Supplemental Appendix

University Research and the Market for Higher Education

Titan Alon, Damien Capelle and Kazushige Matsuda

A Data Sources

A.1 Main Sample and Data Sources

Our primary sample includes all 4-year public and private non-profit institutions in the United States. Data primarily comes from the National Center for Education Statistics Integrated Postsecondary Education Data System (IPEDS) which provides university level microdata for the universe of domestically accredited higher-education institutions. We merge in additional data from the National Science FoundationâĂŹs Higher Education Research and Development Survey (HERD) on research spending by field and source of funds, the Association of University Technology Managers (AUTM) Patent Licensing Survey, and Web of Science (WoS) bibliometric data from the CWTS Leiden Rankings. The paragraphs below provide notes for each data figure in the paper. The sections that follow provide further details on each of the main data sources.

Figure 1. The figure reports total university research expenditures by source of funds in 2012 U.S. dollars. Government sources include research grants and contracts. Non-profit funding is included in the Other category. Underlying data are from the Higher Education Research and Development (HERD) survey, 1972-2018.

Figure 2. 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.

Figure 3. SAT scores are the sum of math and verbal scores calculated as the average of the university's reported 25^{th} and 75^{th} percentiles. Data comes from the Integrated Postsecondary Education Data System with points representing university level averages for 2012-2015.

Figure 4. Tuition is the average tuition revenue the university receives per full-time equivalent student, net of any university discounts or allowances. research expenditure is total university spending for activities specifically organized to produce research outcomes, including by institutes, research centers, and individuals. Data retrieved the Integrated Postsecondary education Data System (IPEDS). Points are 2012-2015 university averages in log-scale.

Figure 5. This figure plots a histogram of 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. Underlying data come from the NCES IPEDS.

A.2 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.

A.3 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. Before 2010, it includes only institutions with degree programs in science and engineering (S&E).

We use the HERD survey primarily to disaggregate university research by source (i.e. government, internal, business) and by the type of expenditure (i.e. equipment or salaries). We also use HERD to construct the instruments for our NIH regressions. In particular, we take microdata on university R&D spending by source and field to calculate the university share of federal life science funding before the policy change.

A.4 Association of University Technology Managers Patent Licensing Survey

The Association of University Technology Managers (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.

A.5 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.

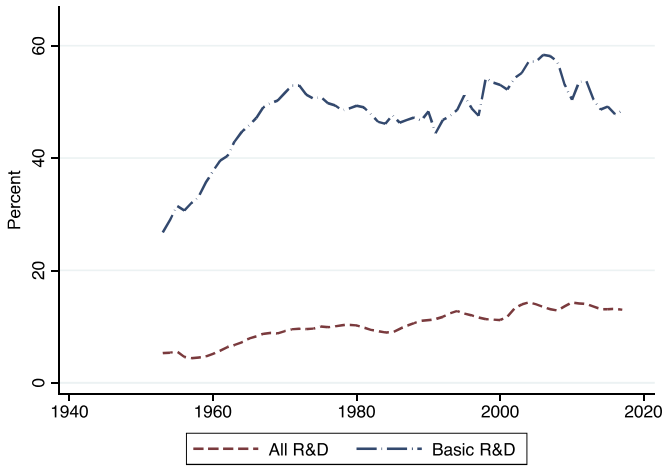
We use the bibliometric micro data underlying the Leiden Rankings to measure university publications and citations.

A.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.

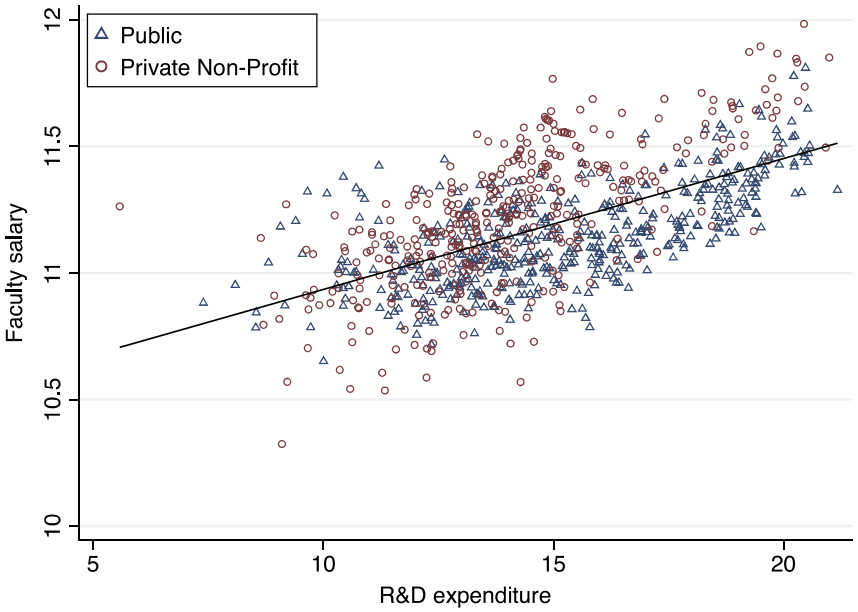
B Additional Figures

Figure B1: Higher education research expenditure as share of national total



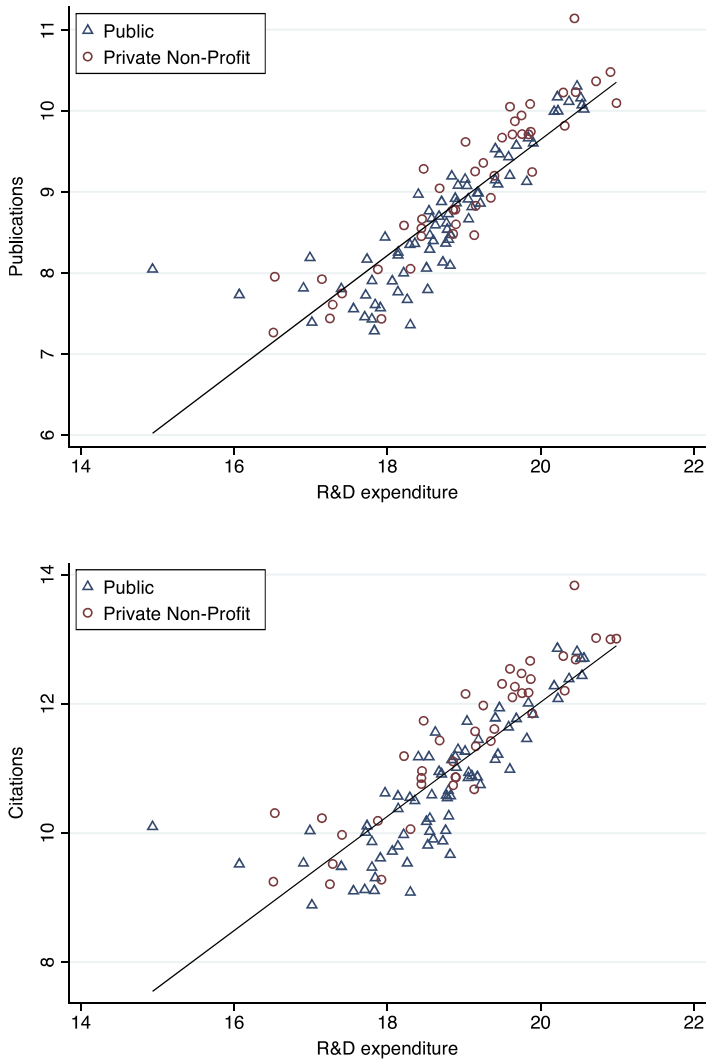
Notes: *Y-axis represents the higher education sector's share of total domestic research and development expenditures, by type. Underlying data come from National Science Board (2018).*

