On the Spatial Distribution of Colleges*

Jacob Wright†
University of Minnesota

June 30, 2025

CLICK HERE FOR MOST RECENT VERSION

Abstract

In the United States, there is substantial cross-state variation in college (1) spending per student, (2) tuition, and (3) capacity. Empirically, I show that larger, more productive states have higher in-/out-of-state tuition, spend more per student, and offer less availability to out-of-state students. Motivated by these facts, I develop a heterogeneous agent lifecycle model embedded within a quantitative spatial framework to better understand cross-state patterns in higher education. Key features of the model include an endogenous distribution of college quality and firm activity across locations, along with the migration-based sorting of students and workers. I estimate the model to replicate cross-state dynamics of migration and education choices, wages, and college characteristics. Key parameters governing college quality and policy choices are calibrated and externally validated using causal, reduced-form evidence. Given the significant diversity and decentralized structure of the U.S. college system, I use the model to compute the optimal spatial distribution of college financing and availability at both the state and federal levels. For the federal policymaker, a key constraint is the historical distribution of colleges: the physical locations of colleges were determined more than 100 years ago, often far from centers of modern economic activity. For a fixed level of expenditures on higher education, optimal state policy increases aggregate welfare by 3.1%, and optimal federal policy increases welfare by 7.2%.

JEL Codes: H52, H75, I23, I28, R12, R13

*A special thanks to Ellen McGrattan and Todd Schsoellman for their invaluable guidance. I also thank Honey Batra, Chris Ellul, Juliana Gamboa-Arbelaez, Thomas Hasenzagl, Kyle Herkenhoff, Karen Kopecky, Ioannis Koutsonikolis, Oksana Leukhina, Anusha Nath, Andrii Parkhomenko, Bipul Verma, Mike Waugh, Alex Wurdinger, Eric Young, and Angela Zheng for many helpful comments and discussions. The author declares that he has no relevant or material financial interests that relate to the research described in this paper. The author did not receive any financial support for completion of this research.

†email: wrig0974@umn.edu. website: jacob-wright.me.

1 Introduction

College systems feature dramatic heterogeneity across the United States. Significant spatial variance exists across states that are otherwise remarkably similar along economic, demographic, or geographic metrics. This paper studies three main aspects in which public systems of higher education differ: (1) in-/out-of-state tuition, (2) spending per student, and (3) in-/out-of-state seat availability. The first question asked is: "Why are college systems so different across states?" Given this large degree of heterogeneity across locations and the clear decentralization of the higher education system, the second question asked is: "What is the optimal distribution of college tuition, spending per student, and capacity across space?"

I begin by documenting two sets of motivating facts in the data. First, states with larger populations and higher wages tend to charge higher tuition, allocate more resources per student, and offer less access to out-of-state students. These states also exhibit lower rates of alumni outmigration alongside greater inflows of college-educated individuals from other states. I develop a quantitative framework that endogenizes the relationship between migration flows, wages, and state-level higher education policies in order to rationalize this correlation in state policies.

Second, I show that the distribution of college quality is highly persistent over time and spatially concentrated. Unlike firms, where the average age is five years, 73% of students attend a public college whose location was fixed over 100 years ago. As a consequence, the spatial distribution of colleges today is not uniform relative to potential students or firm activity. This fact creates two key inefficiencies that motivate the modeling structure of the federal government problem. First, considerable empirical work has established that college attendance is highly elastic to the distance required to travel (Fu et al., 2022; Hillman, 2017; Turley, 2009). Hence, the distribution of college students (or lack of attendance) is likely not optimal. Second, it creates a potential mismatch between college and firm location, whereby the highest-quality colleges (and hence graduates) may not be located in the same places as

the most productive firms.

I formulate a novel theory of U.S. higher education that captures heterogeneity across space and how this variation influences patterns of educational enrollment and firm activity. More specifically, I build a lifecycle model of human capital accumulation, embedded in a general equilibrium spatial framework, featuring migration decisions over education and work locations. High school graduates with differing ability, human capital, and parental income choose whether to attend college and, if so, in which state. Each state features a college of differentiated quality and an admission cutoff, both of which are endogenous with respect to student sorting across locations. College quality affects the trajectory of lifecycle wage growth.

Given an endogenous set of wages and housing prices, college graduates then decide where to work. In each state, a representative firm of differentiated productivity combines human capital from high school and college graduates. Finally, a state government operates in each location, and a federal government operates across the economy as a whole. Both levels of government choose an endogenous labor tax to finance college education expenditures.

In addition to the fixed initial condition governing the spatial distribution of colleges, the model developed here includes several additional frictions relevant to the U.S. economy that allow for welfare-improving government policy. First is moving costs, which are the standard friction featured in quantitative spatial models. Second, public colleges impose significant barriers to student geographical mobility by charging out-of-state tuition premiums ranging from 150% to over 500% of in-state tuition. At a national level, this is unlikely to result in the optimal allocation of students across colleges. Additionally, with state policymakers setting tuition prices and capacity, there is no competitive market allocation or price. Third, in the benchmark model, endogenous peer effects influence college quality and introduce an externality. Agents do not internalize the effect of their own presence on college quality, which subsequently affects the lifecycle earnings of all students attending the college. Finally, in an extension of the model, spillovers from local college quality to firm productivity are

not internalized by either agents or firms.

I calibrate the model to fit a rich set of cross-sectional moments on migration and education choices, human capital accumulation, cross-state wages, and cross-state housing rental rates. A key moment for household and policymaker decisions is the elasticity of future earnings with respect to college expenditures per student. I provide causal instrumental variable estimates of this elasticity and calibrate the model to match. A second set of salient elasticities relevant to the state government's optimal policy decisions are the elasticities of in-state college attendance and in-state work with respect to local wages, college tuition, and college spending per student. The model performs well in replicating these elasticities out of sample in an external validation exercise.

Importantly, the model also generates similar levels of cross-state heterogeneity in college policies and is able to rationalize the documented cross-state empirical correlations between college policies, population levels, and wages. Under the standard assumption of a convex cost function, the positive correlation between state size, in-/out-of-state tuition, and the percentage of the student body that is in-state is driven by scale economies. When population increases, the relative value of an out-of-state student decreases compared to spending per student or reducing in-state tuition. Hence, larger states admit relatively fewer out-of-state students at a higher price.

The model also reproduces the positive correlation between college-educated wages and in-/out-of-state tuition, as well as the percentage of the student body that is in-state, through migration patterns. Consider high-wage state California versus low-wage state North Dakota. For North Dakota (California), the percentage of alumni remaining in-state and the net import of college graduates is low (high). North Dakota will attract out-of-state students by setting tuition low and capacity high. Some of these out-of-state students will then remain in-state after graduation. On the other hand, college graduates migrate in large numbers to California, regardless of whether they are incentivized to attend college in the state. Hence, it is more beneficial for California to spend tax revenues on in-state students than on attracting

out-of-state students.

For the main quantitative exercise, I fix state and federal tax revenues and then compute optimal government policies over tuition, capacity, and spending per student in order to maximize welfare. For the state-level counterfactual, each state maximizes its own aggregate welfare. For federal policy, all state policies are jointly chosen to maximize national welfare.

I find that, for a fixed level of tax revenues spent on higher education, when setting policy parameters optimally, both state and federal governments can significantly improve aggregate welfare. This is true with or without the presence of spillovers from university quality to firm productivity. Moreover, I find room for welfare improvements even when total expenditure on higher education is reduced. The federal government problem yields much larger increases in aggregate welfare or 7.2% compared to the state-level problem of 3.1%. Many states are already acting optimally when maximizing the welfare of agents born in the given state.

The results from the federal problem highlight the importance of mismatch between historical college quality and current firm productivity. I find that governments subsidize tuition more heavily in areas with highly productive firms and colleges. They face (among other trade-offs) a choice between increasing college quality and attendance in highly productive firm areas versus in high college quality areas. More specifically, when high-productivity firms are "matched" with high-quality colleges, the federal government allocates a large majority of revenues to high-quality and high-productivity states. However, when high-productivity firms are mismatched with high-quality colleges, the policymaker distributes funds more evenly.

Finally, I also conduct policy experiments in line with proposed reforms to the U.S. system of higher education—namely, free tuition and the elimination of out-of-state tuition. I find that eliminating out-of-state tuition changes migration patterns by attracting higher-ability students to high-quality colleges. Overall, there are positive but relatively modest effects on aggregate welfare of 1.2%. Free tuition has similar effects on student migration but yields

a much larger increase in aggregate welfare of 4.3%. College quality falls in all locations under this policy, but not by nearly enough to outweigh the positive effects of eliminating the financial frictions induced by tuition.

The remainder of this paper is structured as follows. First, I review the relevant literature in *Section 1.1*. In *Section 2*, I contrast higher education systems and establish the large degree of heterogeneity across U.S. states. I then empirically motivate the trade-offs faced by state and federal policymakers. Given the empirical facts established, the quantitative theory is developed in *Section 3*. In *Section 4*, I show that the model is able to endogenously generate and rationalize key empirical facts. *Section 5* describes the calibration of the quantitative framework. *Section 6* presents the results of the counterfactual exercises. *Section 7* concludes.

1.1 Related Literature

I bridge several strands of the quantitative literature that have so far remained separate. In doing so, I provide a novel theory of the U.S. higher education sector and its relation to firm activity and migration decisions across space.

Structural Models of Higher Education – The modeling of the higher education sector relates to a relatively small literature that develops structural models generating an endogenous distribution of university quality. Those most similar to my paper are Hendricks et al. (2021) and Epple et al. (2017). However, both papers study substantially different questions: the former investigates the rise in college attendance following World War II, while the latter examines the interaction between financial aid at public and private colleges. More closely related to the present question are Brotherhood et al. (2023) and Wright and Zheng (2025), who, as in my work, study the reallocation of college seats across different levels of college quality. However, these papers primarily focus on income-based affirmative action policies and their effects on intergenerational mobility.

The primary methodological difference between my paper and this strand of the literature is that I model a competitive university sector embedded within a general equilibrium spatial framework, featuring a production sector and housing markets. This approach allows me to address a broader set of macroeconomic questions.

Higher Education Financing Across Regions – An extensive empirical literature studies the effects of state-level differences in college financing and tuition levels and is reviewed in Kane (2006). On the quantitative side, this work relates to several papers that examine differences in college tuition subsidies, financing, and quality—however, at the cross-country level rather than the cross-state level. Herrington (2015), Holter (2015), and Wright (2023) find that a significant portion of cross-country differences can be explained by their respective systems of higher education.

Optimal Policy and Higher Education – This paper also relates to a broader quantitative literature that studies optimal education taxation and subsidy policies. For example, Krueger and Ludwig (2016) and Abbott et al. (2019) compute optimal tuition subsidies financed through progressive labor taxes. While this paper also examines optimal tuition subsidization, I do so within a spatial framework, focusing primarily on evaluating the efficiency implications of geographic variation in college tuition, quality, and seat availability.

Migration and Human Capital Accumulation – This paper is closely related to an industrial organization (structural micro) literature studying migration and human capital accumulation in response to spatial variation in college costs and availability. Fu et al. (2022) studies the welfare consequences of "education deserts" and shows that student college demand is elastic with respect to the proximity of colleges. Ishimaru (2023) investigates the link between birth location, migration, and economic outcomes.

Most similar to this paper in both research question and approach to modeling migration

is Kennan (2015), who examines how tuition differences across states explain state-level variation in the proportion of college graduates in the local labor force. He finds that tuition differences have large effects on college enrollment that are not eliminated through migration.

I differ from these works along several lines. First, I model the macroeconomy in a general equilibrium setting. Second, I consider an endogenous college quality distribution across locations, in addition to differences in college financing and availability. Finally, given the macroeconomic structure of my model, I am able to undertake an aggregate welfare analysis and evaluate optimal government policy.

