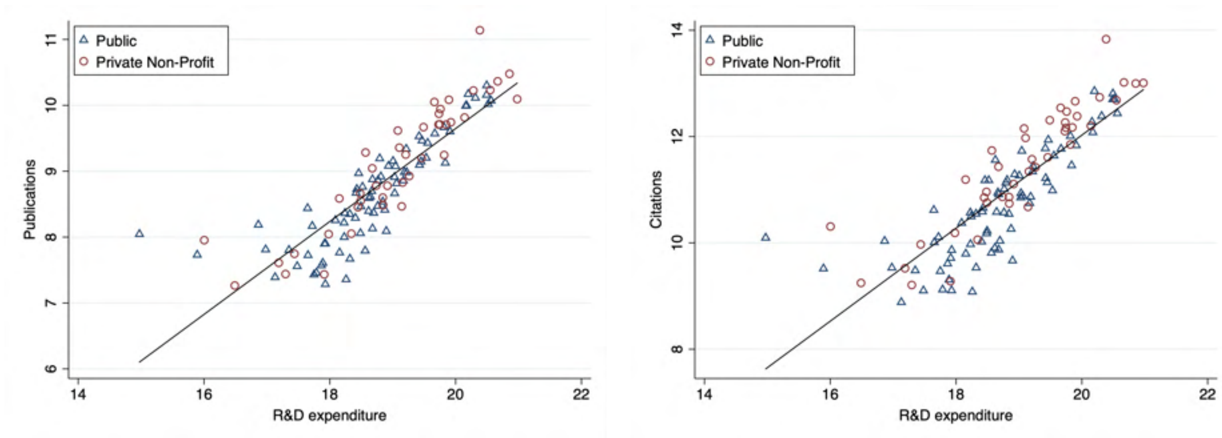


Notes: *Tuition is the average tuition revenue the university receives per full-time equivalent student, net of any university discounts or allowances. research expenditure is total university spending for activities specifically organized to produce research outcomes, including by institutes, research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

of university research. For instance, when student ability and tuition correlate strongly with school quality, fierce competition can drive research spending as universities seek to pull ahead of peer institutions. Conversely, if students were homogeneous or tuition fees were similar across universities, there would be little incentive to invest in research.

In the data, we find strong evidence in favor of an environment similar to the first example: both tuition and student ability rise sharply with an institution's rank. The knowledge that a university possesses and can impart to prospective students constitutes an important part of its rank and the quality of education it offers. That university research spending contributes to the accumulation of this knowledge is widely accepted. It is also apparent in the fact that the universities which spend the most on research regularly top national and international college rankings of the "best colleges" (see Table B1) and in the strong correlations between research expenditures and the number of publications and citations a university generates (see Figure 6). Consistent with the idea that student's perceive these benefits and that universities internalize them when making decisions, we document a strong association between university research spending and the hierarchical sorting and stratification which characterizes the market structure

Figure 6: University research spending and knowledge creation



(a) Publications (b) Citations

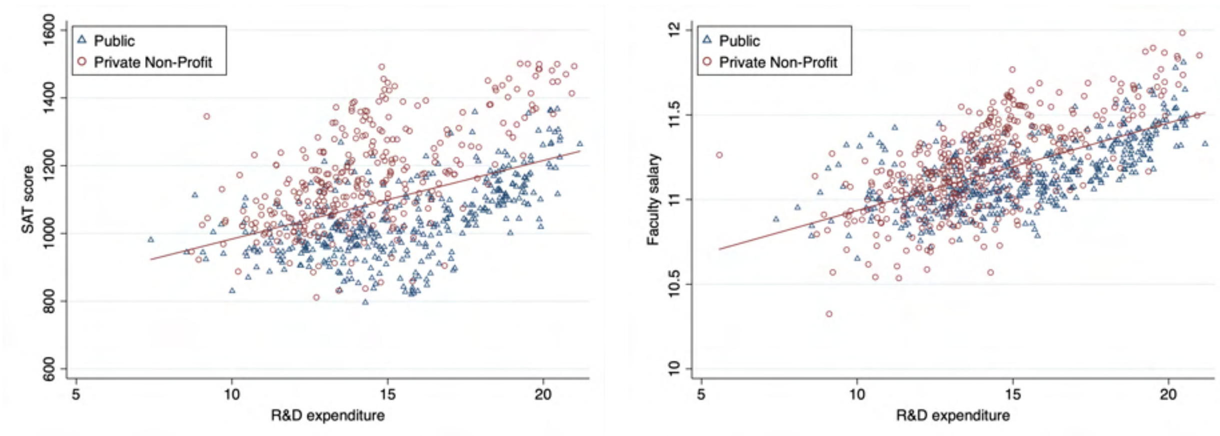
Note: *Publication and citation data come from the CWTS Leiden Ranking derived from the core collection of the Web of Science (WoS) for the years 2015-2018. research expenditure is total university spending for activities specifically organized to produce research outcomes, including by institutes, research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

of higher education in the United States (Hoxby 2009; MacLeod and Urquiola 2015). Figure 7 summarizes the pattern and shows that high ability students (measured by SAT scores) and high quality professors (measured by salaries) sort assortatively into universities with higher research expenditures.

A final important question is the extent to which government funds, which still constitute the bulk of direct funding, lean against or reinforce the competitive motives suggested by the data. Using the National Science Foundation's HERD Survey, we can compare how the amount of external grants and contracts a university is awarded relates to its own internal spending on research. Figure 8 plots the relationship and shows a significant positive, nearly perfectly log-linear relationship between the two. The data suggests that government grants and contracts act primarily as a subsidy, augmenting a university's own internal spending.[3] This likely occurs in part by design and in part as a consequence of the highly competitive and meritocratic process which determines the allocation of federal funds through agencies like the NSF and NIH. An important consequence is that government funding supports a winner-takes-all dynamic by concentrating resources at the universities which already spend the most. Government funding

[3]The relationship remains virtually unchanged if we include *all* grants and contracts, such as those awarded by the business sector or non-profit institutions, rather than just those from the government.

Figure 7: University research spending, student ability, and teacher quality



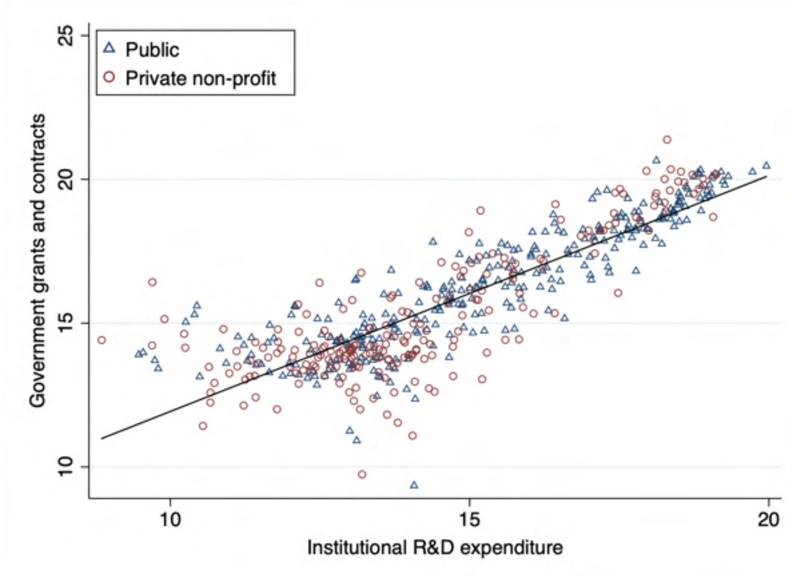
(a) Student Ability (b) Teacher Quality

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.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.*

will therefore reinforce the competitive channels we investigate.

Taken together, the facts above are consistent with the view that university research is incentivized by competition for tuition and talented students in the market for higher education. Universities spend on research to raise the quality of their education and differentiate themselves from competing institutions. Prospective students are willing to pay higher tuition for better quality programs, especially high ability students who experience greater learning complementarities from the new knowledge created through research. The resulting cross subsidization creates an interdependence between a university's research and teaching activities which implies that any study of university research should take into account the structure of education markets it operates in. The facts above provide guidance for this purpose. In the next section, we formalize these channels into a general equilibrium model of the higher education sector which allows us to account for the simultaneity of causality in determining the importance of each economic force in shaping the equilibrium outcomes we observe in the market.

Figure 8: 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 level averages and are plotted in logs. Data on institutional research is from the Higher education Research and Development Survey (HERD). Data on grants and contracts comes from the Integrated Postsecondary education Data System (IPEDS).*

3 A Model of Fundamental Research in Higher Education

The economy is populated by heterogeneous households and colleges. In each period, parents imperfectly transmit human capital to their children and together they decide which college the child will attend. Colleges choose the pool of students to admit, which faculty to hire, and how to allocate equipment and faculty between teaching and research activities. The model's equilibrium features an endogenous hierarchy of universities which differ in their education quality and two-dimensional student sorting by ability and family income. Tuition are market-clearing prices that depend on college quality, government policy, and student characteristics.

3.1 Households

The economy is populated by a unit mass of dynasties, $i \in [0, 1]$. Each individual lives for two periods: one as a child, one as an adult. Each adult begets one child. A generation-t 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}, labor supply ℓ_{it}, and a college quality q_{it} for their child that will determine, along with the labor market shock, 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 " ′ ". The household decision problem can be formulated recursively as

$$\mathcal{U}(h, z) = \max_{c,\ell,q} \left\{ (1-\beta) \left[\ln c - \ell^{\eta} \right] + \beta \mathbb{E} \left[\mathcal{U}(h', z') \right] \right\} \tag{1}$$

where β denotes the intergenerational discount factor.

A child's human capital at the end of high school is modeled as a log-linear combination of parents' human capital h and a birth shock ξ_z capturing the randomness of the intergenerational transmission process

$$z = (\xi_z h)^{\alpha_h} \qquad \text{Child's High School Ability} \tag{2}$$

A household's before-tax labor earnings is the product of the wage per unit of effective labor w, their level of human capital h, and their labor supply ℓ. Household income is subject to a progressive income tax with rate of progressivity τ_y and schedule intercept $(1 - a_y)$, as in Heathcote, Storesletten, and Violante (2017). After tax-and-transfers labor earnings, denoted by y, is therefore given by

$$y = (1 - a_y) \left(w h \ell \right)^{1-\tau_y} . \tag{3}$$

Households are subject to a lifetime budget constraint. Their income y can be spent on consumption and tuition. The tuition schedule is an equilibrium object that depends on college quality q and a child's ability z. Financial aid is need-based, and if $p(q, z)$ denotes the before-aid tuition fee, the out-of-pocket payment by a household with income y is given by

$$\frac{y^{\tau_n}}{1 + a_n} p(q, z) \tag{4}$$

where τ_n is the rate of progressivity of the need-based subsidy and $1 + a_n$ is the intercept of the schedule.[4] Normalizing the price of the final good to one, the household's budget

[4]Government need-based financial aid programs are therefore modelled using the same two-parameter structure as the income tax schedules. See Capelle (2020) for a discussion.

constraint is given by

$$y = (1 + \bar{a}_c)c + \frac{y^{\tau_n}}{1 + a_n} p(q, z) \tag{5}$$

where $\bar{a}_c$ is the consumption tax rate. The budget constraint emphasizes the importance of borrowing constraints in shaping household college choices. In equilibrium, these constraints will lead to student sorting based on family income. Lochner and Monge-Naranjo (2012) provide extensive evidence of the importance of borrowing constraints in the college market and its contribution to student sorting by family income.

The child's human capital after graduating college is a log-linear combination of their pre-college ability, the quality of the college they attended, and a labor market shock. It is given by

$$h' = z q^{\alpha_q} \xi_y \qquad \text{Child's Post-College Human Capital} \tag{6}$$

where α_q parameterizes the earnings elasticity with respect to college quality. There are two sources of randomness in the accumulation process of human capital. The birth shock ξ_z is known before the college quality decision has to be made, while the labor market shock ξ_y is realized once the child enters the labor market. It is assumed that the birth and labor market shocks are i.i.d across generations and households and log-normally distributed.

$$\ln \xi_z \sim \text{i.i.d.}\mathcal{N}\left(-\sigma_z^2/2, \sigma_z^2\right) \qquad \text{Birth Shock} \tag{7}$$

$$\ln \xi_y \sim \text{i.i.d.}\mathcal{N}\left(-\sigma_y^2/2, \sigma_y^2\right) \qquad \text{Labor Market Shock} \tag{8}$$

3.2 Universities

The primary activity of colleges is to educate students. The quality of education a school offers depends on four inputs: (i) the human capital of its teaching faculty $\bar{h}_q$, (ii) its teaching equipment e_q, (iii) its knowledge capital k, and (iv) a student peer effects $\zeta(\phi; p)$. Formally, a college's education quality Q is given by

$$q = \zeta(\phi; p) \times \bar{h}_q^{\omega_h} e_q^{\omega_e} k^{\omega_k} \tag{9}$$

The contribution of faculty time to education quality depends on the average human capital of its teaching faculty,

$$\bar{h}_q = \mathbb{E}_{\mu_q(.)}\,[\,h\,] \tag{10}$$

where $\mu_q(\cdot)$ is the endogenous distribution of teacher ability chosen by the university. The presence of k in the quality function captures the notion that teaching is more effective not only when the faculty is more *skilled*, but also when they are more *knowledgeable*. Hence, as in Akcigit et al. (2018), knowledge and human capital are two different types of productive capital.[5]

Student Peer-Effects. The final input into a college's education quality is the peer-effect generated by it's student body.[6] We model the peer-effects within a university using the following functional form,

$$\ln \zeta(\phi; p) = \omega_z \, \mathbb{E}_{\phi(.)}[\ln(z)] - \sigma_u^2(\phi; p) \tag{11}$$

The first term captures the direct 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. It implies that student heterogeneity has a negative impact on quality relative to an arithmetic average. The parameter ω_z parameterizes the importance of these peer-effects for overall education quality.

A second assumption we make is the presence of a direct cost of heterogeneity: $\sigma_u(\phi; p)$. This term 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 $\sigma_u^2(\phi; p) = \frac{\Omega_t}{2} V_{\phi(.)}(\ln p(q, z))$ where Ω_t is an aggregate constant defined in appendix (57). Defining σ_u^2 in this manner ensures tractability.

Academic Research and Knowledge Capital The second major activity undertaken by universities is research. In our model, universities spend on research to improve their

[5]In theirs, knowledge of inventors is an input in the production of ideas. Ideas are then sold to producers who use them to increase their TFP. In our model, knowledge is used to increase human capital and for research (see below). Human capital is in turn an input to produce goods and services.

[6]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).

education quality by accumulating new knowledge that can be imparted to students.[7] We posit a university research technology which builds on its existing knowledge using research equipment, e_k, and the human capital of its research faculty, $\bar{h}_k$. Formally, the knowledge at college j accumulates over time according to the following law of motion:

$$k' = k^{\gamma_k} e_k^{\gamma_e} \bar{h}_k^{\gamma_h} \tag{12}$$

where $\gamma_k < 1$ captures the idea that the university specific educational value of knowledge discovered through research depreciates over time.[8] As with teaching faculty, the contribution of research faculty depends on their average human capital

$$\bar{h}_k = \mathbb{E}_{\mu_k(.)}\left[\, h \,\right] \tag{13}$$

where $\mu_k(\cdot)$ is the distribution of research faculty ability chosen by the university. Allowing the university to choose different compositions of research faculty, $\mu_k(\cdot)$, and teaching faculty, $\mu_q(\cdot)$, captures the notion that universities can partially specialize these tasks internally by hiring dedicated teaching faculty or increasing teaching loads for research faculty.