Figure B2: University Research Spending and Faculty Compensation



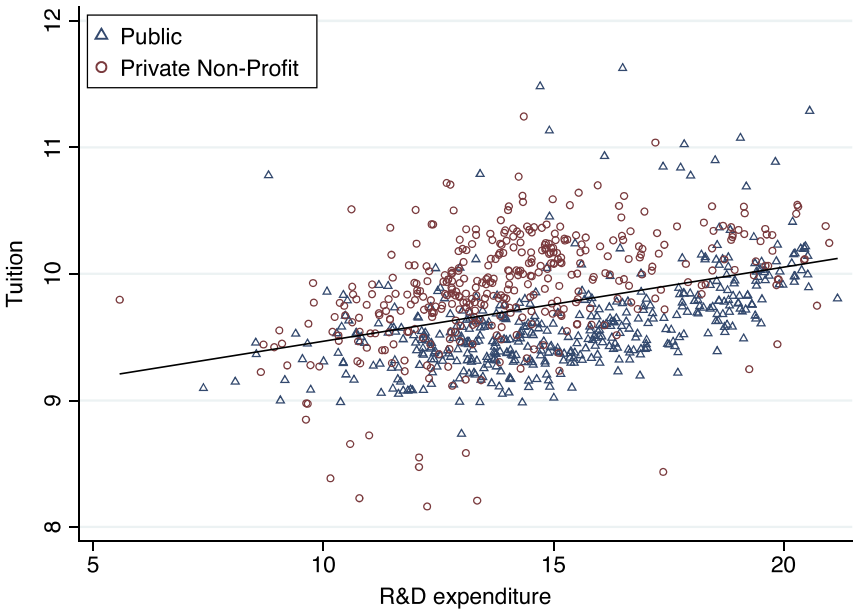
Note: *Faculty salary is the average salary for full-time faculty members on 9-month equated contracts. Research expenditure is total university spending for activities specifically organized to produce research outcomes, including by institutes, research centers, and individuals. Data comes from the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

Figure B3: University Research Spending and Knowledge Creation



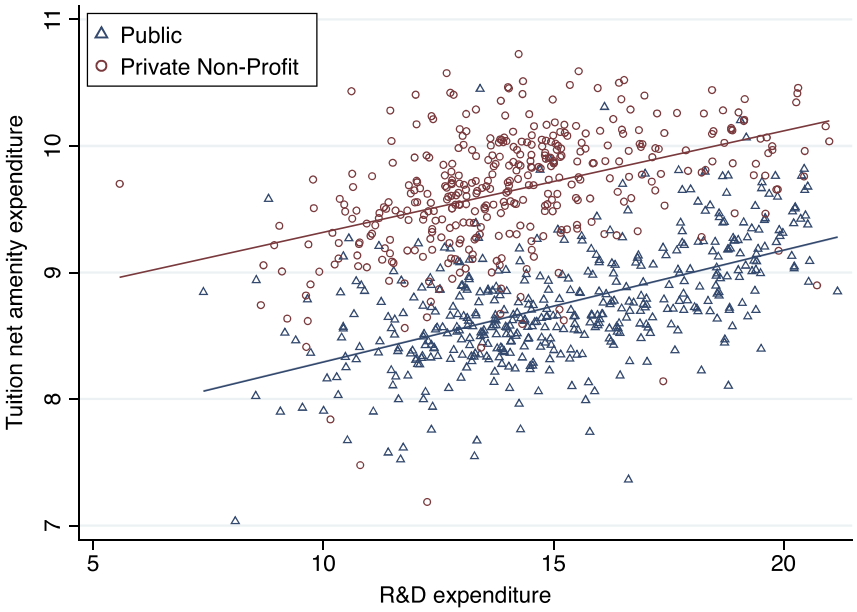
Note: Publication and citation data come from the CWTS Leiden Ranking derived from the core collection of the Web of Science (WoS) for the years 2015-2018. Research expenditure is total university spending for activities specifically organized to produce research outcomes, including by institutes, research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

Figure B4: University research spending and tuition with state and local appropriations



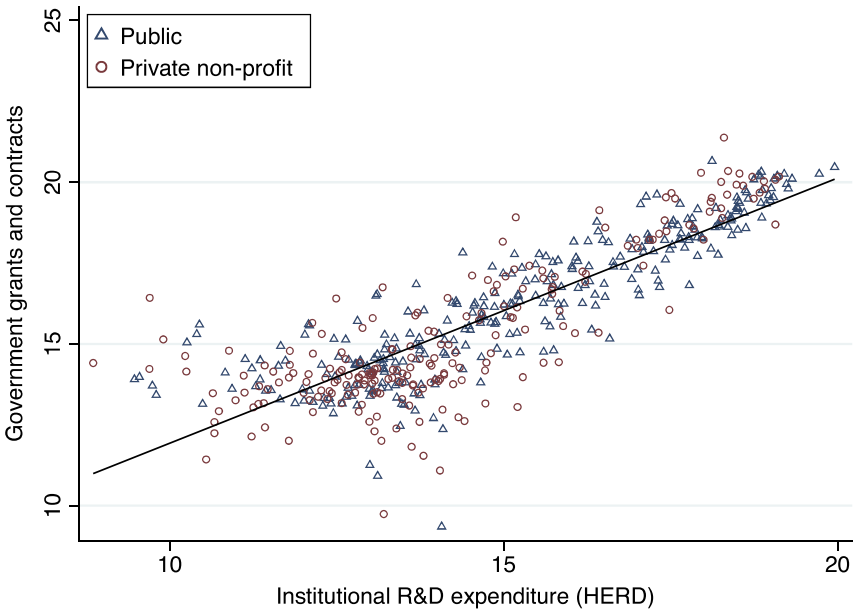
Notes: *Tuition is the average tuition revenue the university receives per full-time equivalent student, net of any university discounts or allowances, but including state and local appropriations per capita. Research expenditure is total university spending for activities specifically organized to produce 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 B5: University R&D spending and tuition net of student services expenditures, by sector