Quantitative Spatial Economics – Finally, this paper relates to a large quantitative spatial literature studying place-based policies, as reviewed in Redding and Rossi-Hansberg (2017). Spatial models and optimal policy typically focus on firms or local infrastructure investments, as in Kline and Moretti (2014) and Fajgelbaum and Gaubert (2020). While my model incorporates spatial firm heterogeneity, it primarily focuses on the higher education sector when evaluating place-based and optimal government policies. My model also differs from this line of work in that most papers in this literature depend on agglomeration economies as the main source of inefficiency. In contrast, my framework does not rely on agglomeration externalities for any of the results.

An important avenue for future work, which could be pursued through the lens of my model, is to consider firm-side policies in the presence of an endogenous college sector and a spatial distribution of skilled and unskilled human capital. Are firm-side policies aimed at improving output in impoverished areas effective without a nearby college to supply the skilled labor necessary for production?

2 The Geography of Colleges and Government Policy

This section documents several empirical facts to motivate the model and subsequent quantitative analysis. First, I describe the large degree of heterogeneity in state college systems across the United States. I then present motivating relationships between state characteristics and education policies. I conclude this section with a brief discussion of the historical distribution of college qualities. The quantitative spatial model developed in *Section 3* provides an explicit theory of these empirical observations. This allows me to better analyze heterogeneity across states and evaluate the aggregate efficiency of alternative policy regimes.

For all data analysis, I focus on public, four-year degree-granting institutions with at least 250 students enrolled. All data are for the 2017–2018 academic year. This amounts to 572 colleges across 50 U.S. states (the District of Columbia is excluded). I choose this time period in order to avoid large changes to the higher education sector during the COVID-19 pandemic. Further results exploring colleges and higher education policy across states are reported in *Appendix A.2*, and details of the empirical calculations and data sources are in *Appendix A.1*.

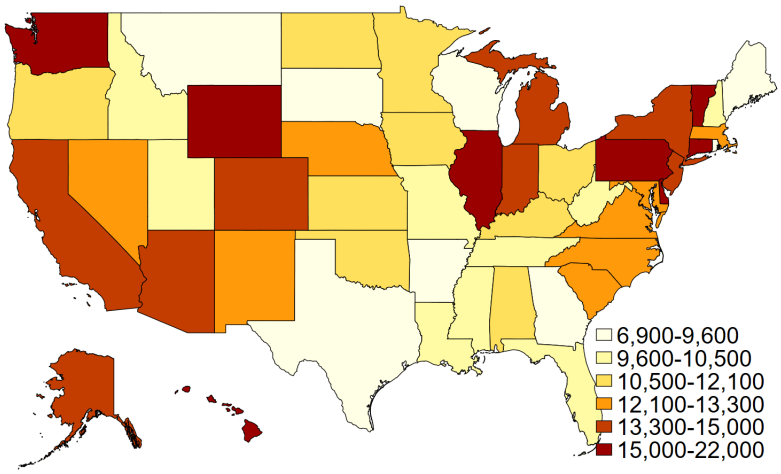
2.1 Cross State Heterogeneity

The state government controls three main aspects of the public college system: (1) financing or spending per student, (2) in-/out-of-state tuition, and (3) in-/out-of-state capacity. In this section, I document the significant cross-state heterogeneity along these three dimensions.[1,2]

[1] While it is true that variation within states can also be substantial, I focus on cross-state variation here.

[2] In practice, the public college budgetary process involves a complicated set of negotiations with the state government and college board. This bureaucratic process also differs significantly across states. However, in the majority of states, the state government has the legal authority to set tuition, capacity, and college budget, which ultimately determines college spending. See Parmley et al. (2009) for an in depth discussion of higher education budget practices across the United States.

Figure 1: **Instructional Spend per FTE Across States**



Notes: This figure plots total instructional spending per full-time equivalent student across the 50 U.S. states. Amounts are reported in 2017 real dollars. Data are taken from the Integrated Postsecondary Education Data System (IPEDS) for the 2017-2018 academic year.

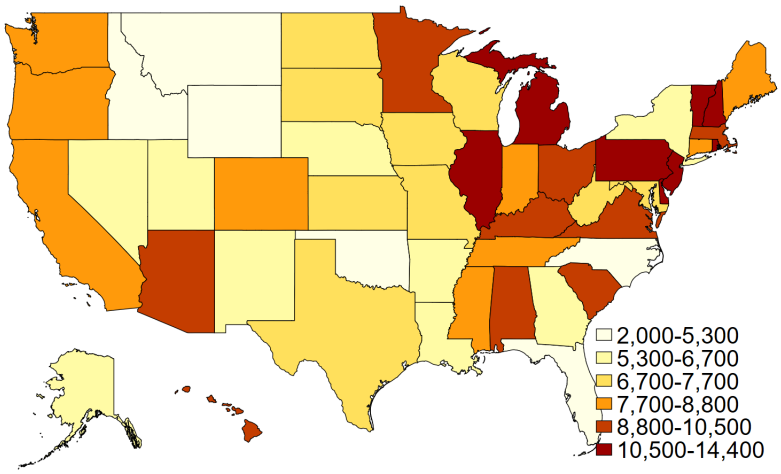
Financing – State financing of colleges vary along several dimensions. First, the amount spent per student. *Figure 1* plots instructional expenditures per full-time equivalent (FTE) student across states. There is substantial variation across the United States, with Pennsylvania spending nearly twice as much as California and three times as much as Texas.

It is not immediately clear which states "value" education most based on their spending patterns. Texas (California) spends a small (large) amount per student but allocates a large (small) portion of total government expenditures to education. Florida, by contrast, spends a relatively small amount per student, yet more than half of the college budget is financed through state tax revenues.[3] This implies that other forms of revenue, such as tuition or donations, are much more important to college budgets in some states than in others.

Tuition – *Figure 2* plots (per-semester) in-state sticker tuition prices. Again, there is large variation in tuition levels. For example, in-state tuition in Pennsylvania is more than double that of neighboring New York. All states also set out-of-state tuition levels higher than

[3]Federal appropriations represent a small portion of college budgets in all states. The vast majority of federal college funding comes in the form of Title IV aid.

Figure 2: **In-state Sticker Tuition**



Notes: This figure plots in-state sticker tuition across the 50 U.S. states. Amounts are reported in 2017 real dollars. Data are taken from the Integrated Postsecondary Education Data System (IPEDS) for the 2017-2018 academic year.

in-state tuition.[4]

Figure 3 plots the ratio of out-of-state to in-state tuition. The figure shows similarly large disparities across states but with a pattern entirely different from that in *Figure 2*. At one end of the spectrum, Minnesota and Illinois charge roughly 75% more to out-of-state students, whereas North Carolina and Florida charge over 450% more.

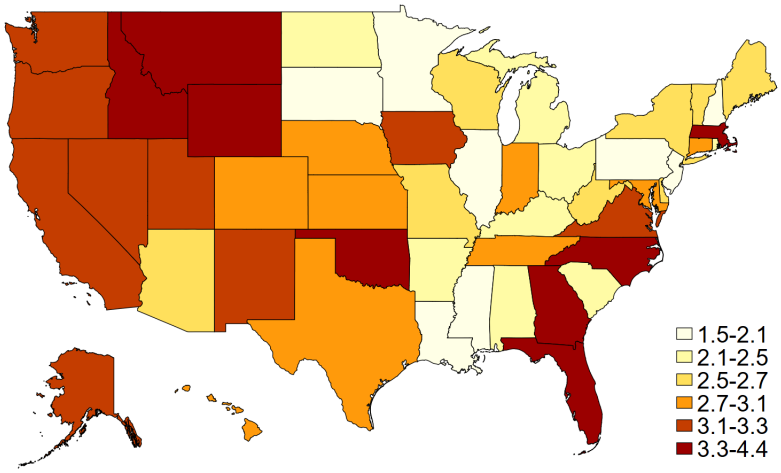
Capacity – Available college seats at existing institutions, relative to in-state applicant demand, are more difficult to represent directly with observable data. However, I present several summary statistics that suggest capacity also varies widely across states. First, in-state admission rates range from below 50% in California to above 95% in Iowa. *Figure 4* plots total admission rates.[5]

A more direct measure of immediately available public college seats is total college enrollment per high school student. While this measure does not account for demand or the availability of private options it serves as a good proxy. Given differences in tuition levels,

[4]Reciprocity agreements exist for certain regional groupings of states (e.g., Minnesota and Wisconsin). I will account for these arrangements when taking the quantitative model to the data in *Section 5*.

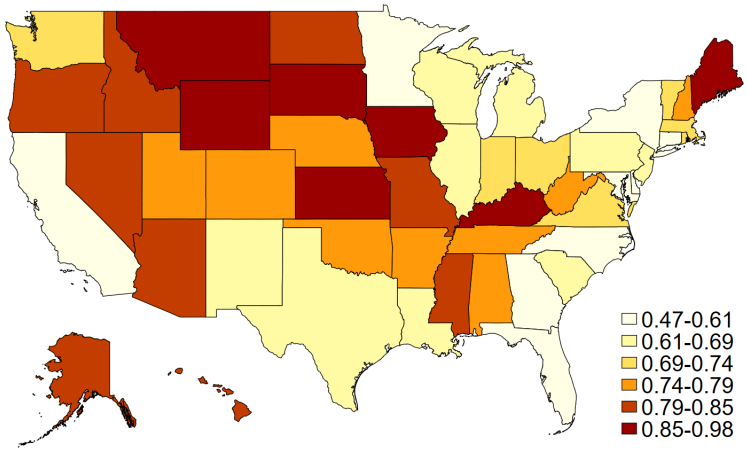
[5]Due to data limitations, I am unable to plot in-state admission rates for all public colleges and states.

Figure 3: **Out-of-state to In-state Tuition Ratio**



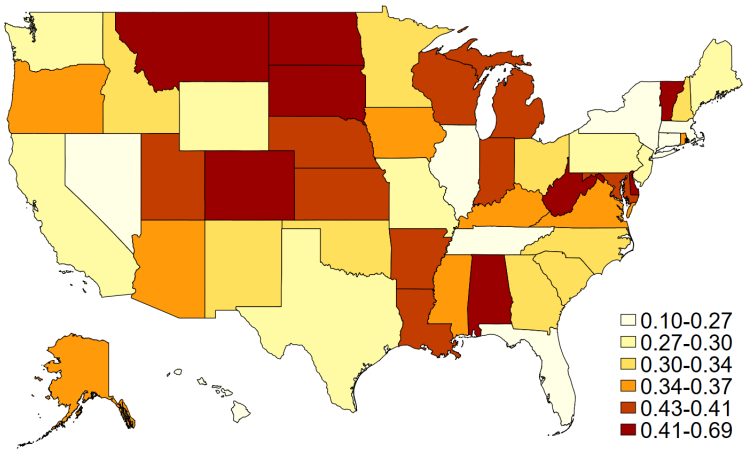
Notes: This figure plots the ratio of out-of-state to in-state sticker tuition across the 50 U.S. states. Data are taken from the Integrated Postsecondary Education Data System (IPEDS) for the 2017-2018 academic year.

Figure 4: **Admission Rate**



Notes: This figure plots the admission rate across the 50 U.S. states. Data are taken from the Integrated Postsecondary Education Data System (IPEDS) for the 2017-2018 academic year.

Figure 5: **College Enrollment per High School Student**



Notes: This figure plots the number of full-time equivalent college students per high school student across the 50 U.S. states. Data are taken from the Integrated Postsecondary Education Data System (IPEDS) for the 2017-2018 academic year.

a college will strictly prefer out-of-state students. Hence, the state government must set a cap on out-of-state students or a minimum on in-state students. Assuming these caps bind, then in-state capacity can be inferred as in-state enrollment to high school graduates.

Figure 5 plots this statistic and shows that, unsurprisingly, states with low admission rates also tend to have a relatively low number of available college seats per high school student. Again, California and New York are among the lowest in the country, with fewer than 0.3 spots per high school student. At the other end of the spectrum are the Dakotas, with nearly 0.7 spots per high school student.

While these trends could be driven by the relative availability of public versus private colleges, *Figure D.1* in *Appendix D* suggests this is not the case. Although there is variation across states, public college enrollment as a percentage of total enrollment is high on average at 79.1%. I discuss the role of private colleges further when introducing the quantitative model in *Section 3*.