Government Research Grants In reality, government subsidization of basic research occurs through a host of meritocratic programs managed by various government departments, including the National Institute of Health, the Department of Defense, NASA, the National Science Foundation, and others (see National Science Board (2018), *Expenditures and Funding for Academic R&D*). We model the net effect of these programs parsimoniously through a reduced-form allocation rule for government grants that captures both the level of subsidy and its meritocratic distribution. Specifically, we assume that government grants cover a fraction $1 - G(k)$ of a universities research (but not teaching) expenditures. We parameterize the grant function with the two-parameter family

$$G(k) = \bar{G} k^{-\tau_G} \tag{14}$$

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

[8]The technology can be extended to allow for the slow diffusion of research discoveries across colleges. Higher levels of diffusion will generally reduce the university's private incentive to invest in R&D. We omit detailing this process due to the availability of good data which could discipline the mechanism. Instead, we capture these effects through the more general depreciation factor γ_k.

where $\bar{G}$ and τ_G capture the average subsidy and its distribution across universities, respectively. This approach allows us to account for the role of government subsidies on university research when we attempt to quantify the importance of the new channels we introduce. This modelling of government research grants also highlights another important distinction between our work and much of the literature: in most models of basic research in the literature, research is undertaken *only* because of government funding, whereas in our model universities have an independent private incentive to engage in basic research, with government subsidies augmenting them.

The University's Problem. Following the literature, we assume colleges value the quality of education they deliver to their students (e.g. Epple, Romano, and Sieg 2006). The presence of a research investment decision makes the university problem dynamic. Letting the instantaneous flow payoff for a college delivering education of quality q be $\ln q$, we can formulate the university problem recursively as

$$V(k) = \max_{\phi, \mu_q, \mu_k, e_q, e_k} \ln q + \beta V(k') \tag{15}$$

where we assume for simplicity that the discount factor of colleges is the same as that of households.

The timing of events is as follows. First, colleges choose the composition of the students' body $\phi_{jt}(z, y)$—a density over (z, y) given a tuition schedule $p_{jt}(q, z)$. Simultaneously students apply given the tuition schedule and an expectation of a quality q. Once they have been admitted and paid their tuition fees, students commit to staying. Second, colleges choose the composition of its instructional and research faculty $\mu_{qjt}(h), \mu_{Ijt}(h)$ and the amount of teaching and research equipment e_{qjt}, e_{Ijt} to purchase, given their tuition and grants revenues and the average ability of students. Colleges make these choices to maximize (15) subject to the education technology in (9), the peer-effect (11), the research technology, (12), and a flow budget constraint,

$$G(k_j)\left(e_{jI} + w\bar{h}_I\right) + e_{jq} + w\bar{h}_q = \mathbb{E}_{\phi_j(.)}[p(q, z)]. \tag{16}$$

which includes the contribution of government subsidies.

3.3 Final Goods and Spillovers of Fundamental Research.

Much of the literature on academic research emphasizes the productivity spillovers it generates for the business sector. Several well received empirical studies have found these effects to be quantitatively large (Jaffe 1989; Adams 1990; Mansfield 1995). To account for these effects, we introduce a goods-producing business sector subject to productivity spillovers from the knowledge discovered in academic research.

Formally, we assume there is a competitive final goods producing sector whose output is used for consumption by households or equipment for universities. Firms operate a constant returns to scale 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 in the higher education sector, so that $A = \bar{A}K^{\iota_k}$ where $K = \mathbb{E}[k_j]$. The elasticity ι_k parameterizes the strength of spillovers from academic research. The profit-maximizing representative firm solves

$$\max_{H_F} \; F(H_F) - wH_F. \tag{17}$$

Government The government implements four kind of taxes: two are specific to higher education—merit-based research grants and need-based financial aid to college students—and two that are more standard—a linear consumption tax and a progressive income tax. The government balances its budget every period with the consumption tax.[9]

3.4 Market Clearing and Equilibrium

Labor Market Universities hire faculty from the pool of workers. Each worker can either work in the goods producing sector or in the higher education sector as either a teacher or a researcher. Labor market clearing requires,

$$H_F + \int h\, \mu_q(h|k)g(k)dhdk + \int h\, \mu_k(h|k)g(k)dhdk = \int h\, f(h,z)dhdz$$

where $f(h,z)$ is the endogenous distribution of household human capital and the left side terms capture aggregate effective labor in production, teaching, and research activities, respectively. Optimal household occupation choice dictates each choose the sector offering the highest income, leading the effective wage to be equated across sectors.

[9]See Appendix D.1 for the full government budget constraint.

Final Goods Market Goods market clearing requires that the total flow of final goods produced be either consumed by households or used as teaching or research equipment in the higher education:

$$F(H_F) = \int_i c_i \, di + \int (e_{kj} + e_{qj}) \, dj$$

Higher Education Market Since college quality is endogenous, market clearing in the higher education sector requires college admission decisions to be consistent with choices of income students.Formally, for any $Q \subset \mathcal{R}^+$,

$$\int \mathbb{1} \left[q(h, z) \in Q \right] f(h, z) dh dz = \int \mathbb{1} \left[q(k) \in Q \right] g(k) dk$$

where $f(h, z)$ and $g(k)$ are the endogenous distributions of household human capital and college quality, and where $q(h, z)$ and $q(k)$ are the optimal policy functions for households and colleges.

Equilibrium Definition. An equilibrium path is a sequence of tuition schedules $\{p_t(q, z)\}_{t=0}^{\infty}$, household's policy functions $\{c_t(h, z), \ell_t(h, z), q_t(h, z)\}_{t=0}^{\infty}$, colleges' policy functions $\{\phi_t(k, z), \mu_{It}(k, h), \mu_{qt}(k, h), e_{It}(k), e_{qt}(k), q_t(k)\}_{t=0}^{\infty}$, and distributions of human capital and scientific knowledge across colleges $\{f_t(h, z), g_t(k)\}_{t=0}^{\infty}$ such that (i) the household's policy functions $c_t(h, z), \ell_t(h, z), q_t(h, z)$ are a solution to (1), (ii) the college's policy functions $\mu_{qt}(k, h), \mu_{It}(k, h), \phi_t(k, z), e_{It}(k), e_{qt}(k)$ are a solution to (15), (iii) education markets, labor markets, and goods markets clear, (iv) the evolution of the distribution of human capital, $f_t(h, z)$, is consistent with the intergenerational law of motion of human capital and the sorting rule, $q_t(h, z)$, (v) the evolution of the distribution of knowledge capital, $g_t(k)$, is consistent with the law of motion of knowledge and research activities.

4 Equilibrium Properties of the Market for Higher Education

In this section we highlight key properties for understanding our mechanism and calibration strategy.

4.1 Equilibrium Market Structure in the Higher Education Sector

The equilibrium market structure of the higher education sector is summarized by a tuition schedule $p_t(q, z)$ and sorting rule $q_t(e, z)$. The sorting rule summarizes the quality of college attended by a student of ability z whose parents spent e on schooling (including government financial-aid) . The tuition schedule is the price charged to a student of ability z at a college of quality q, before any financial aid. Both the tuition schedule and the sorting rule are determined in equilibrium to balance supply and demand in the higher education sector.

Proposition 1. *The equilibrium before-financial-aid tuition schedule is given by*

$$p_t(q, z) = \underline{p}_t q^{\frac{1}{\epsilon_{1t}}} z^{-\frac{\epsilon_{2t}}{\epsilon_{1t}}} \tag{18}$$

and the equilibrium student sorting across colleges is given by,

$$q_t(e, z) = \left(\frac{e}{\underline{p}_t}\right)^{\epsilon_{1t}} z^{\epsilon_{2t}} \tag{19}$$

where $\underline{p}_t$, $\epsilon_{1t}, \epsilon_{2t}$ are non-negative time-varying aggregates defined in Appendix D.

Proposition 1 shows that in equilibrium, the sorting rule and the tuition schedule are characterized by three endogenous parameters common to all households: $\underline{p}_t$, ϵ_{1t}, ϵ_{2t}. The parameter $\underline{p}_t$ can be thought of as the endogenous average price level for higher education. The elasticity ϵ_{1t} captures the strength of student sorting based on family background (since expenditure e will be proportional to parental income, see Proposition 2). The income sorting arises because wealthier households are willing to spend more on education and colleges are willing to trade-off financial resources and student ability. The elasticity ϵ_{2t} captures the strength of sorting on student ability, which is driven by colleges desire to attract high ability students because of the peer-effects they generate.

The result in Proposition 1 illustrates the interdependence of university research and the market structure of the higher education sector. Consider the ratio $\epsilon_{1t}/\epsilon_{2t}$ which captures the extent to which the college market is stratified by family income relative to student ability. Larger values represent stronger sorting on income relative to ability, leading high quality colleges to have more wealthy students and fewer talented ones. Using the expressions for ϵ_{1t} and ϵ_{2t}, we have that

$$\frac{\epsilon_{1t}}{\epsilon_{2t}} = \frac{\omega_e + \omega_h + \beta\left(\gamma_e + \gamma_h\right) v_{t+1}}{\omega_z} \tag{20}$$

where v_t is the marginal value of research for a college.[10]. The variable v_t measures the strength of the university's incentive to invest in research. The expression in equation (20) shows that the ratio rises when the returns to research increase, either through an anticipated rise in the value of research (v_{t+1}) or through increases in the productivity of research through lab equipment (γ_e) or research faculty (γ_h). This is because universities favor wealthy student – who bring financial resources that support increases in research expenditures – over talented ones. As a result, a rise in the returns to university research increases stratification by family income and reduces the positive assortative matching of high ability students and high quality colleges. A similar intuition holds for the other inputs into the university's education quality. A rise in the productivity of inputs that require money – such as instructional equipment (ω_e) or teaching faculty (ω_h) – will increase sorting on family income, while increases in the strength of peer-effects (ω_z) will increase sorting on student ability.

4.2 Household Education Expenditure

The following proposition characterizes household spending on higher education.

Proposition 2. *In equilibrium, all households spend the same share of income y on education, $e_t = s_t \cdot y_t$, where s_t given by*

$$s_t = \frac{\beta \alpha_q \epsilon_{1t} u_{t+1}}{1 - \beta + \beta \alpha_q \epsilon_{1t} u_{t+1}} \tag{21}$$

$$\textit{where} \quad u_t = (1-\beta) \sum_{k=0}^{\infty} \beta^k \prod_{m=0}^{k-1} \rho_{t+m}$$

$$\textit{and} \quad \rho_t = \alpha_h + \alpha_q \left[\epsilon_{2t} \alpha_h + \epsilon_{1t}(1-\tau_n)(1-\tau_y)\right]$$

The education expenditure share in (21) admits a natural intuition. It increases with u_{t+1}, which measures the marginal utility to households of increasing the child's human

[10] More formally it is the marginal increase in the university value function $V_t(k)$ in (15) from a 1% increase in its k_t. We show in Appendix D that there exists at all points in time two aggregate variables $\{\bar{v}_t, v_t\}$ such that the university value function (15) can be written as $V_t\left(k_i\right) = \bar{v}_t + v_t \ln k_i$.

capital by 1%. The value of u_t depends on the future path of ρ_t, the intergenerational elasticity of income (IGE), which captures the persistence of economic status across generations. Spending on education also increases when households are more altruistic toward their children (β) or when attending high quality colleges deliver larger gains in human capital (α_q). Education expenditure increases when there is greater student sorting on financial background (ϵ_{1t}); both directly, by increasing the marginal college quality improvement that can be purchased through increased spending, and indirectly through the IGE, since these gains can be propagated to one's offspring. Improvements in the sorting of high ability students into high quality colleges (ϵ_{2t}) increase education spending by improving the overall quality of education offered in the higher education sector through peer-effects.

4.3 Universities and the Cross-Subsidization of Research

An important feature of our model is that universities have a private incentive to spend on research even in the absence of government subsidies (or patents). In the model, we can summarize the strength of this incentive by the rate of cross-subsidization–the share of tuition revenue that a university diverts to pay for research activities. Proposition 3 provides a first characterization of the cross-subsidization rate, s_R.

Proposition 3. *The share of tuition revenue institutions of higher education spend on research (equipment and faculty wages) is given by*

$$s_{Rt} = \frac{\beta(\gamma_e + \gamma_h)v_{t+1}}{\omega_e + \omega_h + \beta(\gamma_e + \gamma_h)v_{t+1}} \tag{22}$$

where the ratio of spending on research faculty to lab equipment is γ_h/γ_e.

The cross-subsidization rate falls when instructional equipment or faculty become more important for education quality (e.g. a rise in ω_e and ω_h). The rate rises when research faculty or lab equipment become better at producing research output (e.g. a rise in γ_e or γ_h, or when the equilibrium marginal value of academic capital, v_t, increases. The latter depends, in a complicated way, on equilibrium characteristics of the higher education market, including the variances of college quality, household income, and student ability. Before providing an expression for v_t, we present a simpler version of our model which admits an analytical solution.

A Simplified Model without Peer-Effects Consider a simplification of the model in section 3 which omits the peer-effect by setting $\omega_z = 0$. In this case, sorting in the higher education occurs only by family income. To further simplify the notation and focus on the salient characteristics of the model, we also eliminate government taxes ($\tau_y = \bar{a}_y = 0$) and research subsidies ($\tau_G = \bar{G} = 0$); research and teaching equipment ($\omega_e = \gamma_e = 0$); the variance of labor market shocks ($\sigma_y = 0$); and set the elasticities of the earnings and research technologies to unity ($\alpha_q = \gamma_h = 1$). The following proposition characterizes the simplified model's steady state.

Proposition 4. *In the steady state, the share of tuition revenue spent on research is given by,*

$$s_R = \frac{\beta\, v}{\omega_h + \beta\, v} \tag{23}$$

where v, the marginal value of scientific capital to the university, is given by

$$v = \frac{\omega_h \frac{(1-\tau_n)\Sigma_h}{\Sigma_k} + \omega_k}{\left(1 - \beta\left(\gamma_h \frac{(1-\tau_n)\Sigma_h}{\Sigma_k} + \gamma_k\right)\right)}. \tag{24}$$

Households spend a constant share of income on education given by

$$s = \frac{\beta \alpha_q (1-\tau_n)\, \epsilon_1}{1 - \beta \alpha_h}. \tag{25}$$

In the absence of peer-effects, $\epsilon_2 = 0$, and ϵ_1 is given by

$$\epsilon_1 = \omega_h + \omega_k \frac{\Sigma_k}{(1-\tau_n)\Sigma_h}. \tag{26}$$

The Σ_k and Σ_h are the standard deviations of the (log) state variables k and h, and are given by,

$$\Sigma_k = \frac{\gamma_h}{1-\gamma_k}(1-\tau_n)\Sigma_h \tag{27}$$

$$\Sigma_h^2 = \frac{\alpha_h^2 \sigma_z^2}{1 - \left(\alpha_h + \alpha_q (1-\tau_n)\left(\omega_h + \omega_k \frac{1}{1-\gamma_k}\right)\right)^2}. \tag{28}$$

The proposition highlights a key property of the model: the equilibrium interdependence between the household and university optimal choices, and the *dispersion* in the model's state variables Σ_k and Σ_h. The share of their tuition universities assign to research given in equation (23) can be thought of as a best-response function; it determines the university's optimal spending on R&D as a function of the degree of competition, measured by the distance between universities Σ_k, and the characteristics of the demand for higher education, measured by the dispersion in household incomes Σ_h. An exogenous decrease in Σ_k would lead to an increase in university spending on R&D because universities would become more similar to their competing institutions and therefore have a stronger incentive to invest in research to catch up to leading institutions. Similarly, a rise in Σ_h raises university R&D spending by increasing the dispersion in households' willingness to pay for quality education, thereby increasing demand for the highest quality schools.

The simplified model provides some intuition for our quantitative exercise which assesses the impact of increases in the level and progressivity of financial aid programs, $\{a_n, \tau_n\}$, on university research expenditure. Let's consider in isolation the increase in progressivity, τ_n.[11] The policy change impacts research expenditures first of all by decreasing the average household spending on education, as evident in equation 25. This is because the policy redistributes from high to low-income households which discourages the accumulation of human capital. For a given cross-subsidization rate, this leads to a decrease in the average spending on research. In the policy we simulate later, the increase in the average education subsidy $\bar{a}_n$ will however counteract this decrease in private spending.