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

Figure B6: University Research Spending and Government Grants



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

B.1 Model Implications

Figure B7: Quality attended and tuition paid in equilibrium by family income and child ability

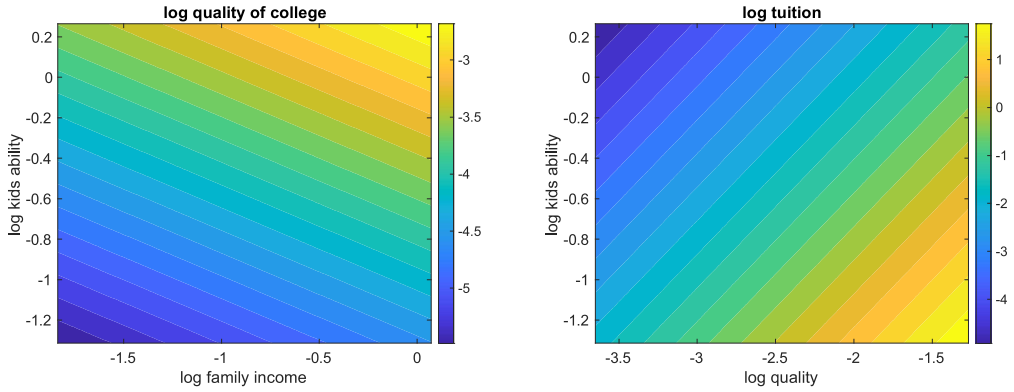
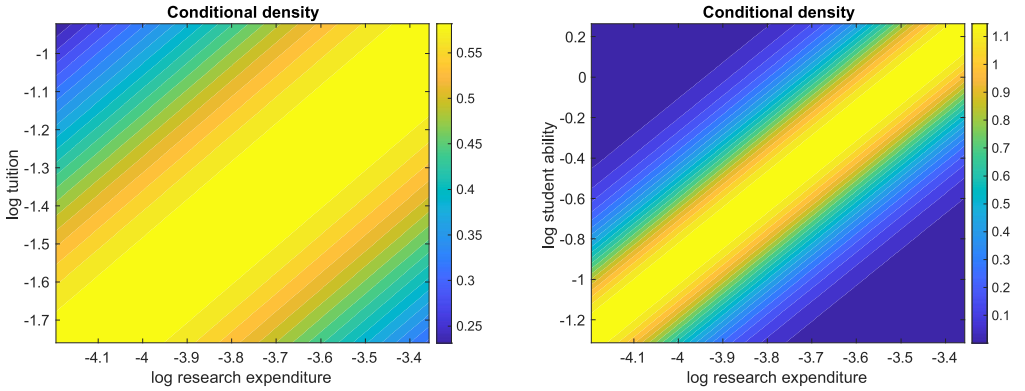


Figure B8: Within-college density of tuition paid and student ability (colleges ranked by their per-student research expenditure)



(a) Density of tuition

(b) Density of student ability

C Additional Tables

Table C1: Characteristics of top 25 research universities, by total research spending

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

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

D Additional Model Properties

D.1 Definition of Market Clearing and Equilibrium

We focus on a steady-state equilibrium in which all cross-sectional distributions are time invariant and prices are constant. An equilibrium consists of a household value function $U(h,z)$ and associated policy functions $\{c(h,z), q(h,z)\}$; the university value function $V(k)$ and associated university policy functions $\{\phi(z|k),\, e_R(k),\, e_T(k),\, q(k)\}$; and an equilibrium tuition price schedule $p(q,z)$ and measures $\mu_h(h,z)$ and $\mu_c(k)$ such that, given the tuition price schedule, $\{U(h,z), c(h,z), q(h,z)\}$ solves the household problem in (2), $\{V(k), \phi(z|k), e_R(k), e_T(k)\}$ solves the university problem in (8), and markets clear for all combinations of quality $Q \subset \mathbb{R}^+$ and student ability $Z \subset \mathbb{R}^+$ so that

$$\int \mathbb{1}\left[q(h,z) \in Q\right] \mathbb{1}\left[z \in Z\right] d\mu_h = \int \left(\int_{z \in Z} d\phi(z|k) \right) \mathbb{1}\left[q(k) \in Q\right] d\mu_c \quad (16)$$

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

D.2 Equilibrium college quality-ladder

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

Proposition 5. *The equilibrium college quality-ladder depends on the market struc-*

ture of the higher education sector and can be expressed as

$$\frac{\Sigma_q}{\Sigma_k} = \sqrt{\epsilon_1^2 \left(\frac{\Sigma_h}{\Sigma_k}\right)^2 + \epsilon_2^2 \left(\frac{\sigma_z}{\Sigma_k}\right)^2} \tag{17}$$

The first term in equation (17) captures the dispersion in household expenditures on education and, through ϵ_1, the extent to which they accrue to colleges of different qualities. The second term measures the variation in student abilities and, through ϵ_2, the extent to high ability students congregate at high quality colleges. University research spending is higher when household education expenditures are more unequal (e.g. high Σ_h) or if there is large variation in the abilities of students (e.g. high σ_z). Conversely, it is low when colleges' intangible capital are ex-ante highly unequal (e.g. high Σ_k), since universities are farther from their competing institutions and so overtaking them in the hierarchy of colleges would require larger and more costly investments in research. Proposition 5 also shows that universities spend more on research when the elasticities of education quality to tuition revenue and ability are high (e.g. high ϵ_1 and ϵ_2).

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

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

Similarly, Proposition 4 shows that the distribution of education quality

supplied is,

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

Market clearing in the higher education sector (equation (16)) requires that the distribution of education demand, Σ_q^D, equals the distribution of education supplied, Σ_q^S, so that

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

which elucidates the two equivalent characterizations of the college quality-ladder in Propositions 4 and 5. The parameters ϵ_1 and ϵ_2, which shape the sorting rule and tuition schedule, are determined in equilibrium to balance supply and demand and clear education markets. More specifically, parameter ϵ_1 adjusts to clear the supply and demand for education in the above equation, while ϵ_2 is simultaneously determined by the marginal rate of substitution between student ability z and monetary inputs, given by

$$\frac{\epsilon_1}{\epsilon_2} = \frac{\omega_e + \beta(1-\beta)^{-1}\gamma_e(\Sigma_q/\Sigma_k)}{\omega_z}. \tag{18}$$

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

D.3 An Equilibrium without Peer Effects

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

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

$$\epsilon_1 = \omega_e + \omega_k \frac{\Sigma_k}{\Sigma_h} \tag{19}$$

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

$$\Sigma_k = \frac{\gamma_e}{1-\gamma_k}\Sigma_h \tag{20}$$

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

$$\Sigma_h^2 = \frac{\sigma_z^2}{1 - \left(\alpha\left(\omega_e + \omega_k \frac{\gamma_e}{1-\gamma_k}\right)\right)^2} \tag{21}$$

In the case without peer-effect, knowledge intangible capital and final goods are the only two inputs to produce education quality. Universities have no incentives to attract talented students and as a result there is perfect sorting across colleges based on family income ($\epsilon_2 = 0$).[31]

These expressions show important equilibrium interactions between the market for higher education and university research. For instance, consider a change in the research technology, such as a rise in γ_k or γ_e, which increases R&D expenditure and leads to an increase in the dispersion of college intangible capital Σ_k in equation (20). From equation (19), we see that the increased dispersion in college intangible capital will decrease the price elasticity of demand, leading to an increase in household education expenditures and greater stratification of students by family background in higher education. In the long-run, the increase in university research spending leads to greater inequality in educational outcomes and more earnings inequality, steepening the college quality-ladder further and reinforcing university incentives to spend on research.