2.2 Motivating Facts

This section documents several facts that motivate the model and quantitative analysis. First I describe patterns observed in state college policy relative to state characteristics across states. Second, I document a set of facts relevant to a counterfactual centralized system of higher education.

2.2.1 The State Government

Given this significant heterogeneity, it is natural to ask: "Why are college systems so different across states?" While I ultimately address this question using a quantitative theory, I first present several relationships observed in the data to motivate this analysis.

I find that the two main correlates of state higher education policy are population size and college wages. *Tables 1* and *2* show the OLS relationships between these variables and the main college policy parameters set by state governments: (1) out-of-state tuition, (2) in-state tuition, (3) out-of-state capacity, and (4) spending per student.

Regression results in columns (1) and (2) show that larger states charge higher in-state and out-of-state tuition on average. Similarly, states with higher college wages also charge higher tuition for both in- and out-of-state students. Column (3) from both tables shows that wealthier and more populous states tend to enroll relatively fewer out-of-state students. Finally, there is no clear relationship between spending per student and state size: column (4) shows a coefficient near zero and not statistically significant. However, there is a strong positive correlation between spending per student and average college wages in a state, as shown in column (4) of *Table 2*.

A large portion of the variation in out-of-state capacity across states is explained by population size, with column (3) reporting an R^2 of 0.48. For instance, on average, states like California educate very few out-of-state students, whereas states like Vermont educate significantly more out-of-state than in-state students. While state size explains little of the

Table 1: **Motivating OLS – State Policy and Population Level**

	(1) $\ln(T^{out})$	(2) $\ln(T^{in})$	(3) ln(% in-state)	(4) ln(spend per)
ln(*population*)	0.101	0.150	0.130	-0.018
	(0.028)	(0.020)	(0.000)	(0.599)
constant	8.39	6.68	-2.34	11.06
	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.11	0.12	0.48	0.01
Observations	50	50	50	50

Notes: P-values show in brackets. Data are taken from the Integrated Postsecondary Education Data System (IPEDS) for the 2017-2018 academic year and the American Community Survey (ACS).

variation in spending per student, a significant portion is explained by state wealth, with an R^2 of 0.32.

Naturally, population size and college wages are highly correlated in the data. However, these correlations remain even when both variables are included in the same regression specification. Moreover, I present these statistics simply as motivating evidence for the question and structure of the quantitative analysis.

I present two additional statistics to provide further insight into the potential causes underlying the observed correlations. First, in column (1) of *Table 3*, I regress the two key correlates discussed above—college earnings and population—on the percentage of a state's college alumni who leave the state to work elsewhere. The regression also includes controls for the unemployment rate and the percentage of students who are in-state. Column (2) replicates the same regression using the net inflow of college graduates as the dependent variable. On average, larger and wealthier states experience lower outflows of their own alumni and higher inflows of alumni from other states.

Finally, *Figure 6* shows that expenditures per student, as presented in *Figure 1*, do not tell the full story. A more relevant measure is the cost of net college graduates—that is, the cost of producing a college graduate who remains in-state or of "importing" a graduate

Table 2: **Motivating OLS – State Policy and College Wages**

	(1) $\ln(T^{out})$	(2) $\ln(T^{in})$	(3) ln(% in-state)	(4) ln(spend per)
ln(*college wages*)	0.661	0.891	0.706	1.441
	(0.041)	(0.106)	(0.008)	(0.000)
constant	2.54	-0.97	-8.10	-4.80
	(0.473)	(0.872)	(0.007)	(0.160)
R^2	0.09	0.06	0.14	0.32
observations	50	50	50	50

Notes: P-values show in brackets. Data are taken from the Integrated Postsecondary Education Data System (IPEDS) for the 2017-2018 academic year and the American Community Survey (ACS).

who earned a degree elsewhere. *Figure 6* plots this cost. For example, while Michigan (West Virginia) spends a large (small) amount per college student, it spends relatively little (a lot) to produce a net college graduate.

2.2.2 The Federal Government

Suppose there were a federal system of higher education in the United States.[6] Would one observe different patterns in the data than presented above, with each state acting individually? Certainly, the federal policymaker has a different objective and faces a different set of trade-offs. If the federal government cared about each state equally (i.e., maximized national welfare), its main consideration would be how to distribute funds across the U.S. spatially. That is, should the federal government spend an extra dollar in California or Idaho? In this section I present two facts which motivate the modeling structure relevant for a counterfactual economy in *Section 6* where college policy is simultaneously chosen for

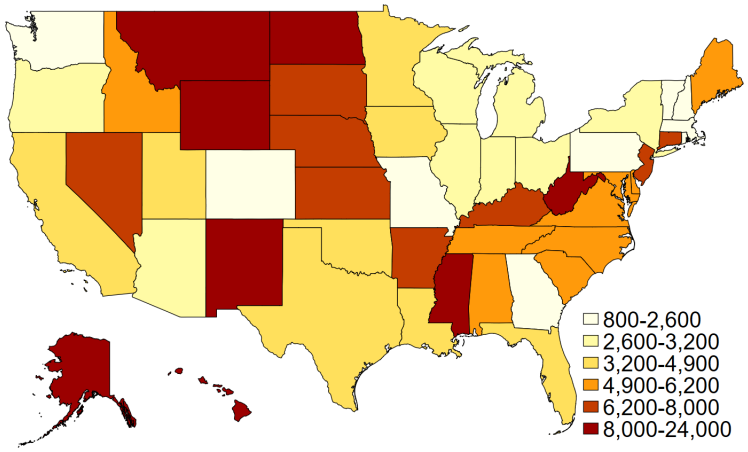
[6] A federal education system was first proposed early in U.S. history by several founding fathers. The strongest proponent was James Madison, who addressed the issue in each State of the Union speech and put forth a bill that passed the Senate but ultimately failed in the House of Representatives. Today in the U.S., the federal government's explicit involvement in higher education is limited to a handful of military colleges. However, fully centralized higher education systems do exist in other advanced or developing countries, such as Germany or Brazil.

Table 3: **Alumni Migration Patterns and State Characteristics**

	(1) ln(% alum leave)	(2) ln(net import)
ln(*college earn*)	-0.735	2.631
	(0.069)	(0.056)
ln(*population*)	-0.134	0.277
	(0.014)	(0.004)
ln(*unemployment*)	0.160	-0.315
	(0.353)	(0.0376)
ln(*%in − state*)	-3.36	
	(0.000)	
constant	8.14	
	(0.048)	
R^2	0.92	0.56
observations	50	50

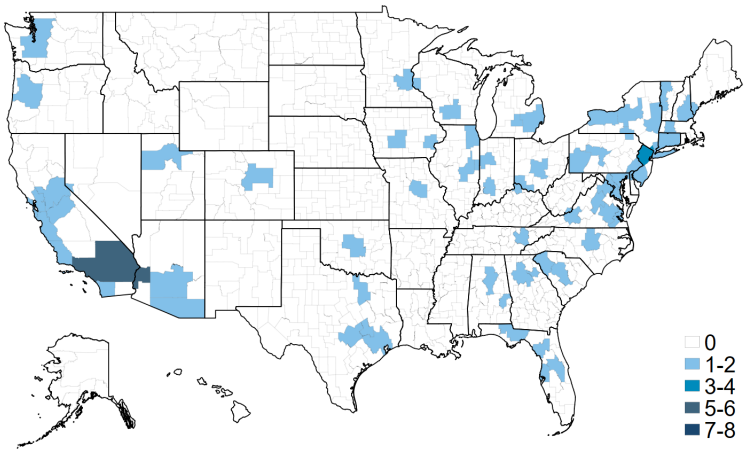
Notes: P-values show in brackets. Data are taken from the Integrated Postsecondary Education Data System (IPEDS) for the 2017-2018 academic year, the American Community Survey (ACS), and Conzelmann et al. (2023).

Figure 6: **Cost of Retained and Imported College Graduates**



Notes: This figure plots state expenditures per graduate student retained (i.e., remaining in-state after graduation) and imported (i.e., moving to the state after graduation). Data are from open access sources provided by Conzelmann et al. (2023).

Figure 7: **Top 150 Ranked Public Colleges by Commuting Zone**



Notes: This figure plots the count of top 150 nationally ranked public colleges (out of all colleges) located within each commuting zone. Rankings are taken from U.S. News (2020). Commuting zones are defined as in Autor and Dorn (2013).

each state by a federal government.

Figure 7 plots the count of public colleges ranked inside the top 150 nationally, by commuting zones. There is an obvious clustering along the coasts, with most central states lacking a high-quality public college. Overall, 411 of 741 commuting zones (18% of the total population) lack a public college of *any* quality. An empirical literature studies these "education deserts" and finds that student demand for college is elastic to distance (Fu et al., 2022; Hillman, 2017; Turley, 2009). Hence, the U.S. may "miss" high-ability students due to a lack of colleges in certain regions.

On the other side of this trade-off, high-productivity firms tend to cluster in dense urban areas, and naturally, it is most efficient to match them with high-ability students graduating from high-quality colleges.[7]

A second key consideration for the federal government in deciding on the location of college spending is the current location of colleges. Unlike firms, whose average age is five years, 73% of students attend a public college whose location was fixed over 100 years ago.

[7]While there is a large literature studying spillovers from universities to firms and agglomeration effects in high-population-density areas, I abstract from this in the benchmark model. See Rosenthal and Strange (2004) or Henderson (2007) for a review of the agglomeration literature. See Jaffe (1989) or Kantor and Whalley (2014) for an empirical analysis of college spillovers.

Consider, for example, Penn State, which was founded in 1862 as the Agricultural College of Pennsylvania. The centralized rural location had a strong policy rationale at the time: to provide agricultural expertise to local farmers. While the federal policymaker is able to move spending and capacity to more urban locations, the stock of intangible capital—name recognition, buildings, or placement networks, etc.—that determines college quality is fixed and cannot be readily relocated.

Table 4 shows that, on average, colleges built over 100 years ago are located in (relatively) less densely populated areas today than when they were founded. Each county is nationally ranked by population density in 1920 and 2020, and then assigned a percentile. I take the difference between the percentile rank in 2020 and 1920 for the county in which a college is located. Hence, if a college is located in a county that became relatively less dense, this difference will be negative. This difference is then regressed on a dummy variable indicating whether the college was built over 100 years ago, and controls for size, land-grant status, and initial density. I find that colleges built over 100 years ago are associated with nearly a 10-percentile decrease in relative county population density.

2.3 Discussion

Taking stock, I have presented three key sets of motivating observations from the data. First, there is dramatic heterogeneity across states in their systems of higher education. In particular, states differ significantly in: (1) in-/out-of-state tuition, (2) in-/out-of-state capacity, and (3) spending per student. My model will feature a government's problem that explicitly optimizes over these five policy levers.

The second set of empirical observations shows that larger and higher-wage states tend to charge higher tuition, admit fewer out-of-state students, and spend more per student. These states also experience lower alumni outflows and larger inflows of college graduates from other states. The novel quantitative framework developed in the next section will generate an endogenous relationship between migration flows, wages, and state college policy.

Table 4: **Changes in Relative Population Density by College Founding**

	(1) Density Difference
$i.100\ years\ old$	-9.31
	(0.000)
FTE	0.001
	(0.000)
$i.land\ grant$	0.55
	(0.776)
$dense\ rank$ 1920	-0.51
	(0.000)
$constant$	43.93
	(0.000)
R^2	0.68
observations	570

Notes: The dependent variable is the difference between the county population density percentile ranking (where the college is located) in 2020 and 1920. P-values are shown in brackets. Founding data are from the author's own collection. Historical county-level population density is taken from the U.S. Census Bureau. Control variables are obtained from the Integrated Postsecondary Education Data System (IPEDS).

Finally, the third set of facts highlights that the distribution of college quality is highly spatially concentrated and persistent over time. The model will analyze these empirical patterns through the lens of a counterfactual federal government optimally choosing college policies simultaneously for each state.

3 Model

Time is discrete and has an infinite horizon. One model period represents five biological years and is denoted by j. The economy is populated by measure one of J overlapping generations with a uniform demographic structure.