The increase in progressivity, τ_n, impacts research expenditures not only through the decline in the average household spending on education but also by changing the cross-subsidization rate, s_R. The immediate impact of the policy is to reduce the incentives to invest in research, v, and a decrease in the rate of cross-subsidization, s_R – as is evident from equation 23 when holding Σ_h and Σ_k constant. This is because the policy decreases the dispersion in the households' willingness to pay for quality education, thereby decreasing the demand for high-quality colleges relative to low-quality ones.

In the long run however, the dispersion in family income Σ_h and academic capital Σ_k also change. Because gaps in education outcomes by family income shrink, so does the resulting dispersion in human capital, Σ_h – as is evident from equation 28. This further decreases the demand for the highest quality of education, which further reduces the incentives to do research. But this decrease in the dispersion of tuition revenues across

[11] In both the simple model and the full model, a change in the average rate of tuition subsidies a_n lead to proportional changes in university R&D expenditure without effecting the research share s_R.

colleges also compresses the dispersion of research spending. Ultimately this decreases Σ_k as is evident from equation 27. By reducing the distance across competing institutions, it gives stronger incentives to catch up to leading institutions, which increases the incentives to do research.

In the simplified model, the decrease in the dispersion in academic capital, Σ_k, exactly offsets the decrease in the dispersion in household's demand for quality in the long-run, $(1-\tau_n)\Sigma_h$. This is directly evident in equation (27) which shows that $(1-\tau_n)\Sigma_h/\Sigma_k$ always equals $\frac{1-\gamma_k}{\gamma_h}$ in the stationary equilibrium. In the full model with peer-effects, this will not be the case. This is because student talent remains scarce relative to monetary inputs, and its supply cannot readily change as market conditions change with the policy. To see this, consider the steady-state expression for the marginal value of research (v) in the full model:

$$v = \frac{(\omega_e + \omega_h)\frac{(1-\tau_n)(1-\tau_y)\Sigma_h}{\sqrt{1-s_z}\Sigma_k} + \omega_z \alpha_h \frac{\Sigma_h}{\sqrt{1-s_z}\Sigma_k} + \omega_k}{1 - \beta\left((\gamma_e + \gamma_h)\frac{(1-\tau_n)(1-\tau_y)\Sigma_h}{\sqrt{1-s_z}\Sigma_k} + \gamma_k + (\gamma_e + \gamma_h)\tau_G\right)} \tag{29}$$

where the dispersion ratio for households and colleges is given by

$$\frac{(1-\tau_n)\Sigma_h}{\Sigma_k} = \frac{1-\gamma_k - \tau_G(\gamma_e + \gamma_h)}{(\gamma_e + \gamma_h)(1-\tau_y)\sqrt{1-s_z}}. \tag{30}$$

These expression are similar to those in the simplified model, given by 24 and 27, except for the presence of two terms. First of all, the term $\omega_z \alpha_h \frac{\Sigma_h}{\sqrt{1-s_z}\Sigma_k}$ captures the additional incentive to do research to attract high-ability students. Second, the endogenous variable s_z measures the share of variance in college quality that stems from differences in student ability.[12] In the simple model without peer-effects, $s_z = 0$. In general it is given by

$$s_z = \frac{\sigma_z^2}{\left(1 + \frac{(1-\tau_n)(1-\tau_y)}{\alpha_h}\frac{\epsilon_1}{\epsilon_2}\right)^2 \Sigma_h^2 + \sigma_z^2}. \tag{31}$$

In the full model, the marginal value of research v is affected by the progressivity of financial aid, even in the long run, through s_z. As τ_n increases, the dispersion in tuition revenues shrinks but the supply of high-ability students – the dispersion of student ability – doesn't change as much. This mitigates the decrease in the dispersion of college

[12] One cannot express the steady-state s_z as a function of parameters of the model. The full model is solved numerically.

quality implied by the decrease in the dispersion in tuition revenues. This in turn alleviates the negative impact of the policy on the incentives to do research. As a result, the net effect in the long-run on the cross-subsidization rate is likely to be positive. This can be seen from equations 30 and 29: the policy results in an increase in the importance of student ability in driving the variation in college quality, s_z, which may lead to an increase in the marginal value of research, v.

Finally, Proposition 5 in the appendix gives an analytical expression for the laws of motion of the distributions of h and k. These expressions fully characterize the dynamic of the economy. The analytical expression will prove useful to efficiently estimate the parameters of the model in the next section and analyze the full transition path to the steady-state implied by the policy change in section 6.

5 Calibration

We calibrate the model to match three key sets of targets pertaining to (i) the level of household and government spending on higher education; (ii) the distribution of those resources across universities, and how they are spent; and (iii) the sorting and stratification of students across schools. We show in particular that the model can replicate many of the important stylized facts in section 2 and validate the calibrated model's predictions on the empirical distribution of university R&D spending.

5.1 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. As in section 2, our primary data source for university level characteristics is the National Center of Education Statistics' IPEDS, which we merge with the National Science Foundation's HERD survey, the AUTM's Patent Licensing Survey, and bibliometric on publications and citations from the CWTS Leiden Rankings. Data on the sorting of students by ability and parental earnings is taken from the National Longitudinal Survey of Youth (NLSY) 1997. Data on government tuition subsidies comes from the National Postsecondary Student Aid Study. Aggregate statistics on income inequality and aggregate household spending on education are from the Congressional Budget Office (CBO) and the OECD's annual Education at a Glance. Appendix C contains additional details on the sample and data sources.

Table 1: Externally calibrated parameters

Parameter	Description	Value	Source
η	Inv. elast. of labor supply	2.00	Standard
ι_K	Spillover of knowledge	0.10	Guellec et al. (2001)
τ_y	Income Tax Progressivity	0.15	Heathcote et al. (2017)
τ_n	Tuition Subsidy Progressivity	0.18	NPSAS

5.2 Model Fit and Parameter Values

We begin by calibrating a number of parameters that can be set externally to the model's equilibrium. These parameters are listed in Table 1. The household's elasticity of labor supply η is set to its commonly assigned value in macroeconomic models. The strength of spillovers from university research to private sector productivity is governed by ι_k. We follow the literature and use values of ι_k from work estimating the importance of spillovers using aggregate data.

The remaining two sets of exogenous parameters correspond to the level and progressivity of income taxes $(\bar{a}_y, \tau_y)$ and higher education tuition subsidies $(\bar{a}_n, \tau_n)$. For the income tax schedule, we take estimates from Heathcote, Storesletten, and Violante (2017) derived from CPS data and the NBER's TAXSIM program. For tuition subsidies, we use micro data from the National Postsecondary Student Aid Study (NPSAS) on student financial assistance, tuition, and parental incomes to estimate a tuition subsidy schedule $(\bar{a}_n, \tau_n)$. Specifically, we estimate the student-aid progressivity parameter τ_n in (4) using the regression,

$$\log(\text{net tuition}) = \tau_n \cdot \log(\text{household income}) + \mathbf{X}'\boldsymbol{\beta} + \epsilon \tag{32}$$

where $\mathbf{X}$ includes the log of ACT scores and a constant.[13] We set the level of the subsidy schedule, $\bar{a}_n$, to match the average public subsidy to higher education.[14]

The model's remaining parameters are jointly calibrated internally to match key equilibrium characteristics of the market for higher education and the resources society invests in it. Table 2 reports the calibrated values and Table 3 summarizes how well the model fits the data.

[13]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).

[14]From the OECD's Education at a Glance (2020), we find $\bar{a}_n = .68$.

Table 2: Internally calibrated parameters

Parameter	Description	Value
a_y	Income tax schedule level	0.43
a_n	Tuition subsidy schedule level	-0.45
β	Inter-generational household preference	0.12
σ_y	Labor market productivity shock	0.29
σ_z	Children ability shock	2.39
α_q	Elasticity of human capital w.r.t. college quality	0.21
α_h	Elasticity of children ability w.r.t. parents' human capital	0.26
ω_k	Elasticity of school quality w.r.t knowledge	0.46
ω_z	Elasticity of school quality w.r.t peer effects	0.36
ω_e	Elasticity of school quality w.r.t equipment	0.08
γ_k	Elasticity of knowledge w.r.t past knowledge	0.74
γ_e	Elasticity of knowledge w.r.t equipment	0.13
γ_h	Elasticity of knowledge w.r.t faculty peer effects	0.11
(a_G,τ_G)	External research grant and contract award schedule	(0.0048, 0.82)

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

For the purpose of exposition, it is useful to think about the parameters in two groups. The first five parameters $(\beta, \sigma_y, \sigma_z, \alpha_q, \alpha_h)$ constitute the first group. Together these parameters govern the process of human capital accumulation, the degree of heterogeneity in ability and income, and preferences for education quality. Appropriately, these parameters are most closely associated with the five household sector data targets, which are the inter-generational dynamics of ability and earnings, the distribution of household income, and the share of household expenditure on education. In particular, to discipline the parameters governing the intergenerational transmission of ability (α_h, σ_z), we run in the NLSY-1997 a regression of children's ASVAB scores on parent's earnings.[15] The slope of the regression is closely related to the elasticity of transmission , α_h, while the share of total variance explained by the variance of parent income, R^2, is inversely related to the standard-error of the ability shock, σ_z.

The remaining eight parameters in Table 2 govern the technologies used by the higher education sector.[16] These parameters govern the relative productivity of teaching and

[15] 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.

[16] For a more parsimonious calibration, we assume the university's technology is constant returns to

Table 3: Jointly fit data targets for internal calibration

Description	Source	Data	Model
Government Sector			
Average household income tax	CBO	0.20	0.20
Average student education subsidy	OECD	0.68	0.68
Household Sector			
Log standard deviation household income	CBO	0.84	0.88
Reg. test-scores on parent's earning (slope)	NLSY	0.12	0.13
Reg. test-scores on parent's earning (R^2)	NLSY	0.11	0.11
Share of household income spent on education	OECD	1.6%	1.6%
Inter-generational elasticity (IGE)	Mazumber (2015)	0.4	0.41
Higher Education Sector			
Log standard deviation university revenues	IPEDS	0.63	0.59
Tuition share in total university revenue	IPEDS	0.83	0.83
Tuition elasticity w.r.t. total revenue	IPEDS	0.64	.71
Research share in total university expenditure	IPEDS & HERD	0.24	0.25
Research elasticity w.r.t. total expenditure	IPEDS & HERD	2.40	2.51
Publication elasticity w.r.t. research expenditure	IPEDS & HERD	0.87	0.87
Equipment expenditure share in teaching	IPEDS	0.40	0.40
Equipment expenditure share in research	IPEDS & HERD	0.54	0.54

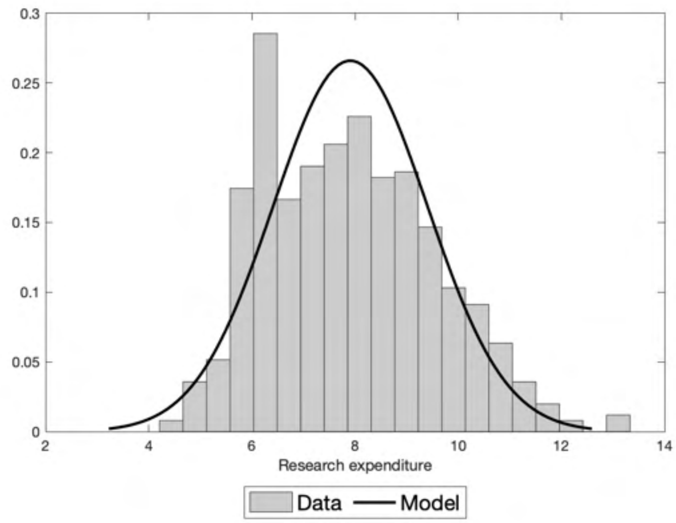
Notes: *Additional details are contained in appendix C.*

research activities, as well as the elasticities of school quality with respect to its various inputs (e.g. student ability, teacher quality, capital equipment, scientific capital). We fit these parameters to match the stylized facts introduced in section 2. Specifically, we include moments that make the model reproduce (i) the observed distribution of financial resources across universities, (ii) the source of those resources (e.g. tuition, government grants), and (iii) the use of those funds (e.g. research v. teaching, faculty v. equipment), in addition to how all three of those allocations varies with university research expenditure.

As Table 3 illustrates, the model does very well at matching both the aggregate and dis-

scale so that $\omega_k + \omega_z + \omega_e + \omega_h = 1$, eliminating a parameter and removing a degree of freedom from the model.

Figure 9: 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. The underlying data come from IPEDS. See appendix C for more details.*

tributional 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 can internally generate an empirically realistic distribution of university research expenditures. Our calibration strategy does not directly target any properties of this distribution. Nevertheless, in light of the propositions in section 4, the model should generate a realistic distribution of research spending provided the research cross-subsidization rate and overall distribution of resources are accurately captured. This makes matching the distribution of research expenditures a natural validation test of the model's internal mechanics. Figure 9 plots the in-sample empirical distribution of university R&D expenditure and compares it to our model's predicted distribution. While the data exhibit slightly more skewness, the calibrated model performs well at replicating the untargetted distribution of university R&D.

6 Quantitative Implications of Tuition Policies for University R&D

To assess the quantitative importance of the economic mechanisms introduced above, we use the calibrated model to measure the impact on university research spending of recently proposed changes to federal tuition subsidies. In particular, we focus on simulating the short-run and long-run consequences of President Biden's recently proposed "free-tuition" policies on university R&D spending and the market structure of higher education.[17]

6.1 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 need-based 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 (32) 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 net tuition payment after the policy, we can compute the average rate of college subsidies $\bar{a}_n = 0.967$.

[17] This version of the program is based on the proposal available on the campaign website, here.

6.2 Results: The Long-Run

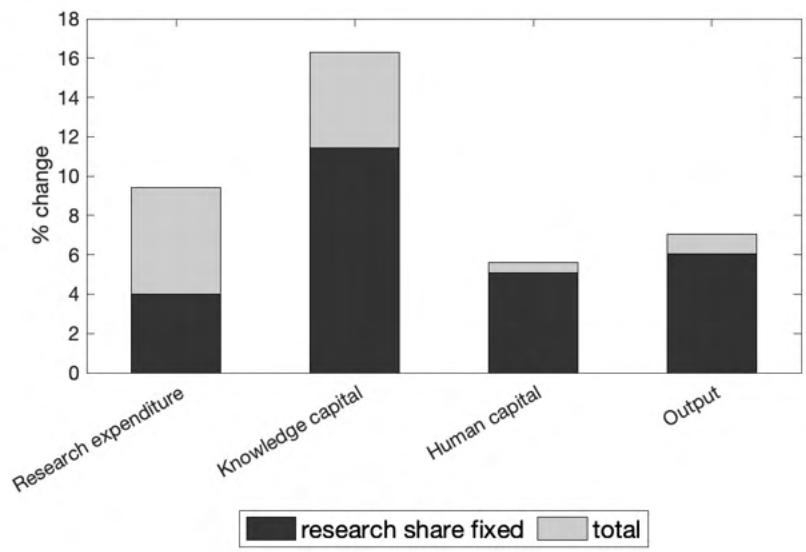
We find that proposed changes to the progressivity of federal student-aid programs will increase university R&D spending by 9.43% and aggregate output by 7.04% in the long-run. The model also predicts that, as a result of the policy, the human capital attained by college graduates will increase by an average of 5.61%, and the scientific output of universities could increase by up to 16.29%.