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

This example illustrates the role of universities as engines of human capital accumulation and innovation, but also of inequality. Equation (21) shows how the education system amplifies initial differences between students, σ_z^2, leading to larger variation in adult human capital. The equation also demonstrates how this amplification depends on the returns to education through α, differences in educational expenditures per students through ω_e, and the contribution of research through $\omega_k \frac{\gamma_e}{1-\gamma_k}$. The final term captures an important equilibrium feedback in the model whereby university research today also influences the future demand-side of the education market.

E Quantitative Extensions and Government Policies

E.1 Quantitative Extensions

Faculty in Teaching and Research. We generalize the university's research and teaching technologies to account for the contribution of faculty. Specifically, the university teaching technology (5) becomes $q = k^{\omega_k} \bar{z}^{\omega_z} e_T^{\omega_e} \bar{h}_T^{\omega_h}$ and the research technology (6) becomes $k' = k^{\gamma_k} e_R^{\gamma_e} \bar{h}_R^{\gamma_h}$, where the contribution of faculty depends on their average human capital, $\bar{h}_x = \mathbb{E}_{\mu_x(.)}[h]$, where $\mu_x(\cdot)$ is the endogenous distribution of faculty chosen by the university to perform task $x \in \{R, T\}$. Allowing the university to choose different compositions of research faculty, $\mu_R(\cdot)$, and teaching faculty, $\mu_T(\cdot)$, captures the fact that they can partially specialize these tasks internally by hiring dedicated teaching faculty or by increasing teaching loads for research faculty. The inclusion of faculty requires an additional labor market clearing condition for labor across the education, research, and production sectors.[32]

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

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

ulty across the college quality-ladder, as summarized in figure B2. Moreover, since faculty are drawn from the adult population, the distribution of faculty human capital is endogenous to the model and depends on Σ_h^2.

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

Household Demographics and Intergenerational Dynamics. We introduce a more general process to determine a child's ability at the start of college, z. Following Capelle and Matsuda (2025), we model z as the result of an intergenerational process given by,

$$z = \xi h^{\varphi} \tag{22}$$

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

shocks to the higher education system.

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

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

E.2 Government Tuition and Research Policies

The government implements two types of policies in the higher education sector: merit-based research grants and need-based student financial aid. Government policies are funded by progressive income taxes, as in Heathcote, Storesletten, and Violante (2017), with excess revenues rebated to households through a linear non-distortive consumption rebate.[33] We take the prevailing tax schemes as given and do not consider the optimal design of government R&D taxation policies within our counterfactuals, as in Akcigit, Hanley, and Stantcheva (2022).

Tuition Policies. Federal tuition policies consist of progressive need-based financial aid. We augment the household budget constraint in (3) so that the out-of-pocket college expense for a household with income y is given

[33] Specifically, $\{a_y, \tau_y\}$ parameterizes the tax-system such that after-tax income is $(1 - a_y) \times (wh)^{1-\tau_y}$.

by $\psi(y) \times p(q, z)$, where $\psi(y)$ represents the government tuition subsidy. While such aid is, in practice, distributed through a variety of policy instruments, we follow the parsimonious approach of Benabou (2002) and model the net effect of these policies in reduced-form using the two-parameter policy schedule

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

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

Research Policies. As with tuition, government subsidization of university research is administered through several different programs and agencies, including the National Institute of Health, the Department of Defense, NASA, the National Science Foundation, and others.[34] As above, we model these programs parsimoniously through a reduced-form allocation rule for government grants that captures both the level of subsidy and its distribution across institutions. Specifically, we augment the university budget constraint (9) so that government grants cover a fraction $1-G(k)$ of a university's research (but not teaching) expenditures. The dependence of research subsidies on k reflects the meritocratic nature of government grant making and allows us to match the distribution of federal research funds observed in Figure B6. We parameterize the government's grant policy schedule using the two-parameter family

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

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

[34] See National Science Board (2018), *Expenditures and Funding for Academic R&D*.

F Proofs of Quantitative Model

F.1 University problem

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

$$\log R(q) = m_R + \chi_R(\log q - m_q) \tag{25}$$
$$\log k(q) = m_k + \chi_k(\log q - m_q) \tag{26}$$
$$\log \bar{h}_R(q) = m_{hR} + \chi_{hR}(\log q - m_q) \tag{27}$$
$$\log \bar{h}_T(q) = m_{hT} + \chi_{hT}(\log q - m_q) \tag{28}$$
$$\log \bar{z}(q) = m_z + \chi_z(\log q - m_q) \tag{29}$$
$$\log \bar{e}_R(q) = m_{eR} + \chi_{eR}(\log q - m_q) \tag{30}$$
$$\log \bar{e}_T(q) = m_{eT} + \chi_{eT}(\log q - m_q) \tag{31}$$

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

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

$$V_t(k) = \bar{v}_t + v_t \ln k \tag{32}$$

Replacing this guess into the expression for the value function and using our guesses (26)-(28) gives

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

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

$$
v_t = \frac{1}{\chi_{kt}} + \beta_c \left(\gamma_k + \gamma_e \frac{\chi_{eRt}}{\chi_{kt}} + \gamma_h \frac{\chi_{hRt}}{\chi_{kt}}\right) v_{t+1} \tag{33}
$$

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

$$
v = \frac{1}{\chi_k} \frac{1}{1 - \beta_c \left(\gamma_k + \gamma_e \frac{\chi_{eR}}{\chi_k} + \gamma_h \frac{\chi_{hR}}{\chi_k}\right)} \tag{34}
$$

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

$$
\max_{q, \phi(\cdot), e_R, e_T, \mu_R(\cdot), \mu_T(\cdot)} \ln q + \beta_c v_{t+1} \ln k' \tag{35}
$$

$$
\text{subject to} \quad \ln q = \ln \bar{h}_T^{\omega_h} e_T^{\omega_e} \bar{z}^{\omega_z} k^{\omega_k} \tag{36}
$$

$$
k' = k^{\gamma_k} e_R^{\gamma_e} \bar{h}_R^{\gamma_h} \tag{37}
$$

$$\mathop{\mathbb{E}}_{\phi(.)}[p(q,z)] = G(k)\left[e_R + \int whd\mu_R(h)\right] + e_T + \int whd\mu_T(h) \tag{38}$$

$$\bar{h}_R = \int hd\mu_R(h) \tag{39}$$

$$\bar{h}_T = \int hd\mu_T(h) \tag{40}$$

$$\ln \bar{z}(\phi; p) = \mathbb{E}_{\phi(.)}[\ln(z)] - \sigma_u^2(\phi; p). \tag{41}$$

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

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

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

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

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

We verify later that the guess that tuition are log-normally distributed within a college is true (see equation (72).