The economy contains a discrete number of locations indexed by $\ell \in \mathfrak{L}$. Each location is populated with an initial mass $\mathcal{M}_\ell$ of high school graduates with heterogeneous family backgrounds, ability, and human capital. Agents first choose whether to attend the local college, attend college elsewhere, or immediately enter the workforce. Agents then choose which location to work in. In order to move across locations, an individual must pay a moving cost. Agents consume and purchase housing services, which are in fixed supply for each location. While working over their lifecycle, individuals are endowed with one unit of time which they supply to the labor markets inelastically and accumulate human capital on-the-job via learning by doing.

A college of differentiated quality operates in each location. The college sets an admissions policy that specifies the ability needed for attendance. In each location, a representative firm aggregates human capital from high school or college graduates in order to produce consumption goods. A state government operates in each location and a federal government operates across the economy as a whole. Both governments tax labor market earnings proportionally.

3.1 Agents

In this section, I describe an agent's lifecycle. I first lay out preferences and the static optimization problem. I then describe the lifecycle optimization problem. Lastly, I present the sequence of discrete choices over education and migration.

3.1.1 Static Problem

An agent who chose education type e, working location ℓ_w, and has a current human capital level h, solves the following static optimization problem to allocate labor market earnings over consumption c, and housing d,

$$\mathcal{U}(e, \ell_w, h) = \max_{c,d} \left\{ \ln \left(\left(\frac{c}{1-\xi} \right)^{1-\xi} \left(\frac{d}{\xi} \right)^{\xi} \right) \right\}$$

$$\text{subject to,} \tag{1}$$

$$c + p(\ell_w) \cdot d = \big(1 - \tau(\ell_w)\big) \cdot \omega(e, \ell_w) \cdot h$$

where, $p(\ell_w)$ is the price of housing in location ℓ_w, $\tau(\ell_w)$ is the total tax liability in location ℓ_w, and $\omega(e, \ell_w)$ is the wage paid to an individual of education type e in location ℓ_w. Given ln utility, this problem can be analytically solved, with the solution given by,

$$\mathcal{U}(e, \ell_w, h) = \ln \Big(\big(1 - \tau(\ell_w)\big) \cdot \omega(e, \ell_w) \cdot h \Big) - \xi \ln \Big(p(\ell_w) \Big) \tag{2}$$

The first term gives the natural logarithm of disposable income, which I will denote as $\Omega(e, \ell_w, h)$. The second term is the natural logarithm of the rental rate on houses multiplied by its income share ξ. Together, this gives the utility of an agent for period j.

3.1.2 Lifecycle Problem

Each agent inelastically supplies one unit of time each period, and accumulates human capital via learning-by-doing. Given the solution to the problem defined by equation (1), agents maximize then receives lifetime utility according to,

$$\begin{aligned} &\max \sum_{j=0}^{J} \beta^j \cdot \mathcal{U}(e, \ell_w, h_j) \\ &\text{subject to,} \\ &\qquad h_{j+1} = \big(a \cdot Q(\ell_c)\big) h_j^{\nu} + (1-\delta) h_j \end{aligned} \tag{3}$$

Clearly this is a trivial maximum problem as static optimality holds each period, and with inelastic labor supply, human capital stock is given. β is the time discounting rate, and $\delta \in [0, 1]$ is the period depreciation rate of human capital. Let the value of this problem for an agent who did not attend and attended college be denoted as, $W^{nc}(a, h, \ell_w)$ and $W^{c}(a, h, \ell_c, \ell_w)$, respectively.

College quality Q_ℓ from location ℓ_c and ability a, determine the efficiency with which human capital is produced. $Q(\ell_c)$ is normalized to one for those who do not attend college. The productivities over the two inputs is given by ψ_1, and ψ_2. The elasticity of substitution is $\frac{1}{1-\phi}$.

A reduced form literature finds that (controlling for observable characteristics) earnings levels and growth rates vary significantly across college qualities.[8] Motivated by this finding, I employ this model structure where college alters the growth of human capital over the agent's lifecycle. That is, for a monetary cost (and opportunity cost), the agent may purchase a more productive technology for producing human capital. This modeling structure and its empirical relevance will be discussed further in *Section 5* when I take the model to the data.

[8] See for example, Chetty et al. (2020), Cellini and Turner (2019), or Kirkeboen et al. (2016).

However, it is important to note that I do not impose that $Q(\ell_c) > 1$ and instead allow for the possibility that college quality plays no role in lifetime earnings dynamics.

3.1.3 Discrete Choices

All discrete choices in the model are subject to idiosyncratic preference shocks. These shocks are distributed according to the Type I Extreme Value distribution with shape parameter equal to zero and scale parameter of one. Given this, all choice problems can be solved analytically following McFadden (1972).[9]

At age $j = 1$ an agent's initial state is characterized by ability a, human capital h, parental resources p, and birth location ℓ_b. The initial endowments are drawn from the distribution $F(a, h, p)$, which is constant across locations. The probability of being born in a given location is $Pr(\ell_b)$.

Ability a affects how much an individual learns and benefits from college, as well as the efficiency with which they accumulate human capital over the lifecycle. Human capital h determines labor market earnings. Parental resources p provide financing for consumption and tuition if the agent chooses to attend college.

Given the initial state, agents make an irrevocable decision to attend college or work as a high school graduate. This determines their education type e_c for college and e_{nc} for non-college. That is, high school graduates first solve,

$$V(a, h, \ell_b, p) = \mathbb{E}_{\epsilon_e}\Big[\max_{c,nc}\Big\{V^c(a, h, \ell_b, p) + \sigma_e\epsilon_e, V^{nc}(a, h, \ell_b) + \sigma_e\epsilon_e\Big\}\Big] \tag{4}$$

where, $V^c(a, h, \ell_b, p)$ is the optimal continuation value of attending college, and $V^{nc}(a, h, \ell_b)$ is the optimal continuation value of entering the workforce immediately as a high school graduate. ϵ_e is drawn from the Type I Extreme Value distribution and the parameter σ_e controls the variance of these shocks. The solution to this problem is given by the choice probabilities and corresponding value function,

[9] Throughout I omit the Euler-Mascheroni constant in the solution to each discrete choice problem.

$$Pr[e = \bar{e}|a, h, \ell_b, p] = \frac{\exp\left(\frac{V^{\bar{e}}(a,h,\ell_b,p)}{\sigma_e}\right)}{\sum_{\hat{e}} \exp\left(\frac{V^{\hat{e}}(a,h,\ell_b,p)}{\sigma_e}\right)} \tag{5}$$

$$V(a, h, \ell_b, p) = \sigma_e \ln\left(\sum_{\hat{e}} \exp\left(\frac{V^{\hat{e}}(a, h, \ell_b, p)}{\sigma_e}\right)\right) \tag{6}$$

For those who choose to attend college, they spend one period studying and graduate with certainty. While in college they cannot work and must consume using parental resources p. The value of attending college is given by,

$$Z(a, h, \ell_b, p, \ell_c) = \ln\left(p - T^i(\ell_c) - T^o(\ell_c)\mathbb{1}_{\ell_b \neq \ell_c}\right) - \delta(\ell_b, \ell_c) + \beta V^{cw}(a, h, \ell_c) \tag{7}$$

where, ℓ_c is the location of college attendance, $T^i(\ell_c)$ is in-state tuition, and $T^o(\ell_c)$ is the additional cost incurred for out-of-state tuition. Moving costs from the location of birth to the location of college attendance are given by the function $\delta(\ell_b, \ell_c)$. The optimal continuation value of working with a college degree is given by $V^{cw}(a, h, \ell_c)$. The agent chooses in which location to attend college by solving,

$$V^c(a, h, \ell_b, p) = \mathbb{E}_{\epsilon_{cq}}\left[\max_{\ell_c}\left\{Z(a, h, \ell_b, p, \ell_c) + \sigma_{cq}\epsilon_{cq}\right\}\right] \tag{8}$$

This gives choice probabilities and value function,

$$Pr[\ell_c = \bar{\ell}_c|a, h, \ell_b, p] = \frac{\exp\left(\frac{Z(a,h,\ell_b,p,\bar{\ell}_c)}{\sigma_{cq}}\right)}{\sum_{\hat{\ell}_c} \exp\left(\frac{Z(a,h,\ell_b,p,\hat{\ell}_c)}{\sigma_{cq}}\right)} \tag{9}$$

$$V(a, h, \ell_b, p) = \sigma_{cq} \ln\left(\sum_{\hat{\ell}_c} \exp\left(\frac{Z(a, h, \ell_b, p, \hat{\ell}_c)}{\sigma_e}\right)\right) \tag{10}$$

Given the solution to equation (3), choice problems over work locations for college and high school graduates are given by,

$$V^{cw}(a,h,\ell_c) = \mathbb{E}_{\epsilon_{cw}}\Big[\max_{\ell_w}\Big\{W^c(a,h,\ell_c,\ell_w) - \delta(\ell_c,\ell_w) + \sigma_{cw}\epsilon_{cw}\Big\}\Big] \tag{11}$$

$$V^{nc}(a,h,\ell_b) = \mathbb{E}_{\epsilon_{ncw}}\Big[\max_{\ell_w}\Big\{W^{nc}(a,h,\ell_w) - \delta(\ell_b,\ell_w) + \sigma_{ncw}\epsilon_{ncw}\Big\}\Big] \tag{12}$$

The resulting choice probabilities and value functions are,

$$Pr[\ell_w = \bar{\ell}_w|a,h,\ell_c] = \frac{\exp\left(\frac{W^c(a,h,\ell_c,\bar{\ell}_w)-\delta(\ell_c,\bar{\ell}_w)}{\sigma_{cw}}\right)}{\sum_{\hat{\ell}_w}\exp\left(\frac{W^c(a,h,\ell_c,\hat{\ell}_w)-\delta(\ell_c,\hat{\ell}_w)}{\sigma_{cw}}\right)} \tag{13}$$

$$V^{cw}(a,h,\ell_c) = \sigma_{cw}\ln\left(\sum_{\hat{\ell}_w}\exp\Big(\frac{W^c(a,h,\ell_c,\hat{\ell}_w) - \delta(\ell_c,\hat{\ell}_w)}{\sigma_{cw}}\Big)\right) \tag{14}$$

$$Pr[\ell_w = \bar{\ell}_w|a,h,\ell_b] = \frac{\exp\left(\frac{W^{nc}(a,h,\bar{\ell}_w)-\delta(\ell_b,\bar{\ell}_w)}{\sigma_{ncw}}\right)}{\sum_{\hat{\ell}_w}\exp\left(\frac{W^{nc}(a,h,\hat{\ell}_w)-\delta(\ell_b,\hat{\ell}_w)}{\sigma_{ncw}}\right)} \tag{15}$$

$$V^{nc}(a,h,\ell_b) = \sigma_{ncw}\ln\left(\sum_{\hat{\ell}_w}\exp\Big(\frac{W^{nc}(a,h,\hat{\ell}_w) - \delta(\ell_b,\hat{\ell}_w)}{\sigma_{ncw}}\Big)\right) \tag{16}$$

Given the initial distribution and choices probabilities, one can derive the steady state distribution of individuals working in the economy. The mass of individuals of education type e_{nc} and e_c, ability a, human capital h, education location ℓ_c, and working location ℓ_w can be found by,

$$\mu(e_{nc},a,h,\ell_w) = Pr(\ell_b)\cdot F(a,h,p)\cdot Pr[e=nc|a,h,\ell_b,p]\cdot Pr[\ell_w=\bar{\ell}_w|a,h,\ell_b] \tag{17}$$

$$\begin{aligned}\mu(e_c,a,h,\ell_u,\ell_w) =& Pr(\ell_b)\cdot F(a,h,p)\cdot Pr[e=c|a,h,\ell_b,p]\\ &\cdot Pr[\ell_c=\bar{\ell}_c|a,h,\ell_b,p]\cdot Pr[\ell_w=\bar{\ell}_w|a,h,\ell_c]\end{aligned} \tag{18}$$

In a similar manner, I will use $\mu(\cdot)$ to denote the mass of individuals with some set of characteristics throughout this paper. When it does not cause confusion, and with a slight

abuse of notation, I will use $\mu(e, a, h, \ell_u, \ell_w)$ to represent the mass of either college or non-college educated workers, with a redundant state variable $\ell_c = \ell_w$ for non-college workers. Finally, since moves are not permitted across the lifecycle and human capital accumulates deterministically, I suppress j as a state in the mass of agents.