To better understand why research increases, Table 10 decomposes the long-run effects into a "market size" effect and one due to changes in the rate of cross-subsidization, s_R. The market size effect captures the impact of general equilibrium changes in the level of university revenues, holding the cross-subsidization rate s_R constant. The second component of the decomposition isolates the impact of changes in the equilibrium cross-subsidization rate s_R, holding university revenues at their initial levels. Together the two components comprise the total long-run effect of the policy change. Table 10 shows that the bulk of the effect, particularly on human capital and output, come from expansions in spending on higher education. However, a substantial portion of the increase in university research expenditure and scientific output comes from increases in the intensity of R&D spending by individual universities. This is due to the cross-subsidization rate, s_R which increases by over 13% from an initial steady state level of 10.2% of revenue to 11.6% after the policy change. This increase in research *intensity* accounts for more than half of the rise in research expenditure, 5.42% of the 9.43% rise, and nearly a third of the rise in scientific output, 4.86% of the 16.29% increase.

Finally, it should be noted that the overall impact of both channels is large. As a comparison, holding constant the higher education sector's share in aggregate R&D (see Figure 1), the results imply an increase in aggregate R&D expenditure of roughly 1.23% and an increase in aggregate basic R&D of roughly 4.57%.

The source of these long-run changes can be traced back to two direct effects of more progressive financial aid policies on the market for higher education. First, they alleviate the borrowing constraint faced by poor households, increasing household spending on education and improving the positive assortative matching of the best students to the best colleges. More spending on education and improving the sorting of students increases the efficiency of the higher education system and leads students to graduate with more human capital. This leads to a more skilled workforce, with higher incomes that increase the economy's absolute spending on education in the long-run. These adjustments are the driving force behind the market-size effect in our decomposition.

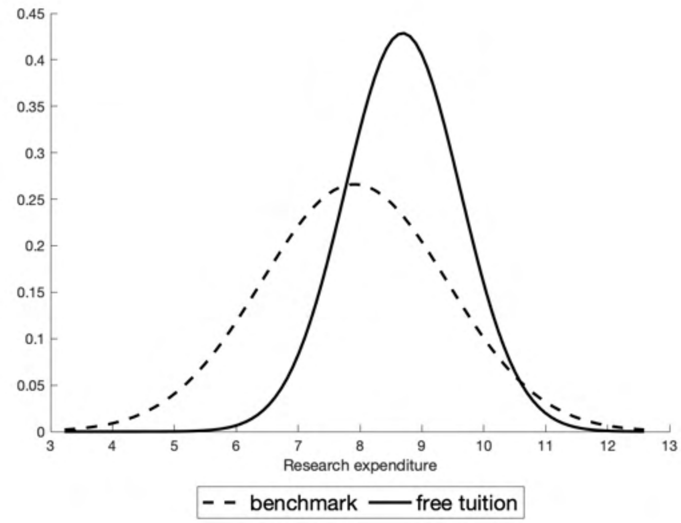
Figure 10: Long-run consequences of increasing the progressivity of student-aid



Notes: *Bars represent the average percent change relative to the initial calibrated steady state. The light colored partition of each bar indicates the contribution of changes in the cross-subsidization rate (s_R) to the total outcome.*

Second, increasing the progressivity of student-aid will reduce the dispersion in household spending on higher education. This happens by design as more progressive aid programs are a transfer from the households which spend the most on education to those who spend the least. In the long-run, this compression in household willingness to spend on education exerts opposing forces on university research spending. On the one hand, since in equilibrium higher income households send their children to higher quality colleges, the compression in spending also reduces the difference in tuition revenue between high- and low-quality colleges. This discourages university investment in quality enhancing initiatives, such as well-funded research programs, leading to a fall in the cross-subsidization rate s_R. On the other hand, the compression in spending shrinks the ex-ante differences between colleges in the long-run, making it easier for low-quality colleges to catch up with high-quality colleges and exposes high-quality colleges to more competition from lower ranked institutions. This increased competition reinforces the university incentives to invest in research and leads to an increase in the cross-subsidization rate s_R. In the simple model, theses two forces cancel out in the long-run and the rate s_R is left unchanged. In the complete model, the latter effect

Figure 11: The Distribution of University Research Expenditures



Notes: *The dashed line depicts the calibrated model's fit for the empirical distribution of university R&D validated in Figure 9. The solid line depicts the counter-factual distribution of university R&D under the simulated policy.*

dominates and the cross-subsidization rate s_R rises. This is because the progressive policies undue part of the misallocation of talent implied by the financial friction and lead to stronger assortative matching of talent students and high quality programs. In the presence of peer-effects, this increases the university's incentive to invest in quality enhancing initiatives as they will be reinforced by attracting a more talented student body. This increases the incentive faced by universities to spent competitively on research relative to the simple model, leading to an ultimate rise in s_R. Figure 11 illustrates how the aggregate increase in university R&D is impacted by this competitive channel, as a substantial part of the total increase is driven by a compression in spending driven by a rise in spending at lower-quality schools.

6.3 Results: The Transition Path

A useful property of our model is that the transition path can be analytically characterized (see Proposition 5) which allows for efficient computation. This applies to both the first and second moments of the state variable distributions (k and h), and to the equilibrium sorting rule and tuition schedule in Proposition 1. Examining the economy in transition reveals that the short-run and long-run effects of more progressive tuition policies can look quite different. Most notable is the behavior of the cross-subsidization

s_R which falls in the short-run, despite accounting for over half of the long-term rise in university research expenditure. An important implication of this is that high-frequency studies of policy interventions in the market for higher education can be misleading as to its permanent long-run impact, as is the case here.

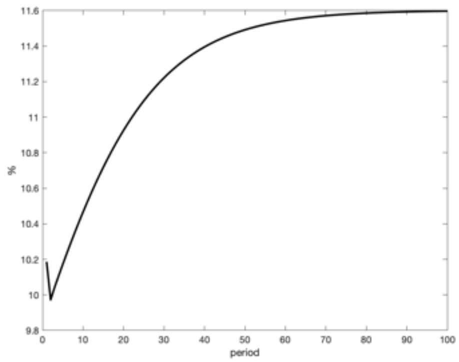
Figure 12 illustrates this point by plotting the transition path of several key equilibrium variables. The transition is triggered by an unexpected one-time and permanent increase in the progressivity of the need-based aid schedule. The first point in each graph corresponds to the pre-intervention calibrated steady state.

The two top panels illustrate the main components of the long-run decomposition in section 6.2: the cross-subsidization rate, measured by s_R, and the market-size effect, which depend on university revenues. Together these constitute the long-run dynamics of university research expenditure, depicted in panel (c). One noteworthy feature of the transition is that while university research expenditure reaches it new steady level very quickly, the forces which sustain it vary over time. The initial rapid increase in university R&D is driven by the market-size effect with a jump in university revenue flowing directly from the increased generosity of federal tuition subsidies ($\bar{a}_n$). Over time these revenues are eroded by competition and universities sustain research expenditures through a rise in the cross-subsidization rate s_R.

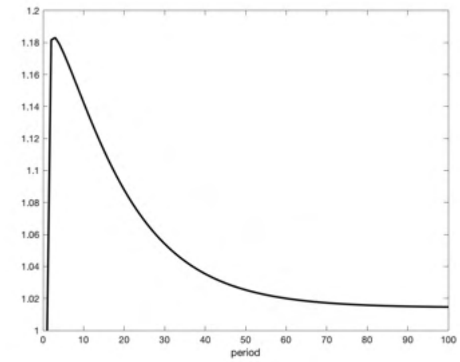
These dynamics emerge from two opposing forces triggered by more progressive financial aid programs which unfold deferentially over time. The first stems from the fact need-based aid programs compress the distribution of household expenditure on education. This occurs by design, as such programs are essentially a transfer from high-spending households to low-spending ones. Over time, this compression reinforces itself–evident in panel (e)–as more equitable spending on education leads to more equitable labor market outcomes, which further influences the future distribution of household spending on education. An immediate consequence of this compression in spending is a fall in demand for the most expensive, highest quality schools which, in equilibrium, are disproportionately attended by the children of wealthy families. This fall in demand at the top reduces university incentives to climb the quality latter by investing in research, leading to the short-term fall in the cross-subsidization rate s_R.

Over time, the differences between higher-quality and lower-quality schools shrinks (evident in the fall in Σ_k) driven by both the homogenization of household abilities and education spending, as well as the initial disinvestment in university research. As colleges become more similar, a competition effect emerges as lower ranked institutions

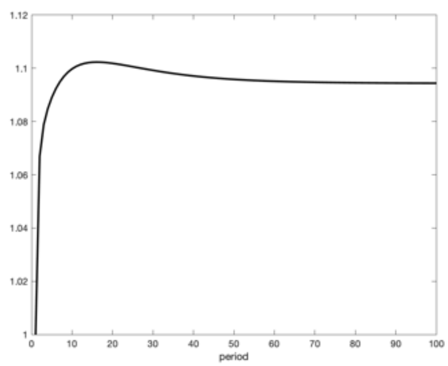
Figure 12: Transition Path following Student-Aid Policy Intervention



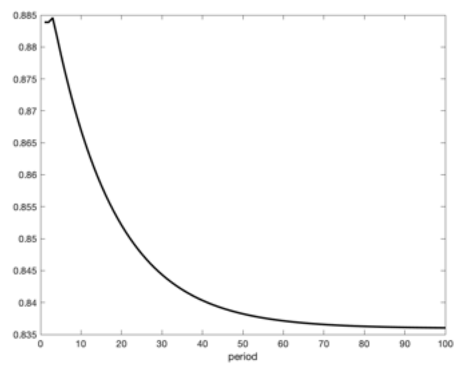
(a) Cross-Subsidization Rate (s_R)



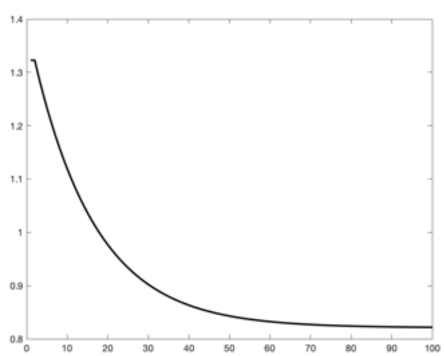
(b) University Revenue



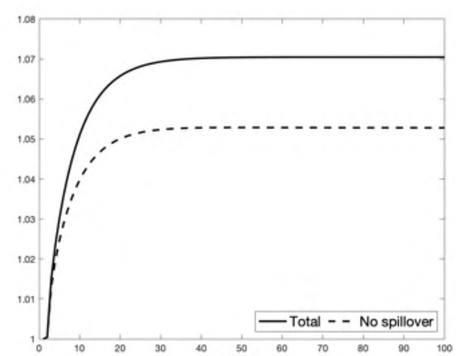
(c) Research Expenditure



(d) Household Income Inequality (Σ_h)



(e) Std. Knowledge Capital (Σ_K)



(f) Aggregate Output (w/ spillover)

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.*

suddenly find it possible to catch up, exposing higher quality schools to intensified competition. This incentivizes universities to spend more aggressively on research, leading to a rise in the cross-subsidization rate s_R in the long-run. Since the increase in competition materializes slowly over time, the fall in demand dominates initially, causes the initial drop in s_R before it's gradual long-run rise.

In addition to compressing household education expenditure, more progressive financial-aid programs also undo some of the misallocation of talent implied by the financial friction. An increase in the need-based targeting of financial-aid makes it possible for high ability children in low income households to attend better colleges. This reduction in misallocation makes the higher education sector more efficient, leading to higher human capital amongst graduates, which improves labor productivity and boosts output. The resulting expansion in income and production (owed also in part to research spillovers) reinforces the market-size channel and sustains the long-run gains in university revenues.

7 Conclusion

This paper argues that a key motivation for higher education R&D spending is the competition between universities for tuition revenue and high ability students. University research is an investment that increases the teaching quality of its faculty, improving the quality of education and enabling a university to command higher tuition and attract better students. Consequently, university research is cross subsidized with tuition revenue. A key implication is that the market structure of higher education will have a substantial influence on the provision of university research. Our calibrated model suggests these channels are important in practice; recently proposed changes to federal tuition policies could increase the higher education sector's research spending by 9.43%.

An important task for future research is investigating the role of government research grants in a framework where education and research activities are interdependent. While our analysis has focused on the equilibrium effects of tuition policies on the provision of research, one could similarly investigate the equilibrium impact of government research subsidies on the structure of education markets and its outcomes. The on-going debates in the European Union about the reforms of research grants towards a more competitive allocation of resources contrast with the widely accepted and unquestioned allocation system of grants in the United States. But little is known about the aggregate implications of these systems. More research is needed to inform the design of optimal allocations of research grants that takes a comprehensive view of the interdependent functions of the higher education sector.

More work is also needed to understand the relationship between fundamental research conducted by universities and private sector innovation. Making progress requires a better understanding of the precise role that academic research and fundamental knowledge play in private sector innovation. An important unanswered question is the extent to which the provision of fundamental research is optimal. Given the distance between the university's, the entrepreneur's and the society's objectives, it is likely that only serendipity would lead the higher education sector's investment in fundamental research to be socially optimal. The framework developed in this paper is a natural starting point to account for the distinct dual contribution of universities to the economy's long-run innovation: discovering fundamental knowledge and building human capital.

References

Adams, James D. 1990. "Fundamental Stocks of Knowledge and Productivity Growth." *Journal of Political Economy* 98 (4): 673–702.

Aghion, Philippe, Mathias Dewatripont, and Jeremy C. Stein. 2008. "Academic freedom, private-sector focus, and the process of innovation." *The RAND Journal of Economics* 39 (3): 617–635.

Akcigit, Ufuk, Santiago Caicedo, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi. 2018, March. "Dancing with the Stars: Innovation Through Interactions." Nber working papers 24466, National Bureau of Economic Research, Inc.

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.

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

Cai, Zhifeng, and Jonathan Heathcote. 2020. "College Tuition and Income Inequality." Staff report 569, Federal Reserve Bank of Minneapolis.

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

Colyvas, Jeannette, Michael Crow, Annetine Gelijns, Roberto Mazzoleni, Richard R. Nelson, Nathan Rosenberg, and Bhaven N. Sampat. 2002. "How Do University Inventions Get into Practice?" *Management Science* 48 (1): 61–72.

Cozzi, Guido, and Silvia Galli. 2014. "Sequential RD and blocking patents in the dynamics of growth." *Journal of Economic Growth* 19 (2): 183–219.

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.

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).

Henderson, Rebecca, Adam Jaffe, and Manuel Trajtenberg. 1998. "Universities As A Source Of Commercial Technology: A Detailed Analysis Of University Patenting, 1965-1988." *The Review of Economics and Statistics* 80 (1): 119–127.

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).

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

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

Jensen, Richard, and Marie Thursby. 2001. "Proofs and Prototypes for Sale: The Licensing of University Inventions." *American Economic Review* 91 (1): 240–259 (March).

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.

MacLeod, W. Bentley, and Miguel Urquiola. 2015. "Reputation and School Competition." *American Economic Review* 105 (11): 3471–88.

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

Mowery, David, Richard Nelson, Bhaven Sampat, and Arvids Ziedonis. 2001. "The Growth of Patenting and Licensing by U.S. Universities: An Assessment of the Effects of the Bayh-Dole Act of 1980." *Research Policy* 30 (01): 99–119.

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.