The last step before taking the first order conditions is to substitute the peer-effect (41) into the expression for teaching quality and to specify the expression for the cost of heterogeneity σ_u^2.

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

$$\sigma_u^2(\phi; p) = \frac{\Omega_t}{2} V_{\phi(.)}(\ln p(q,z)) \tag{42}$$

$$with \quad \Omega_t = \frac{\omega_e + \omega_h + \beta v_{t+1}(\gamma_e + \gamma_h)}{\omega_z} \tag{43}$$

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

Using the expression for the cost of heterogeneity (43), the definition of the peer-effect (41) and the guess that tuition are log-normally distributed, the college problem becomes fully log-linear in tuition and student' ability

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

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

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

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

and hence that tuition be equal to

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

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

tion are highly valued by colleges, v_{t+1}. In other words, research increases the valuation of financial resources relative to student ability.

We now consider the optimal choice of spending on research and teaching equipment and faculty wages. Taking tuition revenues R as given, we take the F.O.C. w.r.t. s_{eR}, s_{eT}, s_{hT} and s_{hR}:[35]

$$s_{eR} = \frac{\beta v_{t+1}\gamma_e}{\beta v_{t+1}(\gamma_e+\gamma_h)+\omega_e+\omega_h} \tag{44}$$

$$s_{eT} = \frac{\omega_e}{\beta v_{t+1}(\gamma_e+\gamma_h)+\omega_e+\omega_h} \tag{45}$$

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

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

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

Starting from the expression for tuition we just derived and using the definition of teaching quality (36), together with our guesses (25)-(31), one gets

$$\begin{aligned}
p(q,z) &= R\left(\frac{z}{\bar{z}}\right)^{-\frac{\omega_z}{\omega_e+\omega_h+\beta_c v(\gamma_e+\gamma_h)}} \\
&= R\left[\bar{h}_T^{\omega_h} e_T^{\omega_e} \bar{z}^{\omega_z} k^{\omega_k}\right]^{\frac{1}{\omega_e+\omega_h+\beta_c v(\gamma_e+\gamma_h)}} \left(\bar{h}_T^{\omega_h} e_T^{\omega_e} k^{\omega_k}\right)^{\frac{-1}{\omega_e+\omega_h+\beta_c v(\gamma_e+\gamma_h)}} z^{-\frac{\omega_z}{\omega_e+\omega_h+\beta_c v(\gamma_e+\gamma_h)}} \\
&= \bar{p} q^{\frac{1}{\omega_e+\omega_h+\beta_c v(\gamma_e+\rho_h)}+\chi_R-\frac{(\omega_e+\omega_h)\chi_R}{\omega_e+\omega_h+\beta_c v(\gamma_e+\gamma_h)}-\chi_k\frac{\omega_k}{\omega_e+\omega_h+\beta_c v(\gamma_e+\gamma_h)}} z^{-\frac{\omega_z}{\omega_e+\omega_h+\beta_c v(\gamma_e+\gamma_h)}} \\
&= \bar{p} q^{\frac{1}{\epsilon_1}} z^{-\frac{\epsilon_2}{\epsilon_1}}
\end{aligned} \tag{48}$$

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

where we have defined the following aggregate endogenous variables:

$$\log \bar{p} = \frac{\left(\beta_c v(\gamma_e+\gamma_h) m_R - \omega_k m_k - \left(\frac{\omega_e+\omega_h+\beta_c v(\gamma_e+\gamma_h)}{\epsilon_1} - 1\right) m_q - \log\left(s_{eT}\right)^{\omega_e}\left(\frac{s_{hT}}{w}\right)^{\omega_h}\right)}{\omega_e+\omega_h+\beta_c v\left(\gamma_e+\gamma_h\right)} \tag{49}$$

$$\epsilon_1 = \frac{\omega_e+\omega_h+\beta_c v\left(\gamma_e+\gamma_h\right)}{1+\chi_R\beta_c v\left(\gamma_e+\gamma_h\right)-\chi_k\omega_k} \tag{50}$$

$$\epsilon_2 = \frac{\omega_z}{1+\chi_R\beta_c v\left(\gamma_e+\gamma_h\right)-\chi_k\omega_k} \tag{51}$$

where we use $w\bar{h}_T = s_{hT}R$, $e_T = s_{eT}R$, (25), and (26).

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

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

Using equation (10), the household problem is given by

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

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

$$\ln q = \epsilon_1 \ln(swh) - \epsilon_1 \ln \bar{p} + \epsilon_2 \ln z \tag{52}$$

and combining with the guess on the value functions gives

$$\begin{aligned} \bar{u} + u\ln h + u_z \ln z &= \ln((1-s)wh) + \beta(\bar{u} + u\ln h' + u_z \mathbb{E}\ln z') \\ &= \ln((1-s)wh) + \beta(\bar{u} + u\ln z + u\alpha \ln q + u_z \mathbb{E}\ln z') \end{aligned}$$

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

where we use $\ln z' \sim (-\sigma_z^2/2, \sigma_z^2)$ in the last term.

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

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

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

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

and thus

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

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

Finally, using the expression for u, we obtain

$$s = \frac{\beta u \alpha \epsilon_1}{1 + \beta u \alpha \epsilon_1} = \beta \alpha \epsilon_1. \tag{53}$$

F.4 Household problem in the quantitative model

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

The household problem can be formulated recursively as

$$U(h,z) = \max_{c_a,q} \left\{ \left[\sum_{a=1}^{10} \delta^{a-1} \ln c_a \right] + \beta \mathbb{E}\left[U(h',z') \right] \right\} \tag{54}$$

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

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

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

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

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

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

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

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

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

$$U(h,z) = \max_{c,q} \left\{ \ln c \sum_{a=1}^{10} \delta^{a-1} + \beta \mathbb{E}\left[U(h',z') \right] \right\} \tag{55}$$

$$\text{s.t.} \quad y = \delta^{-9}(1+a_{ct})c\sum_{a=1}^{10}\delta^{a-1} + \frac{y^{\tau_n}}{1+a_n}p_{t+10}(q,z) \tag{56}$$