3.2 Colleges

Each location has one public college.[10] The quality Q_ℓ of each college is given by,

$$Q_\ell = \bar{q}_\ell \big(\bar{a}_\ell\big)^{\theta} \big(m_\ell\big)^{1-\theta} \tag{19}$$

where, $\bar{a}_\ell$ represents the average ability of the student body and is commonly referred to as "peer effects". θ controls the elasticity of college quality with respect to peer effects. Spending per student is given by m_ℓ. Lastly, $\bar{q}_\ell$ is a fixed college quality term representing intangible capital such as reputation or job placement networks. This location-specific term represents historical university quality discussed in *Section 2*.

Colleges set an admissions criterion that specifies a minimum student ability for acceptance as $\bar{z}$. The objective function of a college is lexicographic. Its first priority is to maximize enrollment until it hits capacity. Once capacity is met, the college's objective is to maximize quality, which implies setting the highest value of $\bar{z}$ that maintains full enrollment.

The remaining objects are taken as given by the college, and decided by the state policy maker. After subsidies, net tuition for in-state and out-of-state students is fixed at T^i_ℓ and $T^i_\ell + T^o_\ell$, respectively. Likewise capacity is fixed by the policymaker at $\mathcal{C}^i_\ell$ and $\mathcal{C}^o_\ell$. Given separate capacity for in and out-of-state students, colleges will set separate admission cutoffs $\bar{z}^i_\ell$ and $\bar{z}^o_\ell$.

[10] Abstracting here from private colleges is an innocuous assumption. First, the majority of students attending private colleges are out-of-state. Second, as seen in *Figure D.3* in *Appendix D* average out-of-state tuition exhibits significant overlap with private tuition. Hence, one can also interpret migrating out of state for college as attending a private college.

3.3 Firms

Production is performed in each location by a representative firm that combines labor inputs from workers with a college degree and those without a college degree. The production technology is given by the CES function,

$$Y_\ell = \left(\bar{A}(e_{nc}, \ell)Y(e_{nc}, \ell)^\eta + A(e_c, \ell)Y(e_c, \ell)^\eta\right)^{\frac{1}{\eta}} \tag{20}$$

The elasticity of substitution between the two types of labor is given by $\frac{1}{1-\eta}$. The exogenous skill-augmenting productivity for non-college labor is given by, $\bar{A}(e_{nc}, \ell)$. The skill-augmenting productivity for college labor is a function of college quality in location ℓ and a fixed component,

$$A(e_c, \ell) = \bar{A}(e_c, \ell)\big(Q_\ell\big)^\lambda \tag{21}$$

Here, λ governs the strength of spillovers from the state college to firms. The benchmark model will set $\lambda = 0$, and $\lambda > 0$ is considered as an extension.

Labor is aggregated according to,

$$Y(e_{nc}, \ell_w) = \sum_{j=1}^{J} \iint \mu(e_{nc}, a, h, \ell_w)h(e_{nc}, j, a, \ell_w)\, da\, dh \tag{22}$$

$$Y(e_c, \ell_w) = \sum_{j=2}^{J} \sum_{\ell_c} \iint \mu(e_c, a, h, \ell_u, \ell_w)h(e_c, j, a, \ell_c, \ell_w)\, da\, dh \tag{23}$$

3.4 Governments

The economy contains two types of governments. Each location features a state government and a federal government operates at the economy-wide level. State and federal governments

must finance S_ℓ and F in fixed expenditures on education using proportional taxes. In equilibrium, state and federal governments set taxes τ_ℓ^s and τ^f, respectively, such that,

$$\tau_\ell^s \Big(\sum_{\ell_c} \sum_{e} \int_a \int_h \mu(e, a, h, \ell_c, \ell_w) \Omega(e, \ell_w, h) \, da \, dh \Big) = S_\ell \tag{24}$$

$$\tau^f \Big(\sum_{\ell_w} \sum_{\ell_c} \sum_{e} \int_a \int_h \mu(e, a, h, \ell_c, \ell_w) \Omega\big(e, \ell_w, h\big) \, da \, dh \Big) = F \tag{25}$$

At the federal level, government expenditures consist of transfers to the state government. The federal government provides fraction $\pi_\ell \in [0, 1]$ of F to each locality, such that $\sum_\ell \pi_\ell = 1$. At the state level, government expenditures consist of three components related to colleges: (1) capacity $\mathcal{C}_\ell$, (2) subsidized tuition T_ℓ^i, T_ℓ^o, and (3) spending per student m_ℓ. The state government's accounting identity is given by,

$$\begin{aligned} K\big(\mathcal{C}_\ell, m_\ell, \ell\big) = T_\ell^i \int_a \int_h \int_p \mu(a, h, \ell_b = \ell, p, \ell_c = \ell) \, da \, dh \, dp \\ + \big(T_\ell^i + T_\ell^o\big) \sum_{\ell_b \neq \ell} \int_a \int_h \int_p \mu(a, h, \ell_b, p, \ell_c = \ell) \, da \, dh \, dp + S_\ell + \pi_\ell \cdot F \end{aligned} \tag{26}$$

3.5 Housing

Location ℓ_w has a fixed supply of housing $D(\ell_w)$. It is assumed that housing is owned by an outside landlord. Housing prices are determined by the solution to the household's static optimality problem, equation (2). That is, total spending on housing services is equal to ξ times total disposable income in a location,

$$p(\ell_w) D(\ell_w) = \xi \sum_{e} \sum_{\ell_c} \int_a \int_h \mu(e, a, h, \ell_c, \ell_w) \Omega\big(e, \ell_w, h\big) \, da \, dh \tag{27}$$

3.6 Equilibrium

Given model exogenous objects, an equilibrium consists of prices and allocations,

- College qualities: $\{Q_\ell\}_\ell$
- Admissions cut offs: $\{z^i_\ell, z^o_\ell\}_\ell$
- Wages: $\{\omega(e,\ell)\}_{e,\ell}$
- Firm labor demands: $\{Y(e,\ell)\}_{e,\ell}$
- Firm productivities: $\{A(e_c,\ell)\}_\ell$
- House prices: $\{p_\ell\}_\ell$
- State government tax rates: $\{\tau^s_\ell\}_\ell$
- Federal government tax rate: τ^f
- Value functions: $V(a,h,\ell_b,p), V^c(a,h,\ell_b,p), V^{nc}(a,h,\ell_b), V^{cw}(a,h,\ell_u)$, $W^{nc}(a,h,\ell_w), W^c(a,h,\ell_u,\ell_w)$
- Choice probabilities: $Pr[e=\bar{e}|a,h,\ell_b,p], Pr[\ell_c=\bar{\ell}_c|a,h,\ell_b,p]$, $Pr[\ell_w=\bar{\ell}_w|a,h,\ell_c], Pr[\ell_w=\bar{\ell}_w|a,h,\ell_b]$

such that,

1. Given college qualities, admissions cut offs, wages, and house prices, the value functions, and choice probabilities are consistent with agent optimization.
2. Labor demands, wages, and outputs are consistent with firm optimization problem.
3. Given choice probabilities, college qualities are consistent with college optimization.
4. State and federal government budgets balance.
5. $\forall \ell \in \mathfrak{L}$, markets clear,
   (a) Consumption goods markets.
   (b) Labor markets.
   (c) Admissions markets.
   (d) Housing markets.

3.6.1 Equilibrium Selection

Given the presence of peer effects, multiple equilibria may arise in this model. At the college level, I follow Epple et al. (2017) and Hendricks et al. (2021) and consider what are referred to as "hierarchical adherence". Computationally, I find that a stable and unique equilibrium exists in the relevant parameter space.

3.7 Frictions and Inefficiencies

There are several frictions and sources of inefficiency in this model that make room for welfare improving policies. I discuss those relevant to the main counterfactual exercise of studying optimal government policy.

First, moving costs act as frictions preventing perfectly competitive migration decisions. Hence the distribution of human capital and abilities relative to college qualities and firm productivities will not be optimal. Second, the distribution of fixed college historical productivities $\bar{q}$ need not be matched perfectly with the distribution of population, or firm productivities. That is, in a competitive equilibrium where the first-welfare theorem holds, $\bar{q}$ would be free to adjust across regions. Third, both the college and firm production function features externalities. Agents do not internalize the effect that their own ability has on college quality. Fourth, the college market is not complete. There is not a competitive market allocation for tuition, spending per student or capacity. These are fixed by the state government.

Note, the key source of inefficiency in most quantitative spatial models are agglomeration and congestion forces. Given my current framework, I am able to abstract from agglomeration and do not rely on them for any key results.

3.8 Government Problems

In this section I lay out the government problems which determine optimal state and federal level decisions. These problems will be used for the main counterfactual exercises presented in *Section 6*. I allow for choices over spending per student, capacity, in-state, and out-of-state tuition in order to address the main sources of cross-state heterogeneity observed in the data and presented in *Section 2*.

Note, neither state nor federal problems presented here are social planner allocations. The government in these problems is not able to choose all allocations as a planner would. In particular, the government is unable to dictate, which agents attend college, where each agent attends college, and where each agent is employed. Given peer-effects (and possible college spillovers), the solution to the government problems may see welfare improvements over the competitive equilibrium, but will not be the welfare maximizing allocation of a social planner.

The cost of capacity faced by both governments is given by the function,

$$K\left(\mathcal{C}_\ell, m_\ell, \ell\right) = \bar{f} + \mathcal{V}(\mathcal{C}_\ell, \ell) + m_\ell \mathcal{C}_\ell \tag{28}$$

where $\bar{f}$ represents fixed costs, and $\mathcal{V}(\mathcal{C}_\ell, \ell)$ denotes variable costs. Note that variable costs depend on a college's location through land/housing prices. Fixed and variables costs are independent of college quality and I will refer to them as "custodial costs". The last term represents the sum of spending per student m_ℓ, again fixed by the policymaker.

3.8.1 State Government

Each state solves its optimal decision problem taking as given the choices of all other states. I solve for an equilibrium where state's decisions are consistent with each other. That is, given all other state's actions, no individual state has a profitable deviation.

The state government l, chooses capacity $\mathcal{C}_l$, spending per student m_l, in-state tuition

T_l^i, and out-of-state tuition T_l^o, in order to maximize aggregate utilitarian welfare of agents born in state l, by solving,

$$\max_{\{\mathcal{C}_l, m_l, T_l^i, T_l^o\}} \int_a \int_h \int_p \mu(a, h, p, \ell_b = l) V(a, h, p, \ell_b = l)\, da\, dh\, dp$$

subject to, (29)

$$K\big(\mathcal{C}_l, m_l, l\big) = T_l^i \int_a \int_h \int_p \mu(a, h, \ell_b = l, p, \ell_c = l)\, da\, dh\, dp$$
$$+ \big(T_l^i + T_l^o\big) \sum_{\ell_b \neq l} \int_a \int_h \int_p \mu(a, h, \ell_b, p, \ell_c = l)\, da\, dh\, dp + S_l + \pi_l \cdot F$$

$$Q_l = \bar{q}_l \big(\bar{a}_l\big)^{\theta} \big(m_l\big)^{1-\theta}$$

$$\int_a \int_h \int_p \mu(a, h, \ell_b = l, p, \ell_c = l)\, da\, dh\, dp + \sum_{\ell_b \neq l} \int_a \int_h \int_p \mu(a, h, \ell_b, p, \ell_c = l)\, da\, dh\, dp \leq \mathcal{C}_l$$

+ equilibrium conditions

Total state expenditures are given by cost of capacity and spending per student $K(\mathcal{C}_l, m_l, l)$. Revenues consist of in-state and out-of-state tuition, a fixed state spending S_l generated by equilibrium labor taxes τ_l^s, and some fraction π_l of fixed federal expenditures F.

The second constraint is the quality production function and the third constraint is the capacity constraint. The government's problem must also be consistent with all conditions which characterize a competitive equilibrium as defined in *Section 3.6*.