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

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

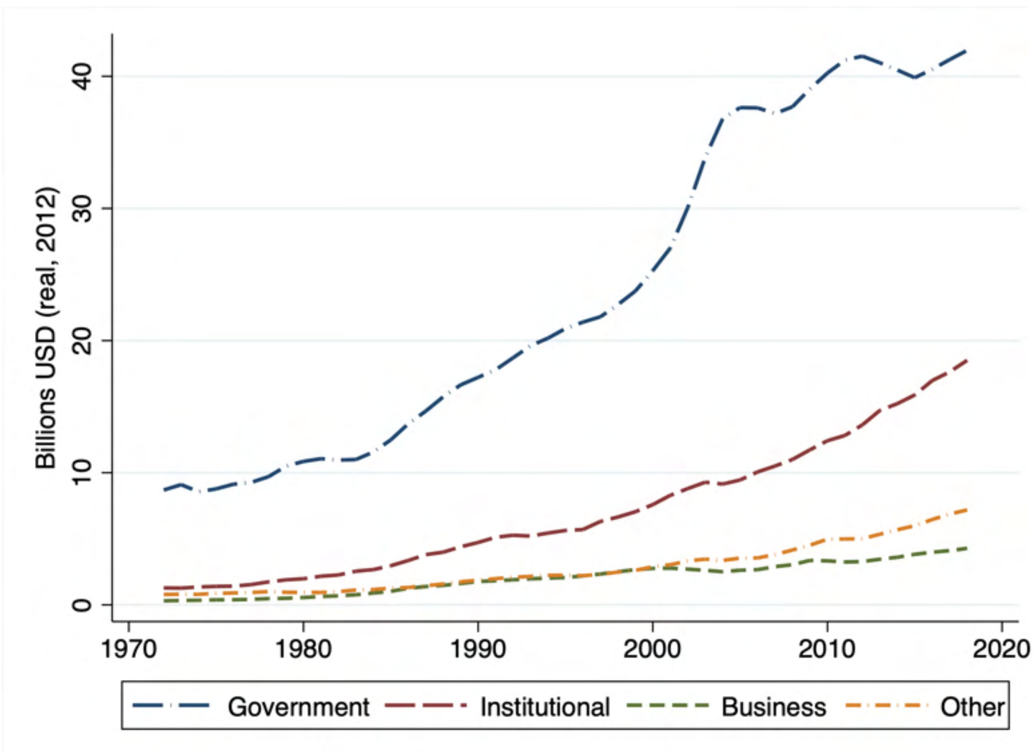
Sampat, Bhaven N. 2006. "Patenting and US academic research in the 20th century: The world before and after Bayh-Dole." *Research Policy* 35 (6): 772–789. Property and the pursuit of knowledge: IPR issues affecting scientific research.

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

Appendix

A Additional Figures

Figure A1: University research spending by source of funds

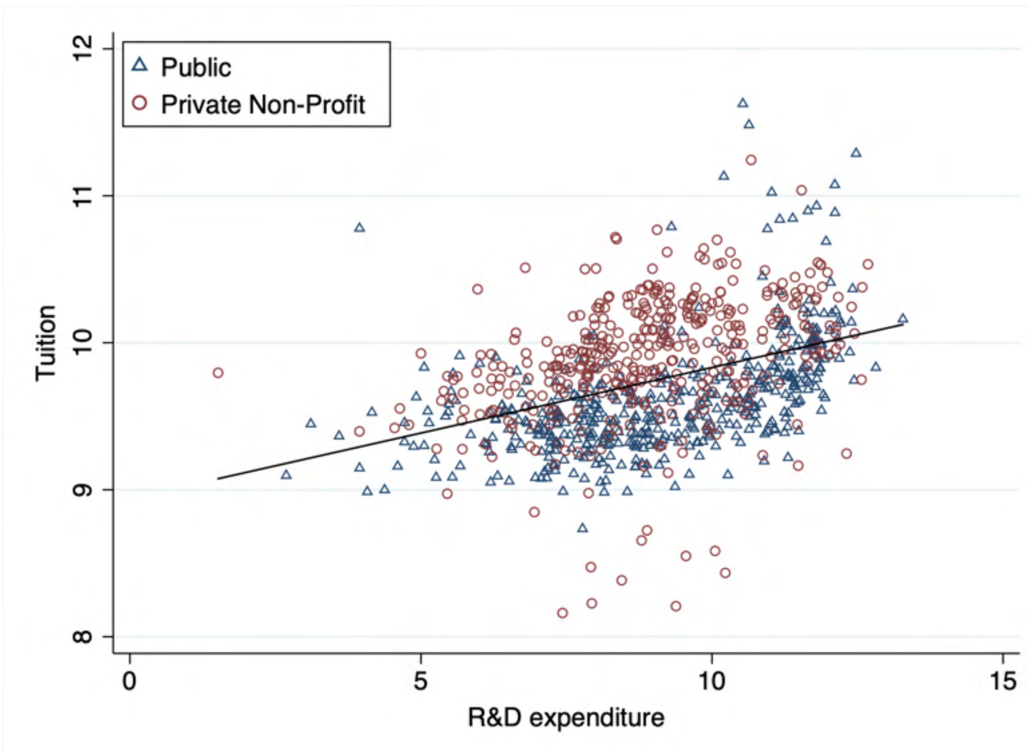


Notes: *Y-axis reports share of total university research spending by source of funds. Government sources includes federal, state, and local grants and contracts, though the federal component represents the vast majority. Non-profit funding included in Other category. Data from National Science Board (2018).*

Figure A2: University research spending and new start-ups

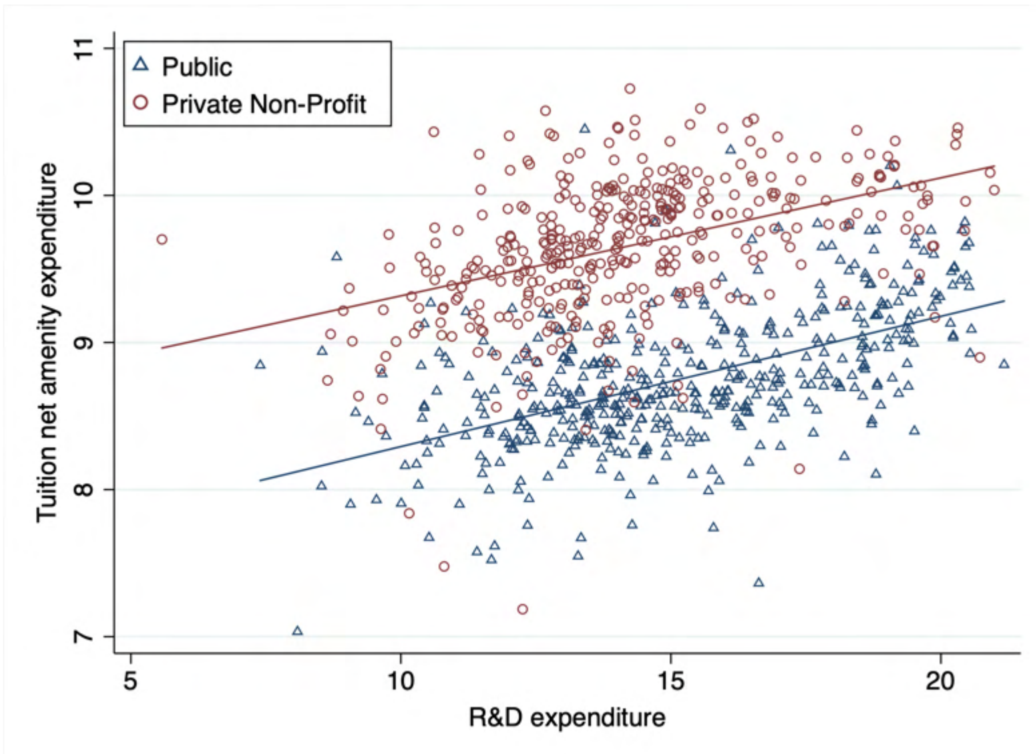
Notes: *Data from the AUTM Licensing Activity Survey. Points represent university level averages for 2012-2015.*

Figure A3: 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 A4: University research spending and tuition net 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.*

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 come the from National Science Foundations 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. [add details on each variable]

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). [add details]

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. [add details]

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. [add details]

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. [add details]

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. [add details]

D Proof Appendix

D.1 Government Budget Constraints

We denote $\bar{G}$ the total spending on research grants, $\bar{a}_y$ and $\bar{a}_n$ the average tax rate on income and the average rate of tuition subsidy:

$$\bar{G} = \mathbb{E}_j \left[\left[1 - G k_j^{-\tau_G} \right] \left(e_{Ij} + \mathbb{E}_{\mu_{Ij}(.)} \left[w_I(h) \right] \right) \right] \tag{33}$$

$$(1 - \bar{a}_y) \int w h_k \ell_i di = \int (1 - a_y) \left(w h_k \ell_i \right)^{1-\tau_y} di \tag{34}$$

$$(1 + \bar{a}_n) \int \frac{y_i^{\tau_n}}{(1 + a_n)} p_{ui} di = \int p_{ui} di. \tag{35}$$

The government balances its budget every period:

$$\bar{a}_c \int c_i di + \bar{a}_y \int w h_k \ell_i di = \bar{a}_n \int \frac{y_i^{\tau_n}}{(1 + a_n)} p_{ui} di + \bar{G}. \tag{36}$$

D.2 University problem

Let's define $s_{eI}(q), s_{eq}(q), s_{hI}(q), s_{hq}(q)$ the share of per-student tuition revenues at a university with quality q, $R(q) = \mathbb{E}_{\phi(.)} \left[p(q, z) \right]$, devoted to research intermediate goods $s_{eI}(q) = G(k) p_e e_k(q) / R(q)$, to teaching equipment $s_{eq}(q) = p_e e_q(q) / R(q)$ where p_e denotes the price of equipment goods and to research and teaching faculty wages, $s_{hI}(q) = G(k) p_{hI} \bar{h}_I(q) / R(q)$, $s_{hq}(q) = p_{hq} \bar{h}_q(q) / R(q)$ where $p_{hI}(q), p_{hq}(q)$ is the shadow price of research faculty quality at college of quality q. They are defined by

$$\begin{aligned} p_{hI}(q) \bar{h}_I(q) &= R(q) s_{hI}(q) \\ p_{qI}(q) \bar{h}_q(q) &= R(q) (1 - s_{eI}(q) - s_{eq}(q) - s_{hI}(q)). \end{aligned}$$

Since equipment goods are final goods, we set $p_e =$ in the rest of the proof.

Faculty Sorting Colleges choose the density of research and instructional faculty types to hire. Given our (inconsequential) assumption that there is one faculty per student, a college chooses the best possible faculty $\bar{h}_I$ given their chosen expenditures on research faculty wages $s_{hI} R$ and teaching $s_{hq} R$:

$$s_{hI} R = G(k) w(h_k) = G(k) w h_k \iff h_k = \frac{s_{hI} R}{w G(k)} \tag{37}$$

$$\text{and similarly} \quad h_q = \frac{(1 - s_{hI} - s_{eq} - s_{eI})R}{w} \tag{38}$$

The shadow price of teaching and research faculty is given by w.

Solving for v_t We now guess that in equilibrium there is a log-linear mapping between faculty peer-effect, student peer-effect, research intermediate goods, revenue per student, knowledge capital and teaching quality, namely that there exists $\chi_R, \chi_k, \chi_{h_k}, \chi_{h_q}, \chi_z, \chi_{e_k}, \chi_{e_q}$

$$\log R_t(q) = m_R + \chi_R(\log q - m_q) \tag{39}$$

$$\log k_t(q) = m_k + \chi_k(\log q - m_q) \tag{40}$$

$$\log \bar{h}_I(q) = m_{hI} + \chi_{hI}(\log q - m_q) \tag{41}$$

$$\log \bar{h}_q(q) = m_{hq} + \chi_{hq}(\log q - m_q) \tag{42}$$

$$\log \bar{z}(q) = m_z + \chi_z(\log q - m_q) \tag{43}$$

$$\log \bar{e}_I(q) = m_{eI} + \chi_{eI}(\log q - m_q) \tag{44}$$

$$\log \bar{e}_q(q) = m_{eq} + \chi_{eq}(\log q - m_q) \tag{45}$$

Using these guesses we now 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

$$\mathcal{V}(k_i, t) = \bar{v}_t + v_t \ln k_i \tag{46}$$

Replacing this guess into the expression for the value function (15) and using our guesses (42)-(40) gives

$$\begin{aligned}
&(\bar{v}_t + v_t \ln k_{it}) = \ln q + \beta\left(\bar{v}_{t+1} + v_{t+1} \ln k_{it+1}\right) \\
&= \ln q + \beta v_{t+1}\left(\gamma_h \ln \bar{h}_{I,t} + \gamma_e \ln e_{I,t} + \gamma_k \ln k_i\right) + \beta \bar{v}_{t+1} \\
&= \ln q + \beta v_{t+1}(\gamma_e(m_{eI} + \chi_{eI}(\ln q - m_q)) + \gamma_h(m_{hI} + \chi_{hI}(\ln q - m_q)) + \gamma_k \ln k) + \beta \bar{v}_{t+1} \\
&= \ln q + \beta v_{t+1}((\gamma_e \chi_{eI} + \gamma_h \chi_{hI})(\ln q - m_q) + \beta v_{t+1}\gamma_k \ln k + const. \\
&= (1 + \beta v_{t+1}(\gamma_e \chi_{eI} + \gamma_h \chi_{hI}))\frac{1}{\chi_k}(\ln k - m_k) + \beta v_{t+1}\gamma_k \ln k + const.
\end{aligned}$$

Gathering all terms in $\ln k_{it}$ gives

$$v_t = \frac{1}{\chi_{k,t}} + \tilde{\beta}_t v_{t+1} \tag{47}$$

$$\text{with} \quad \tilde{\beta}_t = \beta\left(\gamma_k + \gamma_e \frac{\chi_{e_k t}}{\chi_{kt}} + \gamma_h \frac{\chi_{h_k t}}{\chi_{kt}}\right) \tag{48}$$

whose solution is

$$v_t = \sum_{\tau=0}^{+\infty} \left[\Pi_{j=0}^{\tau} \tilde{\beta}_{t+j}\right] \frac{1}{\chi_{k,\tau}}$$

When all elasticities are constant, it simplifies to

$$v_t = \frac{1}{\chi_k} \frac{1}{1 - \beta\left(\gamma_k + \gamma_e \frac{\chi_{e_k}}{\chi_k} + \gamma_h \frac{\chi_{h_k}}{\chi_k}\right)} \tag{49}$$

In the main text, we use the following notations

$$\begin{aligned} \epsilon_{qk} = \chi_k^{-1} \qquad \epsilon_{e_I k} = \frac{\chi_{e_k}}{\chi_k} \qquad \epsilon_{h_I k} = \frac{\chi_{h_k}}{\chi_K} \\ v_t = \sum_{\tau=0}^{+\infty} \left[\Pi_{j=0}^{\tau} \tilde{\beta}_{t+j}\right] \epsilon_{qK\tau} \\ \text{with} \quad \tilde{\beta}_t = \beta\left(\gamma_k + \gamma_e \epsilon_{e_I k} + \gamma_h \epsilon_{h_I k}\right) \end{aligned} \tag{50}$$

An equivalent static problem. 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

$$\max_{q,\phi(z),s_{hI},s_{eI},\mu(h)} \ln q + \omega_{It} \ln k' \tag{51}$$

$$\text{with} \quad \omega_{It} = \beta v_{t+1} \tag{52}$$

$$\text{subject to} \quad \ln q = \ln \bar{h}_q^{\omega_h} e_q^{\omega_e} \bar{z}^{\omega_z} k^{\omega_k} - \sigma_{u,i}^2 \tag{53}$$

$$k' = k^{\gamma_k} e_k^{\gamma_e} \bar{h}_I^{\gamma_h} \tag{54}$$

$$\mathop{\mathbb{E}}_{\phi_j(.)}[p(q,z)] = G(k)\left[p_e e_k + \int \mu_q(h) w_f(h)\right] + p_e e_q + \int \mu_k(h) w_f(h) \tag{55}$$

This setting is attractive for two reasons: i) the weight on research is endogenous and capture the discounted payoffs of knowledge production, ii) this weight is common across all colleges. A key component of the weight on research is the marginal value of knowledge capital, v_t, which we have solved for in the previous paragraph.