For the next step it is useful to first derive the sorting rule (12). We combine the expression for tuition (48) and financial aid (23) and we denote s the share of net income that households spend on tuition $\bar{p}q^{\frac{1}{\epsilon_1}}z^{-\frac{\epsilon_2}{\epsilon_1}} = s(1+a_n)y^{1-\tau_n}$. Hence, the sorting rule is given by

$$q = \left(\frac{s(1+a_n)y^{1-\tau_n}}{\bar{p}}\right)^{\epsilon_1} z^{\epsilon_2} \tag{57}$$

We guess that the value function is a log-linear function of h and z: $U(h,z) = \tilde{U}(h,\xi) = u_t \ln h_t + u_{zt}\ln\xi + \bar{u}_t$ with $u_t, u_{zt}, \bar{u}_t$ three endogenous and aggregate variables. Recall that ξ denotes the shock to the transmission of human capital from parents to child given in equation (22). Like in section (F.3), we use this guess to substitute for the value function in the current and future period in equation (55) and we find that a necessary condition is that u_t obeys the following forward difference equation:

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

where we have used the sorting rule (57).

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

$$s_t = \frac{\beta\alpha\epsilon_{1t}(1-\tau_n)u_{t+10}}{\sum_{a=1}^{10}\delta^{a-1} + \beta\alpha\epsilon_{1t}(1-\tau_n)u_{t+10}} \tag{58}$$

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

One can easily check that the expression for s obtained in equation (53) cor-

responds to the special case with $\phi = \tau_n = \tau_y = 0$.

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

$$\epsilon_I \equiv \epsilon_1(1-\tau_n)(1-\tau_y) \tag{59}$$

$$\epsilon_A \equiv \varphi\epsilon_2 \tag{60}$$

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

Taking the log of the sorting rule (57) and using the previous definitions (59), (60) and ias well as the transmission of human capital over generation $z = \xi h^{\varphi}$, the optimal college is given by

$$\ln q = \epsilon_1(C_h - \ln \bar{p}) + (\epsilon_I + \epsilon_A)\ln h + \epsilon_2 \ln \xi \tag{61}$$

$$\text{with} \quad C_h \equiv \ln\left(s(1+a_n)\tilde{w}^{(1-\tau_y)(1-\tau_n)}(1-a_y)^{(1-\tau_n)}\right). \tag{62}$$

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

Replacing this expression in the law of accumulation of human capital given by (4) we obtain

$$\ln h' = \alpha(\epsilon_1(C_h - \ln \bar{p}) + (\epsilon_I + \epsilon_A)\ln h + \epsilon_2 \ln \xi) + \varphi \ln h + \ln \xi. \tag{63}$$

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

Law of motion of the distribution of human capital We denote m_{hta} and Σ_{hta} the mean and standard deviation of the log of human capital of adults with age a at time t. We denote m_{ht}, Σ_{ht} the mean and standard deviation

of the log of human capital of the parents who are currently sending their child to college, namely those with $a = 10$:

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

From these definitions and the law of motion of human capital at the individual level (63) we have that the law of motion of the distribution of human capital across households is given by

$$\ln h_{t+1,a} \sim \mathcal{N}\left(m_{h,t+1,a+1}, \Sigma^2_{ht+1,a+1}\right) \tag{64}$$
$$m_{h,t+1,a+1} = m_{h,t,a} \quad \text{for} \quad a < 10 \tag{65}$$
$$\Sigma_{h,t+1,a+1} = \Sigma_{h,t,a} \quad \text{for} \quad a < 10 \tag{66}$$
$$m_{h,t+1,1} = \rho_t m_{ht} - (\alpha\epsilon_{2t} + 1)\frac{\sigma_z^2}{2} + \alpha\epsilon_{1t}(C_{ht} - \ln \bar{p}_t) \quad \text{for} \quad a = 10 \tag{67}$$
$$\Sigma^2_{h,t+1,1} = \rho_t^2 \Sigma^2_{ht} + \sigma_y^2 + \left(\varphi\left(\alpha\epsilon_{2t} + 1\right)\right)^2 \sigma_z^2 \quad \text{for} \quad a = 10 \tag{68}$$
$$\text{where } \rho_t = \varphi + \alpha\left[\epsilon_{2t}\varphi + \epsilon_{1t}(1 - \tau_n)(1 - \tau_y)\right]$$

It is intuitive that the shifter C_{ht} in the law of motion of the mean of the distribution (67) is increasing in the saving rate s_t, in the average education subsidies a_n but decreasing in the intercept of the tuition schedule $\bar{p}_t$. From the expression given by (49), the latter is increasing in the share of resources devoted to research. Finally, the persistence coefficient ρ_t is decreasing in the progressivity of financial aid τ_n.

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

Distribution of college quality. From (61), it follows that the distribution of college quality is given by

$$\ln q \sim \mathcal{N}\left((\epsilon_I + \epsilon_A)m_h + \epsilon_1(C_h - \log \bar{p}) - \epsilon_2 \frac{\sigma_z^2}{2}, \epsilon_2^2\sigma_z^2 + (\epsilon_I + \epsilon_A)^2\Sigma_h^2\right) \quad (69)$$

Thus

$$m_q = (\epsilon_I + \epsilon_A)m_h + \epsilon_1(C_h - \log \bar{p}) - \epsilon_2 \frac{\sigma_z^2}{2}. \quad (70)$$

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

Within college parental income distribution. Using (61), we now solve for the conditional distribution of parental human capital within a college, which is given by

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

where

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

$$\text{with} \quad s_z = \frac{\epsilon_2^2\sigma_z^2}{(\epsilon_I + \epsilon_A)^2\Sigma_h^2 + \epsilon_2^2\sigma_z^2} \quad (71)$$

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

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

Notice that

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

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

$$\ln z = \frac{1}{\epsilon_2} \left(\ln q - \epsilon_I \ln h - \epsilon_1 (C_h - \log \bar{p}) \right) \tag{72}$$

$$\Rightarrow \ln z | q \sim \mathcal{N} \left(\frac{1}{\epsilon_2} \left(\ln q - \epsilon_I m_{h|q} - \epsilon_1 (C_h - \log \bar{p}) \right), \left(\frac{\epsilon_I}{\epsilon_2} \right)^2 \sigma^2_{h|q} \right) \tag{73}$$

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

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

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

This also verifies our guess (25). Finally these results enables us to get an expression for the distribution of revenue per student across colleges:

$$\ln R \sim \mathcal{N} \left(m_R, \Sigma_R^2 \right) \tag{74}$$

$$m_R = C_h + \frac{((1-\tau_y)(1-\tau_n))^2 \sigma^2_{h|q}}{2} + (1-\tau_y)(1-\tau_n) m_h \tag{75}$$

$$\Sigma_R^2 = \chi_R^2 \left[\epsilon_2^2 \sigma_z^2 + (\epsilon_I + \epsilon_A)^2 \Sigma_h^2\right] \tag{76}$$

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

Identifying coefficients with the guess, one gets:

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

Combining this with equations (71) and (76),

$$\Sigma_R = (1-\tau_y)(1-\tau_n)\sqrt{1-s_z}\Sigma_h. \tag{77}$$

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

Verification of guesses. From the definition of $\ln \bar{z}$ and equation (73),

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

Thus (29) is verified. From the technology for college quality

$$\ln q = \ln\left(R s_{hT}/w\right)^{\omega_h} \left(R s_{eT}\right)^{\omega_e} \bar{z}^{\omega_z} k^{\omega_k},$$

equation (29) and (25), our guess (26) is also verified. From $s_{eR} = G(k)e_R/R$, $s_{eT} = e_T/R$, $s_{hR} = G(k)w\bar{h}_R/R$, $s_{hT} = w\bar{h}_T/R$, equations (27), (28), (30), (31) are also verified. This verifies that all college-specific variables are log-linear functions of quality q and log-normally distributed.