An alternative to maximizing the welfare of those born in state l is to maximize the welfare of those living in state l. I chose the former as the latter leads the state government to counterfactual choices with $T_l^o < T_l^i$ in order to attract high-ability out-of-state applicants. While this could also occur under the chosen specification if peer effects were strong enough, it does not arise in the calibrated model.

3.8.2 Federal Government

The federal government's problem is the natural extension of equation (29) where now the vector of capacities, spending per student, and tuition are chosen jointly to solve,

$$\max_{\{\mathcal{C}_\ell, m_\ell, T^i_\ell, T^o_\ell\}_{\ell \in \mathfrak{L}}} \sum_{\ell_b} \int_a \int_h \int_p \mu(a, h, p, \ell_b) V(a, h, p, \ell_b)\, da\, dh\, dp$$

subject to, (30)

$$\sum_{\ell} K\big(\mathcal{C}_\ell, m_\ell, \ell\big) = \sum_{\ell_c} \Big(T^i(\ell_c) \int_a \int_h \int_p \mu(a, h, \ell_b = \ell_c, p, \ell_c)\, da\, dh\, dp \Big)$$

$$\sum_{\ell_c} \Big(\big(T^i(\ell_c) + T^o(\ell_c)\big) \sum_{\ell_b \neq \ell_c} \int_a \int_h \int_p \mu(a, h, \ell_b, p, \ell_c)\, da\, dh\, dp \Big) + \sum_{\ell} S_\ell + F$$

$$Q_\ell = \bar{q}_\ell \big(\bar{a}_\ell\big)^\theta \big(m_\ell\big)^{1-\theta} \qquad \forall \ell \in \mathfrak{L}$$

$$\sum_{\ell_b} \int_a \int_h \int_p \mu(a, h, \ell_b, p, \ell_c)\, da\, dh\, dp \leq \mathcal{C}(\ell_c) \qquad \forall \ell_c \in \mathfrak{L}$$

$+$ equilibrium conditions

Notice here that the federal government is appropriating the sum of state revenues generated by labor taxes τ^s_ℓ. I formulate the federal problem in this way to isolate welfare changes exclusively due to the system of higher education and not any changes in tax distortions.

4 Properties of the Model and Equilibrium

I now numerically demonstrate several properties of the equilibrium in the model, with a focus on the optimal policy choices of states, given model primitives. I then provide economic intuition for the model mechanisms and the motivating facts presented in *Section 2*.

The model primitives that affect the optimal decision over policy choices are: (1) population size, (2) housing stock (relative to population), (3) college and non-college labor productivity, (4) fixed college productivity, (5) federal expenditures on higher education, (6) state expenditures on higher education, and (7) geographical location. These primitives lead to a large degree of heterogeneity in the model solution to the states' optimal problem and the resulting equilibrium, as observed in U.S. data.

In this section, I focus on (1) through (3), as in the empirical analysis. Item (4) is not observable in the data, but its implications will be discussed in the context of the counterfactual federal government's problem in *Section 6*. I will also discuss items (5) through (7) in *Section 6*, before conducting the main welfare analysis.

Consider a stylized model with five states of three different types.[11] First, a large and high-productivity (high-wages) state, such as California or New York. Second, a small and low-productivity (low-wages) state, such as North Dakota or Mississippi. Call these high-type and low-type states, respectively. Finally, three mid-type states of median size and productivity, such as Arizona or Wisconsin. I begin by solving the optimal policy problem (equation (29)) for each state type. I then conduct a numerical comparative static by perturbing the optimal state solution (and resulting equilibrium) with respect to college productivity and state population for the low-type and high-type states separately.[12]

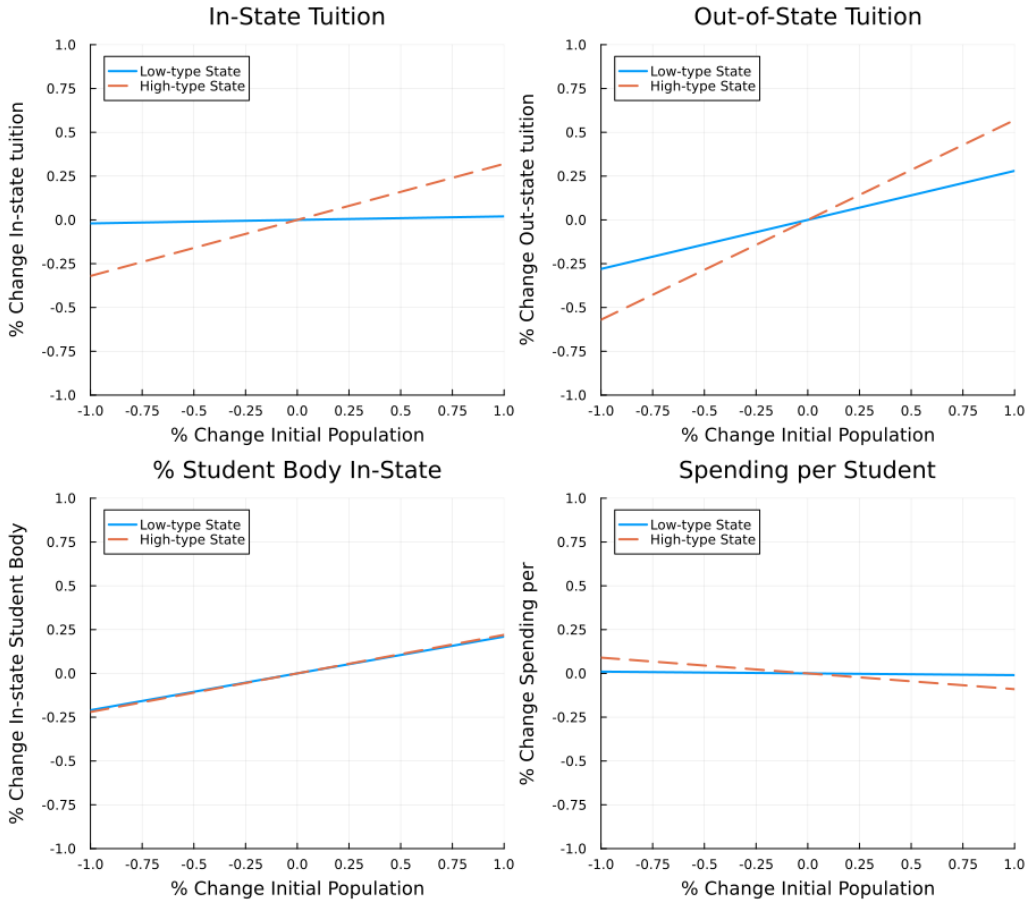
Figures 8 and *9* report the results of this exercise. The model does well at qualitatively replicating the motivating correlations presented in *Section 2*. From *Figure 8* when initial population levels rise, so do in-/out-of-state tuition and the percentage of the student body that is in-state. Spending per student falls, although this correlation in the data is not precisely estimated.

Under the standard assumption of a convex cost function, these changes in the model are

[11] The qualitative dynamics shown here also hold for the full 50-state calibrated model.

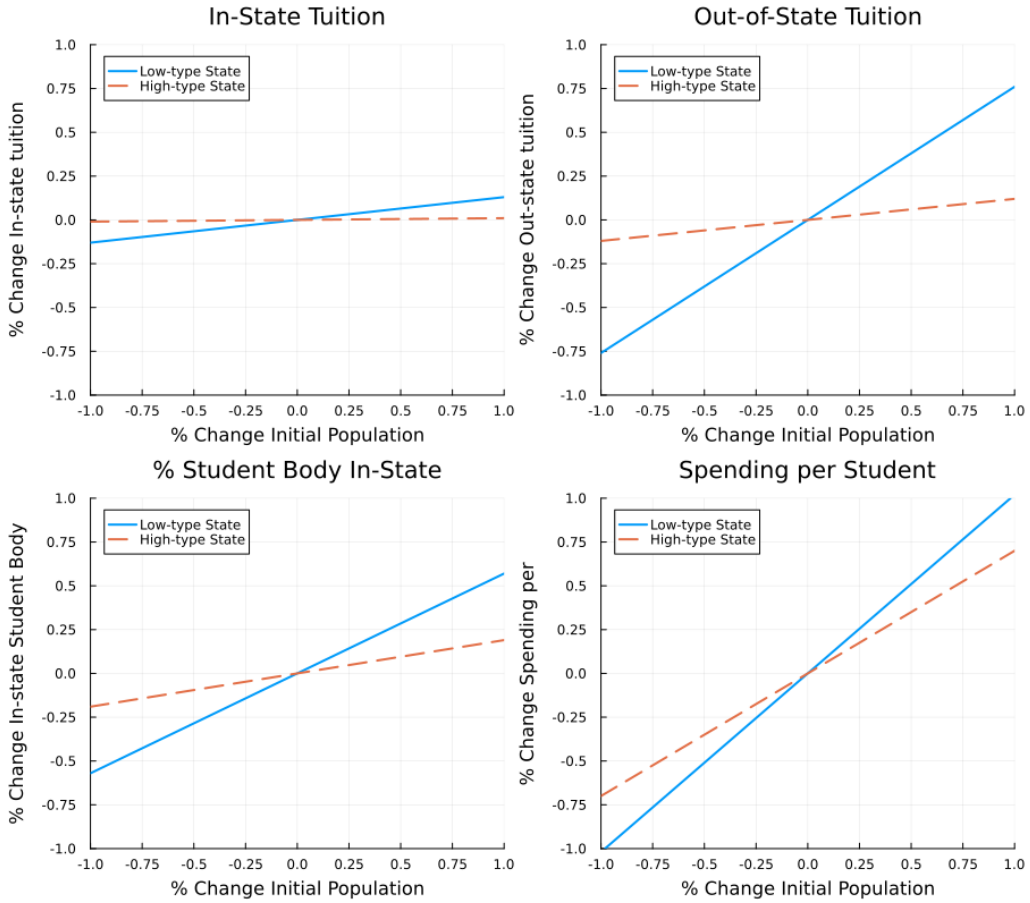
[12] I plot this for small changes. For small values this is equivalent to a central finite difference derivative of the solution to the state's problem. That is, denoting the solution to problem (29) as $x^* = [\mathcal{C}_i^*, \mathcal{C}_o^*, T_i^*, T_i^*, m_i^*]$, the numerical derivative is $\frac{dx^*}{d\bar{A}_c}$ and $\frac{dx^*}{d\mathcal{M}}$. When taking changes with respect to $\bar{A}_c$, I convert to changes in college wages ω_c. When taking the changes with respect to initial population, I adjust the housing stock so that the ratio of population to housing remains unchanged.

Figure 8: **Changes in Optimal Policy Choices with Respect to Initial Population**



Notes: This figure plots the numerical comparative statics for a percentage point change in state optimal policy parameters due to a percentage point change in initial population size.

Figure 9: **Changes in Optimal Policy Choices with Respect to College Wages**



Notes: This figure plots the numerical comparative statics for a percentage point change in state optimal policy parameters due to a percentage point change in college labor productivity (college wages).

driven by scale economies. When population increases, the relative value of an out-of-state student decreases compared to spending per student or reducing in-state tuition. Hence, the state admits relatively fewer out-of-state students at a higher price. The decline in spending per student approaches zero as the cost function approaches linearity. The magnitude of changes around the optimum is also larger for high-type states, as they are further along the respective cost curve.

From *Figure 9* when wages rise, so do in-/out-of-state tuition, the percentage of students who are in-state, and spending per student, consistent with the empirical correlations. The model intuition for this result is as follows. When wages are low (high), the percentage of alumni remaining in-state and the net import of college graduates is low (high). A low-type state will attract out-of-state students by setting tuition low and capacity high. Some of these out-of-state students will then remain in-state after graduation. On the other hand, college graduates migrate in large numbers to high-type states, regardless of whether they are incentivized to attend college in the state. Hence, it is more beneficial for a high-type state to spend tax revenues on in-state students than on attracting out-of-state students.

The dynamics of alumni remaining to work can be seen in *Figure D.18*, where I plot the change in the percentage of alumni who stay to work in-state (top panel) and the net import of college graduates (bottom panel) as college skill prices are exogenously changed. Both lines are proportional to changes in skill prices.