Optimal Policy Functions. Using the definition of the share of expenditures devoted to research wages s_{hI}, and research and teaching equipments s_{eI}, s_{eq}, teaching quality becomes

$$q_i = \left(\underset{\phi_j(.)}{\mathbb{E}} [p(q,z)] s_{hq}/w \right)^{\omega_h} \left(\underset{\phi_j(.)}{\mathbb{E}} [p(q,z)] s_{eq} \right)^{\omega_e} \bar{z}^{\omega_z} k_i^{\omega_k} - \sigma_{u,i}^2$$

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

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

The cost of heterogeneity across students σ_n is assumed to have the following form:

Assumption 1.

$$\sigma_u^2(\phi; p) = \frac{\Omega}{2} V_{\phi(.)}(\ln p(q,z)) \tag{56}$$

$$\textit{with} \quad \Omega = \omega_e + \omega_h + \omega_I (\gamma_e + \gamma_h) \tag{57}$$

This choice for Ω ensures the tractability of the college problem. We verify later that the guess that tuition are log-normally distributed within a college is true as well as the guess that the variance of log-tuition σ_u^2 is independent of colleges.

These two assumptions imply that the college problem becomes fully log-linear in tuition and student' ability

$$\begin{aligned} \max_{q,\phi(z),s_{hI},s_{hq},s_{eI},s_{eq}} \ln & \left(\frac{s_{hq}\tilde{R}}{w} \right)^{\omega_h} \left(\tilde{R} s_{eq} \right)^{\omega_e} \bar{z}^{\omega_z} k_i^{\omega_k} \\ & + \omega_{It} \ln \left(\frac{\tilde{R} s_{eI}}{G(k)} \right)^{\gamma_e} \left(\frac{s_{hI}\tilde{R}}{wG(k)} \right)^{\gamma_h} k_j^{\gamma_k} \\ \text{with} \quad \ln \tilde{R} = & \underset{\phi_j(.)}{\mathbb{E}} [\ln p(q,z)] \end{aligned}$$

where we have also used the expression for faculty quality as a function of college income, (37) and (38).

An equilibrium where colleges are indifferent across students requires that tuition should be equal to

$$0 = (\omega_e + \omega_h + \omega_{It} (\gamma_e + \gamma_h)) \ln \frac{p(q,z)}{\tilde{R}} + \omega_z \ln \frac{z}{\bar{z}}$$
$$\Rightarrow p(q,z) = \tilde{R} \left(\frac{z}{\bar{z}}\right)^{-\frac{\omega_z}{\omega_e + \omega_h + \omega_{It}(\gamma_e + \gamma_h)}}$$

This corresponds to the F.O.C. for $\phi(.)$. Compared to the findings in Capelle (2020), the elasticity of tuition to ability is lower because the introduction of research and knowledge production increases the relative attractiveness of financial resources and lowers the role of students abilities.

We now take the F.O.C. w.r.t. s_{eI}, s_{eq} and s_{hI}. When doing so, we assume that revenues R are given and do not vary with these three controls. This relies on the assumption 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 the equilibrium choices for s_{eI}, s_{eq} and s_{hI} and therefore deviate from its promised quality of education. In another words, we are assuming that there is commitment on the part of students to stay irrespective of the college's choice of s_{eI}, s_{eq} and s_{hI}.

$$s_{eI} = \frac{\omega_{It}\gamma_e}{\omega_{It}(\gamma_e + \gamma_h) + \omega_e + \omega_h} \tag{58}$$
$$s_{eq} = \frac{\omega_e}{\omega_{It}(\gamma_e + \gamma_h) + \omega_e + \omega_h} \tag{59}$$
$$s_{hI} = \frac{\omega_{It}\gamma_h}{\omega_{It}(\gamma_e + \gamma_h) + \omega_e + \omega_h} \tag{60}$$
$$s_{hq} = \frac{\omega_h}{\omega_{It}(\gamma_e + \gamma_h) + \omega_e + \omega_h} \tag{61}$$

D.3 Laws of Motion of Knowledge Capital and Human Capital

From the law of motion of capital, we obtain

$$\ln k_{jt+1} = \ln e_k^{\gamma_e} \bar{h}_I^{\gamma_h} k^{\gamma_k} = \ln \left(\frac{s_{eI}}{G}\right)^{\gamma_e} \left(\frac{s_{hI}}{AG}\right)^{\gamma_h} + (\gamma_e + \gamma_h) \ln R_{jt} + (\gamma_k + \tau_G(\gamma_e + \gamma_h)) \ln k_{jt}$$

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

$$\ln k_{it} \sim \mathcal{N}\left(m_{kt}, \Sigma_{kt}^2\right)$$

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

$$m_{kt+1} = \ln\left(\frac{s_{eI}}{G}\right)^{\gamma_e}\left(\frac{s_{hI}}{AG}\right)^{\gamma_h} + (\gamma_e+\gamma_h)m_{Rt} + (\gamma_k+\tau_G(\gamma_e+\gamma_h))m_{kt} \tag{62}$$

$$\Sigma_{kt+1} = (\gamma_k+\tau_G(\gamma_e+\gamma_h))\Sigma_{kt} + (\gamma_e+\gamma_h)\Sigma_{Rt} \tag{63}$$

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

Proposition 5. *Assume human capital and knowledge capital are log-normally distributed in the first period. They are log-normally distributed along the transition path. The law of motion of the distribution of knowledge capital across colleges is given by*

$$\ln k_{it} \sim \mathcal{N}\left(m_{kt}, \Sigma_{kt}^2\right) \tag{64}$$

$$m_{kt+1} = \ln\left(\frac{s_{eI}}{G}\right)^{\gamma_e}\left(\frac{s_{hI}}{wG}\right)^{\gamma_h} + (\gamma_e+\gamma_h)m_{Rt} + (\gamma_k+\tau_G(\gamma_e+\gamma_h))m_{kt} \tag{65}$$

$$\Sigma_{kt+1}^2 = (\gamma_k+\tau_G(\gamma_e+\gamma_h))^2\Sigma_{kt}^2 + (\gamma_e+\gamma_h)^2\Sigma_{Rt}^2 + 2(\gamma_k+\tau_G(\gamma_e+\gamma_h))(\gamma_e+\gamma_h)\Sigma_{Rt}\Sigma_{kt} \tag{66}$$

The law of motion of the distribution of human capital across households is given by

$$\ln h_{t+1} \sim \mathcal{N}\left(m_{ht+1}, \Sigma_{ht+1}^2\right) \tag{67}$$

$$m_{ht+1} = \rho_t m_{ht} + X_{1t} \tag{68}$$

$$\Sigma_{ht+1}^2 = \rho_t^2\Sigma_{ht}^2 + \sigma_y^2 + \left(\alpha_h\left(\alpha_q\epsilon_{2t}+1\right)\right)^2\sigma_z^2 \tag{69}$$

where $X_{1t} = -\frac{\sigma_y^2}{2} - \alpha_h\left(\alpha_q\epsilon_{2t}+1\right)\frac{\sigma_z^2}{2} + \alpha_q\epsilon_{1t}\ln\left(\frac{s_t(1+a_n)}{\underline{p}_t}(A\ell)^{(1-\tau_y)(1-\tau_n)}(1-a_y)^{(1-\tau_n)}\right)$

D.4 Solving for all guesses

Step 1: ϵ_1**,** $\underline{p}_t$ Using the expression for tuition and the definition of teaching quality (9) together with our guesses one has

$$p(q,z)$$

$$
\begin{aligned}
&= \tilde{R}\left(\frac{z}{\bar{z}}\right)^{-\frac{\omega_z}{\omega_e+\omega_h+\omega_{It}(\gamma_e+\gamma_h)}} \frac{\left(k_i^{\omega_k} e_k^{\omega_e}\right)^{\frac{1}{\omega_e+\omega_h+\omega_{It}(\gamma_e+\gamma_h)}}}{\left(k_i^{\omega_k} e_k^{\omega_e}\right)^{\frac{1}{\omega_e+\omega_h+\omega_{It}(\gamma_e+\gamma_h)}}} \frac{p_h(q) s_{hq}}{p_h(q) s_{hq}} \frac{e^{\frac{V(\ln e)}{2}}}{e^{\frac{V(\ln e)}{2}}} \\
&= \frac{\left[\bar{h}_q^{\omega_h} e^{\omega_e} \bar{z}^{\omega_z} k^{\omega_k}\right]^{\frac{1}{\omega_e+\omega_h+\omega_{It}(\gamma_e+\gamma_h)}} \bar{h}_q^{1-\frac{\omega_h}{\omega_e+\omega_h+\omega_{It}(\gamma_e+\gamma_h)}}}{\left(k_i^{\omega_k} e_k^{\omega_e}\right)^{\frac{1}{\omega_e+\omega_h+\omega_{It}(\gamma_e+\gamma_h)}}} \frac{p_{hq}(q) e^{-\frac{V(\ln e)}{2}} z^{-\frac{\omega_z}{\omega_e+\omega_h+\omega_{It}(\gamma_e+\gamma_h)}}}{s_{hq}} \\
&= \left(q e^{\sigma_n^2}\right)^{\frac{1}{\omega_e+\omega_h+\omega_{It}(\gamma_e+\gamma_h)}} R \frac{\bar{h}^{-\frac{\omega_h}{\omega_e+\omega_h+\omega_{It}(\gamma_e+\gamma_h)}}}{\left(k_i^{\omega_k} e_k^{\omega_e}\right)^{\frac{1}{\omega_e+\omega_h+\omega_{It}(\gamma_e+\gamma_h)}}} e^{-\frac{V(\ln e)}{2}} z^{-\frac{\omega_z}{\omega_e+\omega_h+\omega_{It}(\gamma_e+\gamma_h)}} \\
&= \underline{p}_t q^{\frac{1}{\omega_e+\omega_h+\omega_{It}(\gamma_e+\rho_h)}+\chi_R-\frac{(\omega_e+\omega_h)\chi_R}{\omega_e+\omega_h+\omega_{It}(\gamma_e+\gamma_h)}-\chi_k\frac{\omega_k}{\omega_e+\omega_h+\omega_{It}(\gamma_e+\gamma_h)}} z^{-\frac{\omega_z}{\omega_e+\omega_h+\omega_{It}(\gamma_e+\gamma_h)}} \\
&= \underline{p}_t q^{\frac{1}{\epsilon_1}} z^{-\frac{\omega_z}{\omega_e+\omega_h+\omega_{It}(\gamma_e+\gamma_h)}} \\
&= \underline{p}_t q^{\frac{1}{\epsilon_1}} z^{-\frac{\epsilon_2}{\epsilon_1}} \qquad (70)
\end{aligned}
$$

where we used $p_{hq}\bar{h}_q = s_{hq}R = s_{hq}\tilde{R}e^{\frac{V(\ln e)}{2}}$ and $\bar{h}_q = \left(\frac{s_{hq}R}{w}\right)$ as well as $p_e e_q = s_{eq}R$. Finally we define the new aggregate endogenous variables:

$$
\log \underline{p}_t = \frac{\left(\omega_I(\gamma_e+\gamma_h)m_R - \omega_k m_k - \left(\frac{\omega_e+\omega_h+\omega_{It}(\gamma_e+\gamma_h)}{\epsilon_1} - 1\right) m_q - \log\left(s_e\right)^{\omega_e}\left(\frac{s_h}{A}\right)^{\omega_h}\right)}{\omega_e+\omega_h+\omega_{It}\left(\gamma_e+\gamma_h\right)} \qquad (71)
$$

$$
\epsilon_1 = \frac{\omega_e+\omega_h+\omega_{It}\left(\gamma_e+\gamma_h\right)}{1+\chi_R\omega_{It}\left(\gamma_e+\gamma_h\right)-\chi_{k,t}\omega_k} \qquad (72)
$$

$$
\epsilon_2 = \frac{\omega_z}{1+\chi_R\omega_{It}\left(\gamma_e+\gamma_h\right)-\chi_{k,t}\omega_k} \qquad (73)
$$

where we used $e^{\frac{\sigma_n^2}{\omega_e+\omega_h+\omega_{It}(\gamma_e+\gamma_h)}-\frac{V(\log e)}{2}} = 1$.

Sorting Rule. Recall that financial aid is given by (4),$e(q,z,y) = \frac{y^{\tau_n}p(q,z)}{1+a_n}$. Hence in equilibrium, the sorting rule is given by

$$
q_t = \left(\frac{s_t(1+a_n)y^{1-\tau_{n,t}}z^{\tau_{m,t}}}{\underline{p}_t}\right)^{\epsilon_{1t}} z^{\epsilon_{2t}} \qquad (74)
$$

The elasticity of quality to income and ability which capture the strength of the income-sorting and ability-sorting channel takes into account the progressivity of taxes and financial aid:

$$
\epsilon_{I,t} = \epsilon_{1t}(1-\tau_n)(1-\tau_y) \qquad \text{elasticity to Income} \qquad (75)
$$

$$\epsilon_{A,t} = \alpha_h \epsilon_{2t} \qquad \text{elasticity to Ability} \tag{76}$$

Notice that when there is no income tax and financial aid, the income-sorting coefficient simplifies and becomes $\epsilon_{I,t} = \epsilon_{1t}$.

Household Problem and Law of Motion of Human Capital The problem of the household has been treated in Capelle (2020). From this paper and the expression for tuition obtained above (70), one directly obtains (21), (47), (77):

$$\ell_t = \left[\frac{1}{\eta}\left(1 + \frac{\beta}{1+\beta}\alpha_q \epsilon_{1t} u_{t+1}\right)\right]^{\frac{1}{\eta}} \tag{77}$$

Finally, the sorting rule (19) simply combines the expression for tuition (70) with the policy rule (21).

Similarly, the law of motion of the distribution of human capital (91) and (92) shown in proposition 6 results from taking the mean and standard deviation of the law of motion of human capital and using the policy functions of the households.

Step 2: Within College Parental Income Distribution. In the same paper, we also solved for the distribution of parental human capital within a college, students ability within a college and the distribution of quality in equilibrium.

Distribution of parental income within a college. It is given by

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

where

$$\begin{aligned}
m_{h|q} &= s_z m_h + (1-s_z)\frac{\left(\ln q - \epsilon_1(C_h - \log \underline{p}) + \epsilon_A \frac{\sigma_z^2}{2}\right)}{\epsilon_I + \epsilon_A} \\
\sigma^2_{h|q} &= s_z \Sigma_h^2 \\
C_h &= \ln\left(s(1+a_n)(w\ell)^{(1-\tau_y)(1-\tau_n)}(1-a_y)^{(1-\tau_n)}\right)
\end{aligned}$$

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

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

From the distribution of parents' human capital within a college, the distribution of parents' before tax and transfer income is

$$f(\ln y|q) \sim \mathcal{N}\left(\ln w + m_{h|q} + \ln \ell, \sigma^2_{h|q}\right).$$

Distribution of college quality

Taking log of the sorting rule,

$$\ln q \sim \mathcal{N}\left((\epsilon_I + \epsilon_A)m_h + \epsilon_1(C_h - \log \underline{p}_t) - \epsilon_A \frac{\sigma_z^2}{2}, \epsilon_A^2\sigma_z^2 + (\epsilon_I + \epsilon_A)^2\Sigma_h^2\right) \tag{78}$$

Thus

$$m_q = (\epsilon_I + \epsilon_A)m_h + \epsilon_1(C_h - \log \underline{p}_t) - \epsilon_A \frac{\sigma_z^2}{2} \tag{79}$$

and

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

Students' abilities

From the definition of abilities $\ln z = \alpha_h \ln h + \alpha_h \ln \xi_b$ and the sorting rule used above $\ln q = (\epsilon_I + \epsilon_A)\ln h + \epsilon_A \ln \xi_b + \epsilon_1(C_h - \log \underline{p})$, one gets

$$\begin{aligned}
\ln z &= \frac{\alpha_h}{\epsilon_A}\left(\ln q - \epsilon_I \ln h - \epsilon_1(C_h - \log \underline{p})\right) \\
\Rightarrow \ln z|q &\sim \mathcal{N}\left(\frac{\alpha_h}{\epsilon_A}\left(\ln q - \epsilon_I m_{h|q} - \epsilon_1(C_h - \log \underline{p})\right), \left(\frac{\alpha_h \epsilon_I}{\epsilon_A}\right)^2 \sigma^2_{h|q}\right)
\end{aligned}$$