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

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

We denote m_R and Σ_R (m_k and Σ_k) the mean and standard deviation of log-income (log-knowledge capital) across colleges. Taking the mean and variance of the law of motion above gives that if knowledge capital is log-normally distributed,

then it remains log-normally distributed and the law of motion of the mean and the variance of the associated normal distribution are given by the following proposition.

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

$$\ln k_{it} \sim \mathcal{N}\left(m_{kt}, \Sigma_{kt}^2\right) \tag{78}$$

$$m_{kt+1} = \ln \left(\frac{s_{eRt}}{G_t}\right)^{\gamma_e} \left(\frac{s_{hRt}}{w_t G_t}\right)^{\gamma_h} + (\gamma_e + \gamma_h) m_{Rt} + (\gamma_k + \tau_G(\gamma_e + \gamma_h)) m_{kt} \tag{79}$$

$$\Sigma_{kt+1} = (\gamma_k + \tau_G(\gamma_e + \gamma_h))\Sigma_{kt} + (\gamma_e + \gamma_h)\Sigma_{Rt} \tag{80}$$

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

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

We now express the laws of motion of the distribution of knowledge capitals and human capital in a compact format.[36]

$$m_{kt+1} = \gamma_{kkt} m_{kt} + \gamma_{kht} m_{ht} + \gamma_{kt} \tag{81}$$

$$m_{h,t+1,1} = \gamma_{hht} m_{ht} + \gamma_{hkt} m_{kt} + \gamma_{ht} \quad \text{for} \quad a = 10 \tag{82}$$

$$m_{h,t+1,a+1} = m_{h,t,a} \quad \text{for} \quad a < 10 \tag{83}$$

with

$$\gamma_{kkt} = \gamma_K + (\gamma_e + \gamma_h)\tau_G \tag{84}$$

$$\gamma_{kht} = (\gamma_e + \gamma_h)(1-\tau_y)(1-\tau_n) \tag{85}$$

$$\gamma_{kt} = \ln\left(\frac{s_{eR}}{G}\right)^{\gamma_e}\left(\frac{s_{hR}}{wG}\right)^{\gamma_h} + (\gamma_e+\gamma_h)\left[C_h + \frac{((1-\tau_y)(1-\tau_n))^2\sigma^2_{h|q}}{2}\right] \tag{86}$$

$$\gamma_{hht} = (1-\tau_y)(1-\tau_n)\alpha(\omega_e+\omega_h) + \varphi(1+\alpha\omega_z) \tag{87}$$

$$\gamma_{hkt} = \alpha\omega_k \tag{88}$$

$$\begin{aligned}\gamma_{ht} = &-(1+\alpha\omega_z)\frac{\sigma_z^2}{2} + \alpha(\omega_e+\omega_h)C_h \\ &- \alpha\beta_c v(\gamma_e+\gamma_h)\frac{((1-\tau_y)(1-\tau_n))^2\sigma^2_{h|q}}{2} + \alpha\left[\ln\left(s_{eT}\right)^{\omega_e}\left(\frac{s_{hT}}{w}\right)^{\omega_h}\right]\end{aligned} \tag{89}$$

We now briefly give an intuition for each term from (84)-(89). Looking at equation (84), current average knowledge capital has a strong effect on future knowledge capital when knowledge depreciates slowly (low γ_k). Looking at equation (85), 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 (86), the growth of fundamental knowledge is high when the rate of cross-subsidization is

[36] Recall that

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

where m_R in the first line is given by equation (75).

high (s_{eR}, s_{hR}) or when households spend a large share of their income on tuition, s.

Looking at equation (87), current average human capital has a strong effect on future human capital when the transmission of of abilities from parents to children is strong (φ), the peer effect and the effect of teaching equipment and faculty is high ($\omega_z, \omega_e, \omega_h$). Looking at equation (88), current average knowledge capital has a strong effect on future human capital when knowledge capital matters a lot for teaching equality ω_q. Finally, looking at equation (89), the growth of human capital is high when household spend a significant share of their income on tuition s and universities spend a lot on teaching equality s_{eT}, s_{hT}.

Expressing $\bar{p}$ as a function of m_h and m_k. From the earlier expression for $\bar{p}$ given by equation (49), and using the expressions for the mean of college quality (70), and college revenues (75), we obtain

$$\epsilon_1 \log \bar{p} = (\epsilon_1 - (\omega_e + \omega_h))(C_h + (1-\tau_y)(1-\tau_n)m_h) + \beta_c v(\gamma_e + \gamma_h) \left(\frac{((1-\tau_y)(1-\tau_n))^2 \sigma^2_{h|q}}{2} \right)$$
$$- \omega_k m_k - \left(\frac{\omega_e + \omega_h + \beta_c v (\gamma_e + \gamma_h)}{\epsilon_1} - 1 \right) \left(\epsilon_A m_h - \epsilon_2 \frac{\sigma_z^2}{2} \right) - \log (s_e)^{\omega_e} \left(\frac{s_h}{w} \right)^{\omega_h} \tag{90}$$

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

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

Organizing this equation,

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

From (50), (51), and (75),

$$\omega_k m_k + \omega_e(\ln s_e + m_R) - (\omega_e + \beta v\gamma_e)\left(\frac{\sigma_{h|q}^2}{2}\right) - \omega_z \frac{\sigma_z^2}{2} = \epsilon_1(\ln s + m_h) - \epsilon_1 \log \bar{p} - \epsilon_2 \frac{\sigma_z^2}{2} \tag{91}$$

Since $m_{\bar{z}} = E[\ln \bar{z}] = E[E_{\phi(\cdot)}[\ln z]] - \sigma_u^2(\phi; p)$ and the mean of $\ln z$ and from(42),

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

From (73),

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

Plugging this into (91) and from (61),

$$\omega_k m_k + \omega_e(\ln s_e + m_R) + \omega_z m_{\bar{z}} = \epsilon_1(\ln s + \ln w + m_h) - \epsilon_1 \log \bar{p} - \epsilon_2 \frac{\sigma_z^2}{2} \tag{92}$$

The left hand side is the mean supply of log quality of colleges m_q, where the first, second, and third terms are the mean contributions of log intangible capital, teaching expenditure, and the peer effect. The right hand side is the mean demand of log quality from taking log and the mean of (12). $\log \bar{p}$ is determined to balance the supply and demand of mean quality of colleges.