Finally, to see migration patterns with respect to out-of-state college capacity, *Figure D.19* takes the competitive equilibrium and exogenously increases out-of-state capacity in low- and high-type states by an equal (relative) amount. The resulting increase in the equilibrium population of college graduates working in each state is approximately three-fold higher in low-type than in high-type states. Moreover, due to scale economies, the decrease in spending per student in the low-type state is one-quarter of that in the high-type state.

5 Calibration

The calibration involves choosing parameters so that the model replicates a steady-state competitive equilibrium of the U.S. economy. I calibrate the model to the 50 U.S. states for the 2015-2019 time period. For households the model is calibrated to match college decisions consistent with high school graduates by location and parental income and the ensuing migration decisions and earnings dynamics by college attendance and ability. Importantly to the household and policymaker I ensure the model matches the elasticity of earnings to college expenditures. After calibration, I externally validate the model using out of sample predictions for key elasticities central to the policymakers optimal problem.

The main counterfactual will then involve computing an economy in which state or federal governments optimally choose tuition levels, spending per student, and college capacities. I compare the welfare of the calibrated benchmark economy to these counterfactual economies. This section describes the estimation of all external and internal parameters needed to conduct this exercise.

Table 5 summarizes model parameters that are externally calibrated from the data and *Table 6* summarizes the internally calibrated parameters and parameters derived from model inversion. Each section below describes a block of the economy, in which I discuss both externally and internally calibrated parameters. Details are discussed throughout this section and additional information is reported in *Appendix B*. All moments and parameters estimated for each individual state are omitted here and reported in *Appendix D*.

The calibration follows a two-step procedure. The first step is conducted in partial equilibrium, with skill prices, housing rental rates, and taxes taken from the data. Agent preference parameters, the initial distribution, and college parameters are chosen to match choice probabilities and human capital accumulation dynamics. In the second step, I recover the remaining firm, college, and housing parameters by inverting structural equations from the model.

Table 5: **Externally Calibrated Parameters**

Parameter	Value	Description	Source
Agents			
β	0.98	Time discounting factor	Annual interest rate, 0.02
J	8	Number of model periods	5-year periods, age 18-58
ξ	0.33	Housing share of total expenditures	BLS CEX
$\mu(\ell_b)$	–	Mass of high school students in state ℓ	ACS
μ_h	0.99	Human capital distribution, mean	ACS
σ_h	0.56	Human capital distribution, s.d.	ACS
ν	0.50	Human capital production	Huggett et al. (2011)
δ	0.02	Human capital depreciation	Huggett et al. (2011)
Colleges			
$\mathcal{C}_\ell^i, \mathcal{C}_\ell^o$	–	In- and out-of state capacity	IPEDS
T_ℓ^i, T_ℓ^o	–	In- and out-of-state tution	IPEDS
$\bar{f}$	2.67×10^7	College custodial fixed costs	IPEDS and NHGIS
κ_1	11,817	College custodial FTE linear term	IPEDS and NHGIS
κ_2	0.219	College custodial FTE quadratic term	IPEDS and NHGIS
κ_3	12.37	College custodial FTE×rent interaction term	IPEDS and NHGIS
Firms			
η	0.31	College to non-college CES elasticity	Katz and Murphy (1992)
Government			
S_ℓ	–	State college expenditures in state ℓ	IPEDS
$\pi_\ell F$	–	Federal college expenditures in state ℓ	IPEDS

Notes: This table reports model parameters directly observed and calculated from the data. Values that are not calculated by state are reported here. Values calculated for each state individually are reported in *Appendix D*.

Table 6: **Internally Calibrated Parameters**

Parameter	Value	Description	Source
Preferences			
γ_1^e	0.0012	Moving cost, education	Choice probabilities
γ_2^e	-2.1179	Home bias, education	Choice probabilities
γ_1^w	0.0004	Moving cost, working	Choice probabilities
γ_2^w	9.5459	Home bias, working	Choice probabilities
γ_3^w	4.9769	Unskilled labor, working	Choice probabilities
σ_e	0.1502	Dispersion, education	Choice probabilities
σ_{cq}	0.8910	Dispersion, college location	Choice probabilities
σ_{cw}	1.7335	Dispersion, college work	Choice probabilities
σ_{ncw}	2.5625	Dispersion, non-college work	Choice probabilities
Distributions			
μ_a	0.3176	Ability, mean	Average earnings profile
σ_a	0.5751	Ability, s.d.	Average earnings variance
μ_p	1.0248	Parent resources, mean	College by parent income
σ_p	1.5479	Parent resources, s.d.	College by parent income
Colleges			
θ	0.6112	Elasticity of earnings w.r.t. spending	IPEDS, Grapevine, and Deming and Walters (2017)
$\bar{q}_\ell$	–	College fixed quality college	Earnings profiles by quality. Chetty et al. (2020)
Firms			
$\bar{A}(e,\ell)$	–	Firm fixed productivity	Firm prod. function, skill prices.
Housing			
H_ℓ	–	Housing stock	Housing market clearing, rent price

Notes: This table reports internally calibrated model parameters, their value, a brief description, and the moment target. Values that are not calculated by state are reported here. Values calculated for each state individually are reported in *Appendix D*. Preference parameters, initial distribution parameters the elasticity of earnings with respect to college spending per student are estimated by GMM. The remaining sets of parameters are estimated by model inversion.

5.1 Agents

The set of parameters here is, $[\gamma_1^e, \gamma_2^e, \gamma_1^w, \gamma_2^w, \gamma_3^w, \sigma_e, \sigma_{cq}, \sigma_{cw}, \sigma_{ncw}]$, which consist of the moving cost and dispersion parameters. Agent preference parameters are estimated so that the choice probabilities in the model are closest to those in the data. Given that the sum of probabilities equals one, the ln-odds ratio is used for estimation. From equations (5), (9), (13), and (15), the odds ratios can be derived as follows,

$$\ln\left(\frac{Pr[e=c|a,h,\ell_b,p]}{Pr[e=nc|a,h,\ell_b,p]}\right) = \frac{1}{\sigma_e}\left(V^c(a,h,\ell_b,p) - V^{nc}(a,h,\ell_b)\right) \tag{31}$$

$$\begin{aligned}&\ln\left(\frac{Pr[\ell_c=\bar{\ell}_c|a,h,\ell_b,p]}{Pr[\ell_c=1|a,h,\ell_b,p]}\right)\\ &=\frac{1}{\sigma_{cq}}\Bigg(\ln\left(p-T^i(\ell_c)-T^o(\ell_c)\mathbb{1}_{\ell_b\neq\ell_c}\right)-\delta(\ell_b,\ell_c)+\beta V^{cw}(a,h,\ell_c)\\ &\quad-\left(\ln\left(p-T^i(1)-T^o(1)\mathbb{1}_{\ell_b\neq 1}\right)-\delta(\ell_b,1)+\beta V^{cw}(a,h,1)\right)\Bigg)\end{aligned} \tag{32}$$

$$\begin{aligned}&\ln\left(\frac{Pr[\ell_w=\bar{\ell}_w|a,h,\ell_c]}{Pr[\ell_w=1|a,h,\ell_c]}\right)\\ &=\frac{1}{\sigma_{cw}}\left(W^c(a,h,\ell_c,\bar{\ell}_w)-\delta(\ell_c,\bar{\ell}_w)-\left(W^c(a,h,\ell_c,1)-\delta(\ell_c,1)\right)\right)\end{aligned} \tag{33}$$

$$\begin{aligned}&\ln\left(\frac{Pr[\ell_w=\bar{\ell}_w|a,h,\ell_b]}{Pr[\ell_w=1|a,h,\ell_b]}\right)\\ &=\frac{1}{\sigma_{ncw}}\left(W^{nc}(a,h,\bar{\ell}_w)-\delta(\ell_b,\bar{\ell}_w)-\left(W^{nc}(a,h,1)-\delta(\ell_b,1)\right)\right)\end{aligned} \tag{34}$$

Typically, quantitative spatial models feature location- (and possibly education-) specific

amenity values that enter the preferences of agents, allowing the model to achieve an exact fit to the data. In the quantitative model developed in the previous section, I omit location amenity values, and so the model is over-identified. Matching the choice probabilities instead works entirely through earnings (human capital accumulation), initial distributions, moving costs, and preference shocks.

The data identifying choice probabilities originate from several different sources. The education decision, equation (5), is taken from the geocoded National Longitudinal Survey of Youth 1997 (NLSY97). Ability is defined using Armed Services Vocational Aptitude Battery (ASVAB) scores. Human capital is defined as in Wright and Wurdinger (2025), using the GPA-weighted stock of high school credits.

The choice of university location, equation (9), combines data from the NLSY97 and IPEDS. Due to sample size limitations, I cannot directly estimate this choice probability using NLSY97 data. Instead, I first calculate transition probabilities from state of birth to state of college degree using IPEDS. Second, I assign each college a quality in the NLSY97 using the average later-life earnings of those attending the given college. This is the natural definition which closely aligns with model college quality. I group colleges into deciles of quality and obtain transition probabilities to these college qualities, *contingent* on ability, human capital, and parental resources. Finally, I note that for the full state-by-state transition matrix obtained using IPEDS data, any difference in transition probabilities for states with the same quality college must be due to moving costs or idiosyncratic preference shocks. Hence, I take differences within quality deciles and add them to the transition probabilities obtained from the NLSY97.

For the transition probabilities from college location to working location for each state, I use the Baccalaureate and Beyond (B&B) 2016/2020 survey. I then use the same procedure in the NLSY97 as just described, but group states based on the skill price. Finally, for the transition probabilities from birth location to working location for non–college educated workers, I use the American Community Survey (ACS) and again use the NLSY97, with

groupings based on state skill prices. I discuss calculation of skill prices using the Panel Study of Income Dynamics (PSID) in *Section 5.3*.

5.1.1 Demographics

The demographic structure of agents is uniform across the lifecycle. It would be a simple extension of the model to allow for a different mass of agents at different ages. However, demographic variation is not central to the main research question addressed in this paper. The initial mass of agents in each state is given by the high school population, taken from the 2015-2019 5-year ACS sample.

5.1.2 Preferences

The preference parameters affecting household decisions over college and working locations are the preference shocks for each discrete decision: σ_e, σ_{cq}, σ_{cw}, and σ_{ncw}, and the moving costs. Moving costs for the education migration decision are parameterized as,

$$\delta^c(\ell_b, \ell_c) = \gamma_1^e \cdot d(\ell_b, \ell_c) + \gamma_2^e \cdot \mathbb{1}(\ell_b \neq \ell_c) \tag{35}$$

where $d(\cdot, \cdot)$ is the Euclidean metric measuring the distance between birth location and education location. γ_1^e captures moving costs, and γ_2^e captures home bias, or the extra utility received by an individual studying in their home state. The moving costs associated with the work migration decision are given by,

$$\delta^w(\ell, \ell_w, e) = \gamma_1^w \cdot d(\ell, \ell_w) + \gamma_2^w \cdot \mathbb{1}(\ell_b \neq \ell_c) + \gamma_3^w \cdot \mathbb{1}(e = nc) \tag{36}$$

where $\ell = \ell_b$ for a non–college educated worker ($e = nc$), and $\ell = \ell_c$ for a college educated worker ($e = c$). γ_3^w captures a set of unobservable factors affecting the mobility of non-college educated workers, given the well-known observation that college educated workers have a much higher degree of geographic mobility.

Birth and work locations are defined as the population-weighted centroid of a given state, and education locations as the college student weighted centroid of the state. The former is calculated using population-weighted centroids for each block group from the U.S. Census. The latter is calculated using geographical coordinates and student counts provided for each college in IPEDS.

5.1.3 Initial Distributions

There are six parameters governing the variance and mean of initial distributions for ability, human capital, and parental resources: σ_a, μ_a, σ_h, μ_h, σ_p, and μ_p. The initial distribution of human capital is taken from the mean and variance of initial earnings, either when graduating high school. The initial distribution of ability is internally calibrated to match average earnings growth rates, and average earnings variance over the lifecycle. Finally, the initial distribution of parental resources targets college attendance rates by parental income quintiles.