Step 3: solve for $R(q)$ From the budget constraint, the per-student income in a college is also 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

$$\begin{aligned}
R(q) &= \int p(q,z)\phi(y,z|q)d(y,z) \\
&= \int s(1+a_n)(1-a_y)^{1-\tau_n}\,(w\ell h)^{(1-\tau_y)(1-\tau_n)}\,\phi(h|q)dh
\end{aligned}$$

$$\begin{aligned}\ln R(q) &= C_h + (1-\tau_y)(1-\tau_n)m_{h|q} + \frac{((1-\tau_y)(1-\tau_n))^2\sigma^2_{h|q}}{2} \\ &= C_h + (1-\tau_y)(1-\tau_n)\left[s_z m_h + (1-s_z)\frac{\ln q - \epsilon_1(C_h - \log \underline{p}) + \epsilon_A \frac{\sigma_z^2}{2}}{\epsilon_I + \epsilon_A}\right] \\ &+ \frac{((1-\tau_y)(1-\tau_n))^2\sigma^2_{h|q}}{2}\end{aligned}$$

Finally these results enables us to get an expression for the distribution of revenue per student across colleges:

$$\ln R \sim \mathcal{N}\left(m_R, \Sigma_R^2\right) \tag{80}$$

$$m_R = \ln s(1+a_n)(1-a_y)^{1-\tau_n}\left(w\ell\right)^{(1-\tau_y)(1-\tau_n)} + \frac{((1-\tau_y)(1-\tau_n))^2\sigma^2_{h|q}}{2} + (1-\tau_y)(1-\tau_n)m_h \tag{81}$$

$$\Sigma_R^2 = \chi_R^2\left[\epsilon_A^2\sigma_z^2 + (\epsilon_I+\epsilon_A)^2\Sigma_h^2\right] \tag{82}$$

Identifying coefficients with the guess, one gets:

$$\chi_R = (1-\tau_y)(1-\tau_n)(1-s_z)\frac{1}{\epsilon_I+\epsilon_A} = (1-\tau_y)(1-\tau_n)\frac{\left(\frac{\epsilon_A}{\epsilon_I+\epsilon_A}\right)^{-2}\sigma_z^{-2}}{\left[\Sigma_h^{-2} + \left(\frac{\epsilon_A}{\epsilon_I+\epsilon_A}\right)^{-2}\sigma_z^{-2}\right](\epsilon_I+\epsilon_A)}$$

$$\begin{aligned}&\chi_R\left(\Sigma_h^{-2}\left(\frac{\epsilon_A}{\epsilon_I+\epsilon_A}\right)^2\sigma_z^2\left((1-\tau_y)(1-\tau_n)\left[\omega_e+\omega_h+\omega_I\left(\gamma_e+\gamma_h\right)\right]+\alpha_h\omega_z\right) + (1-\tau_y)(1-\tau_n)\left[\omega_e+\omega_h\right]+\alpha_h\omega_z\right) \\ &= (1-\tau_y)(1-\tau_n)\left(1-\chi_k\omega_k\right)\end{aligned}$$

$$\chi_R = \frac{(1-\tau_y)(1-\tau_n)\left(1-\chi_k\omega_k\right)}{\left(\Sigma_h^{-2}\left(\frac{\epsilon_A}{\epsilon_I+\epsilon_A}\right)^2\sigma_z^2\left((1-\tau_y)(1-\tau_n)\left[\omega_e+\omega_h+\omega_I\left(\gamma_e+\gamma_h\right)\right]+\alpha_h\omega_z\right) + (1-\tau_y)(1-\tau_n)\left[\omega_e+\omega_h\right]+\alpha_h\omega_z\right)} \tag{83}$$

$$\tag{84}$$

$$\text{with}\quad \frac{\epsilon_A}{\epsilon_I+\epsilon_A} = \frac{\alpha_h\omega_z}{(1-\tau_n)(1-\tau_y)(\omega_e+\omega_h+\omega_{It}\left(\gamma_e+\gamma_h\right))+\alpha_h\omega_z}$$

Combining this with equation (82),

$$\Sigma_R = (1-\tau_y)(1-\tau_n)\sqrt{1-s_z}\Sigma_h. \tag{85}$$

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. Then the different income households mix more within colleges and total revenue of colleges get less dispersed, which is faster than the decrease in Σ_h. Thus Σ_K decreases faster than Σ_h as τ_n decreases and that is why the competition effect is greater as τ_n increases.

Step 4: Solving for $\underline{p}_t$ From (71), (79), and (81),

$$\begin{aligned}\epsilon_1 \log \underline{p}_t = {} & \omega_I(\gamma_e+\gamma_h)\left(C_h + \frac{((1-\tau_y)(1-\tau_n))^2\sigma^2_{h|q}}{2} + (1-\tau_y)(1-\tau_n)m_h\right) - \omega_k m_k \\ & - \left(\frac{\omega_e+\omega_h+\omega_{It}\left(\gamma_e+\gamma_h\right)}{\epsilon_1} - 1\right)\left((\epsilon_I+\epsilon_A)m_h + \epsilon_1 C_h - \epsilon_A\frac{\sigma_z^2}{2}\right) - \log\left(s_e\right)^{\omega_e}\left(\frac{s_h}{A}\right)^{\omega_h}\end{aligned} \tag{86}$$

Solving for G. We start from the budget constraint of the agency that distributes research grants and use the guesses (39), (40) and $s_{eI}(q) = G(k)p_e e_k(q)/R(q)$:

$$\begin{aligned}\int G(k)\left[p_e e_k + \int \mu_k(h)w_f(h)\right]dj + \bar{G} &= \int\left[p_e e_k + \int \mu_k(h)w_f(h)\right]dj \\ \iff \int (s_{eI}+s_{hI})R(q)dj + \bar{G} &= \int (s_{eI}+s_{hI})R(q)G^{-1}k^{\tau_G}dj \\ \iff \int (s_{eI}+s_{hI})e^{m_R - \frac{\chi_R}{\chi_k}m_k}k^{\chi_R/\chi_k}dj + \bar{G} &= \int (s_{eI}+s_{hI})e^{m_R - \frac{\chi_R}{\chi_k}m_k}k^{\chi_R/\chi_k}G^{-1}k^{\tau_G}dj \\ \iff (s_{eI}+s_{hI})e^{m_R - \frac{\chi_R}{\chi_k}m_k}\int k^{\chi_R/\chi_k}dj + \bar{G} &= (s_{eI}+s_{hI})e^{m_R - \frac{\chi_R}{\chi_k}m_k}G^{-1}\int k^{\chi_R/\chi_k+\tau_G}dj \\ \iff (s_{eI}+s_{hI})e^{m_R - \frac{\chi_R}{\chi_k}m_k}e^{\left(\frac{\chi_R}{\chi_k}m_k + \left(\frac{\chi_R}{\chi_k}\right)^2\frac{\Sigma_k^2}{2}\right)} + \bar{G} &= (s_{eI}+s_{hI})e^{m_R - \frac{\chi_R}{\chi_k}m_k}G^{-1}e^{\left(\left(\frac{\chi_R}{\chi_k}+\tau_G\right)m_k + \left(\frac{\chi_R}{\chi_k}+\tau_G\right)^2\frac{\Sigma_k^2}{2}\right)}\end{aligned}$$

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

where m_R in the last line is given by equation (81).

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

$$G = \frac{(s_{eI}+s_{hI})e^{\tau_G m_k + \tau_G\left(2\frac{\chi_R}{\chi_k}+\tau_G\right)\frac{\Sigma_k^2}{2}}}{(s_{eI}+s_{hI}) + e^{-m_R-\left(\frac{\chi_R}{\chi_k}\right)^2\frac{\Sigma_k^2}{2}}}\bar{g}Y$$
$$\text{where} \quad Y = w\exp\left(m_h + \frac{1}{2}\Sigma_h^2\right)\ell$$

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

$$\begin{aligned}
1-\bar{a}_y &= \frac{\int (1-a_y)\,(wh_k\ell_i)^{1-\tau_y}\,di}{\int wh_k\ell_i di} = \frac{(1-a_n)\,(w\ell)^{1-\tau_y}\exp\left((1-\tau_y)m_h + ((1-\tau_y))^2\,\frac{\Sigma_h^2}{2}\right)}{(w\ell)\exp\left(m_h+\frac{\Sigma_h^2}{2}\right)}\\
&= (w\ell)^{-\tau_y}\exp\left(-\tau_y m_h + \tau_y(\tau_y-2)\frac{\Sigma_h^2}{2}\right)\\
\iff a_y &= 1-(w\ell)^{\tau_y}\exp\left(\tau_y m_h + \tau_y(2-\tau_y)\frac{\Sigma_h^2}{2}\right)(1-\bar{a}_y) \qquad (87)
\end{aligned}$$

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

$$\begin{aligned}
1+\bar{a}_n &= \frac{\int (1+a_n)e_k y^{-\tau_n}di}{\int e_k di} = \frac{\int (1+a_n)sy^{1-\tau_n}di}{\int sydi} = (1+a_n)(1-a_y)^{-\tau_n}\frac{\int (wh_k\ell_i)^{(1-\tau_n)(1-\tau_y)}}{\int (wh_k\ell_i)^{1-\tau_y}}\\
\iff 1+\bar{a}_n &= \frac{(1+a_n)}{(1-a_y)^{\tau_n}}\frac{(w\ell)^{(1-\tau_y)(1-\tau_n)}\exp\left((1-\tau_y)(1-\tau_n)m_h + ((1-\tau_y)(1-\tau_n))^2\,\frac{\Sigma_h^2}{2}\right)}{(w\ell)^{1-\tau_y}\exp\left((1-\tau_y)m_h + ((1-\tau_y))^2\,\frac{\Sigma_h^2}{2}\right)}\\
\iff 1+\bar{a}_n &= \frac{(1+a_n)}{(1-a_y)^{\tau_n}}(w\ell)^{-(1-\tau_y)\tau_n}\exp\left(-(1-\tau_y)\tau_n m_h + \tau_n(\tau_n-2)\left((1-\tau_y)\right)^2\frac{\Sigma_h^2}{2}\right)\\
\iff a_n &= (1-a_y)^{\tau_n}(w\ell)^{\tau_n(1-\tau_y)}\exp\left(\tau_n(1-\tau_y)m_h + \tau_n(2-\tau_n)((1-\tau_y))^2\frac{\Sigma_h^2}{2}\right)(1+\bar{a}_n)-1
\end{aligned} \qquad (88)$$

Government budget constraint. The government budget constraint is given by

$$\bar{a}_c\int c_i di + \bar{a}_y\int wh_k\ell_i di = \bar{a}_n\int\frac{y_i^{\tau_n}}{(1+a_n)}p_{ui}di + \bar{G}$$

$$\bar{a}_c \int c_i di + \bar{a}_y (w\ell) \exp\left(m_h + \frac{\Sigma_h^2}{2}\right) = \bar{a}_n s(1-a_y)(w\ell)^{1-\tau_y} \exp\left((1-\tau_y)m_h + (1-\tau_y)^2 \frac{\Sigma_h^2}{2}\right) + \bar{G}$$

$$\bar{a}_c \int c_i di + \bar{a}_y (w\ell) \exp\left(m_h + \frac{\Sigma_h^2}{2}\right) = \bar{a}_n s(1-\bar{a}_y)(w\ell) \exp\left(m_h + \frac{\Sigma_h^2}{2}\right) + \bar{G}$$

We can finally express the consumption tax as:

$$\bar{a}_c = \frac{\bar{a}_n s(1-\bar{a}_y) + \bar{g} - \bar{a}_y}{(1-s)(1-\bar{a}_y)} \tag{89}$$

Law of Motion of Human Capital We are ready to derive the law of motion of human capital across households. Using the assumption of log-normality of both shocks, (7) and (8), the distribution of human capitals remains log-normally distributed across generations.

Proposition 6. *If* $\ln h_t \sim \mathcal{N}(m_{ht}, \Sigma_{ht}^2)$ *then*

$$\ln h_{t+1} \sim \mathcal{N}\left(m_{ht+1}, \Sigma_{ht+1}^2\right) \tag{90}$$

$$m_{ht+1} = \rho_t m_{ht} + X_{1t} \tag{91}$$

$$\Sigma_{ht+1}^2 = \rho_t^2 \Sigma_{ht}^2 + X_{2t} \tag{92}$$

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

$$X_{1t} = -\frac{\sigma_y^2}{2} - \alpha_h\left(\alpha_q \epsilon_{2t} + 1\right)\frac{\sigma_z^2}{2} + \alpha_q \epsilon_{1t} \ln\left(\frac{s_t(1+a_n)}{\underline{p}_t}(A\ell)^{(1-\tau_y)(1-\tau_n)}(1-a_y)^{(1-\tau_n)}\right)$$

$$X_{2t} = \sigma_y^2 + \left(\alpha_h\left(\alpha_q \epsilon_{2t} + 1\right)\right)^2 \sigma_z^2.$$

It is intuitive that the shifter X_{1t} in the law of motion of the mean of the distribution (91) is increasing in the saving rate s_n, in the average education subsidies a_n but decreasing in the intercept of the tuition schedule $\underline{p}_t$. From its expression in appendix, the latter is increasing in the share of resources devoted to research. The persistence coefficient ρ_t is decreasing in the progressivity of financial aid τ_n.

Law of motion for m_k, m_h We now express the laws of motion of the distribution of knowledge capitals and human capital in a compact format.[18]

$$m_{kt+1} = \gamma_{kkt} m_{kt} + \gamma_{kht} m_{ht} + \gamma_{kt} \tag{93}$$

$$m_{ht+1} = \gamma_{hht} m_{ht} + \gamma_{hkt} m_{kt} + \gamma_{ht} \tag{94}$$

with

$$\gamma_{kkt} = \gamma_K + (\gamma_e + \gamma_h)\tau_G \tag{95}$$

$$\gamma_{kht} = (\gamma_e + \gamma_h)(1 - \tau_y)(1 - \tau_n) \tag{96}$$

$$\gamma_{kt} = \ln \left(\frac{s_{eI}}{G}\right)^{\gamma_e} \left(\frac{s_{hI}}{AG}\right)^{\gamma_h} \tag{97}$$

$$+ (\gamma_e + \gamma_h) \left[\ln s(1 + a_n)(1 - a_y)^{1-\tau_n} (A\ell)^{(1-\tau_y)(1-\tau_n)} + \frac{((1 - \tau_y)(1 - \tau_n))^2 \sigma_{h|q}^2}{2} \right] \tag{98}$$

$$\gamma_{hht} = (1 - \tau_y)(1 - \tau_n)\alpha_q(\omega_e + \omega_h) + \alpha_h(1 + \alpha_q \omega_z) \tag{99}$$

$$\gamma_{hkt} = \alpha_q \omega_k \tag{100}$$

$$\gamma_{ht} = -\frac{\sigma_y^2}{2} - \alpha_h(1 + \alpha_q \omega_z)\frac{\sigma_z^2}{2} + \alpha_q(\omega_e + \omega_h) \ln \left(s(1 + a_n)(A\ell)^{(1-\tau_y)(1-\tau_n)}(1 - a_y)^{(1-\tau_n)} \right)$$

$$- \alpha_q \omega_{It}(\gamma_e + \gamma_h) \frac{((1 - \tau_y)(1 - \tau_n))^2 \sigma_{h|q}^2}{2} + \alpha_q \left[\ln (s_{eq})^{\omega_e} \left(\frac{s_{hq}}{A}\right)^{\omega_h} \right] \tag{101}$$

We now briefly give an intuition for each term from (95)-(101). Looking at equation (95), current average knowledge capital has a strong effect on future knowledge capital when knowledge depreciates slowly (low γ_K). Looking at equation (96), 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 (98), the growth of fundamental knowledge is high when the rate of cross-subsidization is high (s_{eI}, s_{hI}) or when households spend a large share of their income on tuition, s.