F.7 Proof of Proposition 4

Following our guess for the value function (32), in steady-state it is given by $V(k) = \bar{v} + v \ln k$. Substituting this guess into the recursive formulation of the value function $V(k) = \ln q + \beta V(k')$ and differentiating the latter with

respect to $\ln k$ gives

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

From our guess (26) it follows that

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

Combining both equations gives

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

Substituting this expression of v in equation (44) (and also substituting $\omega_h = 0$ and $\gamma_h = 0$) gives

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

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

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

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

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

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

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

the variance of knowledge.

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

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

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

From Propositions 2 and 3,

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

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

Since $\omega_z = 0$, the latter becomes

$$\epsilon_2 = 0$$

and the former becomes

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

In addition, from equation (12), we have

$$\Sigma_q = \epsilon_1 \Sigma_h$$

Combining the last two equations gives

$$\epsilon_1 = \omega_k \frac{\Sigma_k}{\Sigma_h} + \omega_e = \omega_k \frac{\gamma_e}{1-\gamma_k} + \omega_e. \tag{93}$$

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

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

which gives

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

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

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

Taking the mean of the log of equation (10) and using $swh = p(q, z)$, we get

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

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

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

which gives

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

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

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

and taking the expectation of both sides gives

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

Using the law of accumulation of university capital,

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

it follows that

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

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

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

where the last line uses equation (93).

F.9 Government budget constraints

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

$$\bar{G} = \mathbb{E}_j\left[\left[1 - Gk_j^{-\tau_G}\right]\left(e_{Rj} + \mathbb{E}_{\mu_{Rj}(.)}\left[wh\right]\right)\right] \tag{94}$$

$$(1 - \bar{a}_y)\int wh_{Ri}di = \int (1 - a_y)\left(wh_{Ri}\right)^{1-\tau_y} di \tag{95}$$

$$\int \frac{y_i^{\tau_n}}{(1 + a_n)} e_i di = \int \frac{e_i}{(1 + \bar{a}_n)} di. \tag{96}$$

The government balances its budget every period:

$$\bar{a}_c \int c_i di + \bar{a}_y \int w h_{Ri} di = \frac{\bar{a}_n}{(1+\bar{a}_n)} \int p_i di + \bar{G}. \tag{97}$$

Solving for G. We start from the budget constraint of the agency that distributes research grants and use the guesses (25), (26) and $s_{eR} = G(k)e_R/R$:

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

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

where m_R in the last line is given by equation (75).

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

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

$$\text{where} \quad Y_t = w_t \sum_{a=1}^{10} \exp\left(m_{h,t,a} + \frac{1}{2}\Sigma_{h,t,a}^2\right)$$

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

$$\begin{aligned} 1-\bar{a}_y &= \frac{\int (1-a_y)\left(wh_i\right)^{1-\tau_y} di}{\int wh_i di} \\ &= \frac{(1-a_y)\left(w\right)^{1-\tau_y} \sum_{a=1}^{10} \exp\left((1-\tau_y)m_{h,t,a} + \left((1-\tau_y)\right)^2 \frac{\Sigma^2_{h,t,a,}}{2}\right)}{(w)\sum_{a=1}^{10} \exp\left(m_{h,t,a} + \frac{\Sigma^2_{h,t,a}}{2}\right)} \end{aligned} \tag{98}$$

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

$$\begin{aligned} 1+\bar{a}_n = \frac{\int (1+a_n) e_i y_i^{-\tau_n} di}{\int e_i di} = \frac{\int (1+a_n) s y_i^{1-\tau_n} di}{\int s y_i di} = (1+a_n)(1-a_y)^{-\tau_n} \frac{\int \left(\tilde{w}h_i\right)^{(1-\tau_n)(1-\tau_y)}}{\int \left(\tilde{w}h_i\right)^{1-\tau_y}} \\ \iff 1+\bar{a}_n = \frac{(1+a_n)}{(1-a_y)^{\tau_n}} \frac{(\tilde{w})^{(1-\tau_y)(1-\tau_n)} \exp\left((1-\tau_y)(1-\tau_n)m_h + \left((1-\tau_y)(1-\tau_n)\right)^2 \frac{\Sigma^2_h}{2}\right)}{(\tilde{w})^{1-\tau_y} \exp\left((1-\tau_y)m_h + \left((1-\tau_y)\right)^2 \frac{\Sigma^2_h}{2}\right)} \\ \iff 1+\bar{a}_n = \frac{(1+a_n)}{(1-a_y)^{\tau_n}} (\tilde{w})^{-(1-\tau_y)\tau_n} \exp\left(-(1-\tau_y)\tau_n m_h + \tau_n(\tau_n-2)\left((1-\tau_y)\right)^2 \frac{\Sigma^2_h}{2}\right) \\ \iff a_n = (1-a_y)^{\tau_n} (\tilde{w})^{\tau_n(1-\tau_y)} \exp\left(\tau_n(1-\tau_y)m_h + \tau_n(2-\tau_n)((1-\tau_y))^2 \frac{\Sigma^2_h}{2}\right)(1+\bar{a}_n) - 1 \end{aligned} \tag{99}$$

G Robustness Checks - Calibration Details

In this section, we conduct two robustness checks. The first one looks at the extreme case where the direct contribution of research to teaching quality becomes vanishingly small (e.g. $\omega_k \to 0$). This corresponds to a situation where intangible knowledge acts as a pure signal of the universityâĂŹs reputation and as a coordination device for high ability and wealthy students to congregate at the same colleges. Our second robustness check is motivated by the concern that some expenditures that are classified as "teaching expenses" are also inputs in research, such as buildings. These expenses

could also contribute to the state variable k'. We thus propose a robustness check, where we generalize the production function of the dynamic input k' to incorporate some of the teaching equipment spending: $k' = k^{\gamma_k} e_R^{\gamma_e} \bar{h}_R^{\gamma_h} e_T^{\gamma_T}$.

In both robustness checks, we recalibrate all parameters. We calibrate the same set of moments as in the baseline calibration, with the exception of the innovation-education gap which we omit in the robustness check in which $\omega_k \to 0$. For the calibration of γ_T, we make the extreme and conservative assumption that teaching equipment is as productive for research as research equipment $\gamma_e = \gamma_T$. This is arguably an upper bound for γ_T, and the results should thus be interpreted as the most conservative of possible outcomes.