Looking at equation (99), current average human capital has a strong effect on future human capital when the transmission of of abilities from parents to children is strong (α_h), the peer-effect and the effect of teaching equipment and faculty is high ($\omega_z, \omega_e, \omega_h$). Looking at equation (100), current average knowledge capital has a strong effect on

[18] Recall that

$$m_{kt+1} = \ln \left(\frac{s_{eI}}{G}\right)^{\gamma_e} \left(\frac{s_{hI}}{AG}\right)^{\gamma_h} + (\gamma_e + \gamma_h) m_{Rt} + (\gamma_k + \tau_G(\gamma_e + \gamma_h)) m_{kt}$$

where m_R in the first line is given by equation (81).

future human capital when knowledge capital matters a lot for teaching equality ω_q. Finally, looking at equation (101), 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_{eq}, s_{hq}.

D.5 Additional derivations.

Recall that the variance of (log) quality is given by (78):

$$\Sigma_q^2 = \epsilon_A^2\sigma_z^2 + (\epsilon_I + \epsilon_A)^2\Sigma_h^2$$

From the definition of s_z:

$$s_z = 1 - \frac{(\epsilon_I + \epsilon_A)^2\Sigma_h^2}{\Sigma_q^2} \tag{102}$$

Using the definitions of χ_k and of s_z, we obtain

$$\Sigma_k = \chi_k\Sigma_q = \chi_k\frac{(\epsilon_I + \epsilon_A)\Sigma_h}{\sqrt{1-s_z}} \tag{103}$$

Using (103) as well as (72), (73), (75) and (76),

$$\begin{aligned}
\frac{1}{\chi_k} &= \frac{(\epsilon_I + \epsilon_A)\Sigma_h}{\sqrt{1-s_z}\Sigma_k} \\
&= \frac{((1-\tau_y)(1-\tau_n)\epsilon_1 + \alpha_z\epsilon_2)\Sigma_h}{\sqrt{1-s_z}\Sigma_k} \\
&= \frac{1}{\sqrt{1-s_z}\Sigma_k}\left[(1-\tau_y)(1-\tau_n)\frac{\omega_e+\omega_h+\omega_{It}\left(\gamma_e+\gamma_h\right)}{1+\chi_R\omega_{It}\left(\gamma_e+\gamma_h\right)-\chi_{k,t}\omega_k} + \alpha_z\frac{\omega_z}{1+\chi_R\omega_{It}\left(\gamma_e+\gamma_h\right)-\chi_{k,t}\omega_k}\right]\Sigma_h
\end{aligned}$$

Multiplying both sides by $1+\chi_R\omega_{It}\left(\gamma_e+\gamma_h\right)-\chi_{k,t}\omega_k$ we obtain

$$\begin{aligned}
\frac{1+\chi_R\omega_{It}\left(\gamma_e+\gamma_h\right)-\chi_{k,t}\omega_k}{\chi_k} &= \frac{1}{\sqrt{1-s_z}\Sigma_k}\left[(1-\tau_y)(1-\tau_n)(\omega_e+\omega_h+\omega_{It}\left(\gamma_e+\gamma_h\right)) + \alpha_z\omega_z\right]\Sigma_h \\
\iff \quad \frac{1+\chi_R\omega_{It}\left(\gamma_e+\gamma_h\right)}{\chi_k} &= \frac{1}{\sqrt{1-s_z}\Sigma_k}\left[(1-\tau_y)(1-\tau_n)(\omega_e+\omega_h+\omega_{It}\left(\gamma_e+\gamma_h\right)) + \alpha_z\omega_z\right]\Sigma_h + \omega_k
\end{aligned}$$

Recall from equation (85) that

$$\chi_R = (1-\tau_y)(1-\tau_n)\Sigma_h\sqrt{1-s_z} \iff \frac{\chi_R}{\chi_k} = \frac{(1-\tau_y)(1-\tau_n)}{\Sigma_k}\Sigma_h\sqrt{1-s_z}$$

Hence

$$
\begin{aligned}
\frac{1}{\chi_k} &= \frac{(1-\tau_y)(1-\tau_n)(\omega_e+\omega_h)\Sigma_h + \alpha_h\omega_z\Sigma_h}{\sqrt{1-s_z}\Sigma_k} + \omega_k \\
&+ \frac{(1-\tau_y)(1-\tau_n)\omega_I(\gamma_e+\gamma_h)}{\Sigma_k}\Sigma_h\left(\frac{1}{\sqrt{1-s_z}} - \sqrt{1-s_z}\right) \\
&= \frac{(1-\tau_y)(1-\tau_n)(\omega_e+\omega_h)\Sigma_h + \alpha_h\omega_z\Sigma_h}{\sqrt{1-s_z}\Sigma_k} + \omega_k + \frac{(1-\tau_y)(1-\tau_n)\omega_I(\gamma_e+\gamma_h)}{\sqrt{1-s_z}\Sigma_k}\Sigma_h s_z
\end{aligned}
$$

From (49) and the fact that $\chi_{h,I}/\chi_k = \chi_R/\chi_k + \tau_G$, $\chi_{e_k}/\chi_k = \chi_R/\chi_k + \tau_G$,

$$
\begin{aligned}
v_t = &\left[\frac{(1-\tau_y)(1-\tau_n)(\omega_e+\omega_h)\Sigma_h + \alpha_h\omega_z\Sigma_h}{\sqrt{1-s_z}\Sigma_k} + \omega_k + \frac{(1-\tau_y)(1-\tau_n)\omega_I(\gamma_e+\gamma_h)}{\sqrt{1-s_z}\Sigma_k}\Sigma_h s_z\right] \\
&\times \frac{1}{1-\beta\left(\gamma_k + (\gamma_e+\gamma_h)\left(\frac{(1-\tau_y)(1-\tau_n)}{\Sigma_k}\Sigma_h\sqrt{1-s_z} + \tau_G\right)\right)}
\end{aligned}
$$

Using $\omega_I = \beta v$ we can further simplify

$$
v = \frac{(\omega_e+\omega_h)\frac{(1-\tau_n)(1-\tau_y)\Sigma_h}{\sqrt{1-s_z}\Sigma_k} + \omega_z\alpha_h\frac{\Sigma_h}{\sqrt{1-s_z}\Sigma_k} + \omega_k}{1-\beta\left((\gamma_e+\gamma_h)\frac{(1-\tau_n)(1-\tau_y)\Sigma_h}{\sqrt{1-s_z}\Sigma_k} + \gamma_k + (\gamma_e+\gamma_h)\tau_G\right)} \tag{104}
$$

From the steady-state expression of (66) and (85), we get

$$
\Sigma_k = \frac{\gamma_e+\gamma_h}{1-\gamma_k-\tau_g(\gamma_e+\gamma_h)}\Sigma_R = \frac{\gamma_e+\gamma_h}{1-\gamma_k-\tau_g(\gamma_e+\gamma_h)}(1-\tau_y)(1-\tau_n)\sqrt{1-s_z}\Sigma_h \tag{105}
$$

Therefore

$$
\frac{\Sigma_h}{\Sigma_k} = \frac{(1-\gamma_k-\tau_G(\gamma_e+\gamma_h))}{(\gamma_e+\gamma_h)(1-\tau_y)(1-\tau_n)\sqrt{1-s_z}} \tag{106}
$$

D.6 Proof of simple model

In this section, we solve for the simple version of the model without peer-effect, without government research grants and in steady-state. Let's start with the law of motion of the variance of knowledge. Assuming that $\gamma_I = 1$ and $\gamma_e = 0$ gives

$$
\Sigma_{kt+1}^2 = \gamma_K^2\Sigma_{kt}^2 + \Sigma_{It}^2 + 2\gamma_K\Sigma_{kt}\Sigma_{It} \quad \text{with} \quad \Sigma_{It}^2 = \rho_{ht}^2\Sigma_{Rt}^2
$$

When there is no peer-effect, the variance of college income is equal to the variance of household income times the progressivity of the need-based aid $\Sigma_R = (1-\tau_n)\Sigma_h$. The steady-state of the law of motion of knowledge gives

$$(1-\gamma_K)(1+\gamma_K) = \gamma_h^2\left(\frac{(1-\tau_n)\Sigma_h}{\Sigma_k}\right)^2 + 2\gamma_K\gamma_h^2\left(\frac{(1-\tau_n)\Sigma_h}{\Sigma_k}\right)$$

The (positive) solution to the first quadratic equation is

$$\frac{1-\gamma_K}{\gamma_h} = \frac{(1-\tau_n)\Sigma_h}{\Sigma_k}$$

We now solve for χ_R, χ_k. First we use $\chi_{h,I} = [\chi_R + \tau_G \chi_k]$ and $\chi_{h,q} = \chi_R$ and the the assumptions $\tau_G = 0$ to obtain $\chi_R = \chi_h$. We then use the condition $\Sigma_R^2 = \left(\frac{\chi_R}{\chi_k}\right)^2 \Sigma_k^2$, to solve for χ_k/χ_R given the steady-state ratio $\frac{(1-\tau_n)\Sigma_h}{\Sigma_k}$. The production function of quality also implies $\chi_h = \frac{1-\chi_k\omega_k}{\omega_h}$. Putting everything together gives

$$\chi_k = \frac{1}{\omega_h \frac{(1-\tau_n)\Sigma_h}{\Sigma_k} + \omega_k} \qquad \chi_R = \frac{\frac{(1-\tau_n)\Sigma_h}{\Sigma_k}}{\left(\omega_h \frac{\Sigma_h}{\Sigma_k} + \omega_k\right)}$$

We now solve for v using its steady-state expression $v = \frac{1}{\chi_k\left(1-\beta\left(\gamma_h \frac{\chi_h}{\chi_k} + \gamma_k\right)\right)}$ which gives

$$v = \frac{\omega_h \frac{(1-\tau_n)\Sigma_h}{\Sigma_k} + \omega_k}{\left(1-\beta\left(\gamma_h \frac{(1-\tau_n)\Sigma_h}{\Sigma_k} + \gamma_k\right)\right)}$$

After substitution it gives

$$v = \frac{\frac{\omega_h}{\gamma_h}(1-\gamma_k) + \omega_k}{1-\beta}$$

Now that we have v, using $\omega_I = \beta v$, we can obtain s_{hI}:

$$s_{hI} = \frac{\omega_I \gamma_h}{\omega_I \gamma_h + \omega_h} = \frac{1}{1 + \frac{\omega_h}{\gamma_h \beta} \frac{1-\beta}{\frac{\omega_h}{\gamma_h}(1-\gamma_k) + \omega_k}}$$

We then solve for ϵ_1, ϵ_I:

$$\epsilon_1 = \omega_h + \omega_k \frac{\Sigma_k}{(1-\tau_n)\Sigma_h}$$
$$\epsilon_I = \epsilon_1(1-\tau_n)$$

In order to solve for Σ_h and Σ_k we use the law of motion of the variance of human capital in steady-state and then the expression for the ratio $(1-\tau_n)\Sigma_h/\Sigma_k$ to solve for Σ_k:

$$\Sigma_h^2 = \frac{\alpha_h^2 \sigma_z^2}{1 - \left(\alpha_h + \alpha_q(1-\tau_n)\left(\omega_h + \omega_k \frac{\gamma_h}{1-\gamma_k}\right)\right)^2} \qquad \Sigma_k^2 = \frac{\gamma_h^2}{(1-\gamma_k)^2}(1-\tau_n)^2\Sigma_h^2$$

We now solve for the households' optimal decisions. We solve for the marginal utility of human capital, u, and the spending rate on tuition, s:

$$u = \frac{(1-\beta)}{1-\beta\alpha_h} = \frac{(1-\beta)}{1-\beta\left(\alpha_h + \alpha_q(1-\tau_n)\left(\omega_h + \omega_k \frac{\gamma_h}{1-\gamma_k}\right)\right)}$$
$$s = \frac{\beta\alpha_q(1-\tau_n)}{1-\beta\alpha_h}\left(\omega_h + \omega_k \frac{\gamma_h}{1-\gamma_k}\right)$$

We now solve for the steady-state means of (log) human capital college knowledge:

$$m_{kt+1} = \gamma_K m_{kt} + \gamma_h \left[m_{Rt} + \ln s_{hI}\right]$$
$$m_{ht+1} = \alpha_h m_{ht} + \alpha_q \omega_h \left[m_{Rt} + \ln(1-s_{hI})\right] + \alpha_q \omega_k m_{kt} - \alpha_h \frac{\sigma_z^2}{2}$$

with $m_R = (1-\tau_n)\left[\ln w\ell + m_h\right] + \ln s(1+a_n)$. Using the expression for a_n given by (88), one gets:

$$m_R = \ln w\ell s(1+\bar{a}_n) + m_h + \tau_n(2-\tau_n)\frac{\Sigma_h^2}{2}$$

We now solve for the steady-state expressions of m_k and m_h:

$$m_h = \frac{a_h + \beta_h a_k}{1-\beta_h\beta_k} \quad \text{and} \quad m_k = \frac{a_k + \beta_k a_h}{1-\beta_h\beta_k}$$

with $\quad m_k = a_k + b_k m_h \quad$ and $\quad m_h = a_h + b_h m_k$

$$a_k = \frac{\gamma_h \left(\ln ws(1+\bar{a}_n)s_{hI} + \tau_n(2-\tau_n)\frac{\Sigma_h^2}{2} \right)}{1-\gamma_k} \quad \text{and} \quad \beta_k = \frac{\gamma_h}{1-\gamma_k}$$

$$a_h = \frac{\alpha_q \omega_h \left(\ln ws(1+\bar{a}_n)(1-s_{hI}) + \tau_n(2-\tau_n)\frac{\Sigma_h^2}{2} \right) - \alpha_h \frac{\sigma_z^2}{2}}{1-(\alpha_h + \alpha_q \omega_h)} \quad \text{and} \quad \beta_h = \frac{\alpha_q \omega_k}{1-(\alpha_h+\alpha_q\omega_h)}$$

After simplification, it is equal to

$$\begin{aligned}
m_h = & \frac{(1-\gamma_k)\left[\alpha_q \omega_h \left(\ln ws(1+\bar{a}_n)(1-s_{hI}) + \tau_n(2-\tau_n)\frac{\Sigma_h^2}{2} \right) - \alpha_h \frac{\sigma_z^2}{2} \right]}{(1-(\alpha_h+\alpha_q\omega_h))(1-\gamma_k) - \alpha_q\omega_k\gamma_h} \\
& + \frac{\alpha_q\omega_k\gamma_h \left(\ln ws(1+\bar{a}_n)s_{hI} + \tau_n(2-\tau_n)\frac{\Sigma_h^2}{2} \right)}{(1-(\alpha_h+\alpha_q\omega_h))(1-\gamma_k) - \alpha_q\omega_k\gamma_h} \\
m_k = & \frac{(1-(\alpha_h+\alpha_q\omega_h))\gamma_h \left(\ln ws(1+\bar{a}_n)s_{hI} + \tau_n(2-\tau_n)\frac{\Sigma_h^2}{2} \right)}{(1-(\alpha_h+\alpha_q s_{hI}\omega_h))(1-\gamma_k) - \alpha_q\omega_k\gamma_h} \\
& + \frac{\gamma_h \left[\alpha_q \omega_h \left(\ln ws(1+\bar{a}_n)(1-s_{hI}) + \tau_n(2-\tau_n)\frac{\Sigma_h^2}{2} \right) - \alpha_h \frac{\sigma_z^2}{2} \right]}{(1-(\alpha_h+\alpha_q s_{hI}\omega_h))(1-\gamma_k) - \alpha_q\omega_k\gamma_h}
\end{aligned}$$

Finally, we solve for the price schedule and the intercept of the tuition subsidy schedule:

$$p(q) = \underline{p} q^{1/\epsilon_1} \quad \text{with} \quad \ln \underline{p} \ = (\ln s(1+a_n)w + m_h)\left(1 - \frac{\omega_h}{\epsilon_1}\right) - \frac{1}{\epsilon_1}\left[\omega_h \ln(1-s_{hI}) + \omega_k m_k\right]$$