5.1.4 Human Capital Production

I use the same set of moments defining earnings growth by ability and college quality from the first step of the calibration process to estimate the function determining human capital accumulation. Recall, this function is given by,

$$h_{j+1} = \big(a \cdot Q\big)h_j^{\nu} + (1-\delta)h_j \tag{37}$$

Period depreciation and the curvature of human capital production are taken to be standard values from the literature, with $\delta = 0.02$ and $\nu = 0.50$. I discuss the estimation of college quality Q and college production function below.

5.2 Colleges

For college calibration, I focus on public, primarily four-year, degree-granting institutions with at least 250 students enrolled. All IPEDS data are for the 2017-2018 academic year. This amounts to 572 colleges across the 50 U.S. states. I choose this time period in order to avoid large changes to the higher education sector during the COVID-19 pandemic. Data details are reported in *Appendix A*.

5.2.1 Tuition

I construct a measure of net tuition price for in-state and out-of-state students. IPEDS includes an estimate of net price for each college; however, it is not reported separately for in-state and out-of-state students. Instead, I use gross sticker tuition and subtract average federal and state tuition grants per student. At the federal level, this includes Pell Grants, which I assume are awarded uniformly to both in-state and out-of-state students. Federal Title IV funding is not included in the college/state budget constraint; however, it is included in the calculation of net tuition price. That is, while Title IV funds affect the net tuition price, these funds are not readily available for colleges to spend on other purposes, as is the case with direct federal funding.

I account for several tuition reciprocity schemes. Most states are not part of a reciprocity agreement. For those that are, agreements differ by region. For example, the New England Regional Student Program includes Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont. It provides an average per-semester tuition discount of $4,300 to out-of-state students within this grouping. However, no allowances are made for stricter out-of-state admission requirements. On the other hand, Minnesota, Wisconsin, and North Dakota have a reciprocity agreement under which out-of-state students pay in-state tuition and face the same admissions requirements as in-state students. *Appendix A* reports details of all current reciprocity agreements.

5.2.2 Spending per Student

For spending per student, I follow Epple et al. (2006) and consider m_ℓ to consist of "quality-enhancing" expenditures. This is in contrast to custodial costs of operating at size $\mathcal{C}_\ell$ in location ℓ, which are discussed and parameterized in the next section. Inputs to expenditures per student include instructional spending in the benchmark calibration. I perform robustness analysis using broader definitions, which are reported in *Appendix B*.

5.2.3 Costs

Recall from equation (26) that the state government's budget constraint can be summarized as,

$$\bar{f} + \mathcal{V}(\mathcal{C}_\ell, \ell) + m_\ell \mathcal{C}_\ell = T_\ell + S_\ell \pi_\ell \cdot F \tag{38}$$

where m_ℓ is spending per student, as discussed in the previous section. "Custodial costs" are given by a fixed cost $\bar{f}$ and a variable cost function $\mathcal{V}(\mathcal{C}_\ell, \ell)$. Revenues are then a function of total tuition revenue T_ℓ and total state and federal appropriations.[13] Given observable data, the budget constraint then provides a direct expression for custodial costs, which I denote as K^c_ℓ.[14] Custodial costs are estimated following a similar approach to Epple et al. (2006). Assume they can be parameterized as,

$$K^c_{i,\ell} = \bar{f} + \kappa_1 \mathcal{C}_{i,\ell} + \kappa_2 \mathcal{C}^2_{i,\ell} + \kappa_3 (p_\ell \cdot \mathcal{C}_{i,\ell}) + \epsilon_{i,\ell} \tag{39}$$

The identification assumption necessary to estimate each parameter using ordinary least squares (OLS) is that measurement error in custodial costs, $\epsilon_{i,\ell}$ (at college i in state ℓ),

[13] While colleges do receive revenues from gifts, endowment income, or other ventures, an investigation of heterogeneity across states along these categories is beyond the scope of the present paper. Moreover, *Figure D.7* in *Appendix D* shows that tuition and government revenues represent the vast majority of the budget constraint across all states.

[14] In the benchmark calibrated U.S. economy, this equation is simply an accounting identity. It becomes a relevant constraint in the context of problem (29) when the state government chooses an optimal level of tuition, spending per student, and capacity.

is uncorrelated with college size. The industrial organization literature takes a different view and interprets $\epsilon_{i,\ell}$ as an unobserved shock to the cost function (Berry, 1994). Under this interpretation, it is likely that $\mathbb{E}[\epsilon_{i,\ell} \mid \mathcal{C}_{i,\ell}] \neq 0$. To overcome this issue, I employ an instrumental variable (IV) strategy, where college size is instrumented with endowment level.

College capacity (size), $\mathcal{C}_{i,\ell}$, is taken to be the total full-time equivalent (FTE) undergraduate enrollment. *Section 5.4* estimates p_ℓ as the average rent price across each state, as is common in this class of models. While rent prices are the relevant measure for scaling student residential buildings, land prices would be a better measure for other types of physical expansion. However, rent prices are highly correlated with land prices and serve as a reasonable proxy in equation (39) to capture cross-state differences in costs. I do not estimate a separate cost function for each state and assume that there are no unobserved cost differences across colleges.

Results from estimating equation (39) are reported in *Table 7*. Under all specifications, the cost function is convex in the number of students. Column (1) reports the estimation omitting the interaction between college size and rent prices. Column (2) adds this interaction. Finally, Column (3) reports the results of the IV regression.

Throughout, fixed costs (the regression constant) are large but not precisely estimated. The linear and quadratic terms, κ_1 and κ_2 on college size are stable across specifications but lose power under the IV regression. Finally, the effect of rent prices, κ_3 on college costs is large and an important predictor of custodial costs.

5.2.4 Capacity

For capacity, I use the total count of full-time equivalent students enrolled, disaggregated by in-state and out-of-state status. I do not count international students as out-of-state and instead drop them from the analysis. International student enrollment has risen dramatically over the last 20 years. However, they still represent a small share of the total student body, at less than 4% on average.

Table 7: **Custodial Cost Function Estimation**

	(1) Custodial Costs	(2) Custodial Costs	(3) Custodial Costs
κ_1	10,084	8,582	11,817
	(0.000)	(0.000)	(0.271)
κ_2	0.314	0.311	0.219
	(0.000)	(0.000)	(0.210)
κ_3		1.34	12.37
		(0.264)	(0.000)
$\bar{f}$	1.44×10^7	1.52×10^7	2.67×10^7
	(0.236)	(0.212)	(0.520)
Observations	543	543	505
IV	No	No	Yes

Notes: This table reports the results of the custodial cost function estimation. Column (3) instruments for full-time equivalent students using endowment levels. P-values are reported in brackets. Data are taken from IPEDS for college variables and from NHGIS for local rent prices.

5.2.5 Production

College quality fixed component $\bar{q}$ is chosen to match earnings growth profiles over the years 23-35. With data taken from Chetty et al. (2020). This implies a college quality which is then matched in the second step of the estimation procedure.

The remaining parameter to calibrate for colleges in the second stage is θ, the elasticity of college quality with respect to spending per student. θ is internally calibrated to match the elasticity of median graduate earnings with respect to college spending per student. θ is calibrated after the two-step procedure, so that the counterfactual economy when increasing spending per student matches changes graduate earnings. This introduces a fixed point in the calibration process, I first guess θ, and then update after the calibration procedure is completed. I then re-calibrate the model, and update θ until convergence is achieved.

In the data, the estimate of interest for college i at time t is,

$$y_{i,t} = \beta \ln\left(spend\ per_{i,t}\right) + \gamma X_{i,t} + \delta_t + \lambda_i + \epsilon_{i,t} \tag{40}$$

where $y_{i,t}$ is ln-median income of cohort t eight to ten years post graduation, $X_{i,t}$ is a vector of college controls, δ_t are time fixed effects, and λ_i are college fixed effects. I use IPEDS data for for each variable. Earnings data is available for the years 1997-2008. To be consistent with the college cost function estimation, I use total instructional spending per FTE as spending per student.

Estimating equation (40) through OLS would likely result in a biased estimate of the coefficient of interest, β, due to unobserved individual-level heterogeneity in ability. Instead, I employ an instrumental variables (IV) strategy similar to Deming and Walters (2017) and instrument for ln-spending per student, $\ln\left(spend\ per_{i,t}\right)$.

I consider two instruments. The first is a shift-share (i.e., Bartik, 1991) style instrument based on shocks to state budgets. The idea is that state budget cuts made "across the board" will have a larger impact on colleges where state appropriations represent a greater share of the total budget.[15] I construct the instrument using initial 1997 shares of total college revenues accounted for by state appropriations, using IPEDS data. I then interact this with yearly state appropriations allocated to higher education, per college-aged (19-23) person. Yearly state appropriations for higher education are taken from Grapevine, and the population of college-aged individuals is taken from the Current Population Survey (CPS). That is, the instrument is given by,

$$Z^b_{i,t} = \left(\frac{approp_{i,1997}}{rev_{i,1997}}\right) \cdot \left(\frac{state\ approp_{s,t}}{pop_{s,t}}\right) \tag{41}$$

where it is implicit in the notation that state s is the state in which college i is located.

The second instrument uses variation in tuition increase caps or freezes across states. I use data compiled by Deming and Walters (2017), again for the period 1997-2008. Over this

[15] State budget cuts are typically (though not always) applied across all colleges. See Deming and Walters (2017) for a full discussion of state budgetary processes.

Table 8: **Elasticity of Graduate Earnings to Spending per Student**

	(1) ln(median earnings)	(2) ln(median earnings)
spend per	0.173	0.182
	(0.000)	(0.000)
Observations	2,826	2,865
F-stat	267.1	220.3
Controls	Yes	Yes
Time Effects	Yes	Yes
Fixed Effects	Yes	Yes

Notes: This table reports the results of the IV regressions estimating the elasticity of college graduates' median earnings with respect to total instructional spending per student. Earnings are measured eight to ten years after graduation. The first column reports results using the shift-share instrument, and the second column reports results using the tuition caps instrument. Data are taken from IPEDS (1997-2008), Grapevine, the CPS, and Deming and Walters (2017).

period, seventeen states enacted legislation capping or freezing tuition increases at all public colleges. I define the instrument $Z^f_{i,t}$ to equal one if the state instituted a tuition increase freeze or cap between periods $t-1$ and t, and zero otherwise.

The results of each instrumental variables regression are reported in *Table 8*. Column (1) reports results using the shift-share instrument, and Column (2) reports results using the tuition freeze/cap instrument. Both coefficients are of similar magnitudes and are highly significant. Together, they indicate that a 1% increase in spending per student translates into a 0.18% increase in median earnings of graduates.

5.3 Firms

Consider first the benchmark model with no spillovers ($\lambda = 0$). The elasticity of substitution between college- and non-college educated labor, η, is externally calibrated to $0.31 = 1 - 1/1.44$, following Katz and Murphy (1992) and Heckman et al. (1998). Firm fixed productivities, $\bar{A}(e_{nc}, \ell)$ and $\bar{A}(e_c, \ell)$, are then obtained by inverting the production function.

Defining total output in location ℓ as Y_ℓ, the solution to the firm's profit maximization problem gives,

$$\omega(e,\ell)Y(e,\ell) = \left(\frac{\omega(e,\ell)}{A(e,\ell)}\right)^{\frac{\eta}{\eta-1}} Y_\ell \tag{42}$$

ln-linearizing, productivity of education type e can be expressed as,

$$\begin{aligned}\ln\big(A(e,\ell)\big) = \frac{1}{\eta}\ln\big(\omega(e,\ell)\big) + \frac{1-\eta}{\eta}\ln\big(Y(e,\ell)\big) \\ - \frac{1-\eta}{\eta}\ln\big(\omega(e_{nc},\ell)Y(e_{nc},\ell) + \omega(e_c,\ell)Y(e_c,\ell)\big)\end{aligned} \tag{43}$$

Given data for skill prices, $\omega(e,\ell)$, and labor supplies, $Y(e,\ell)$ (implied by calibrated choice probabilities), fixed productivities $A(e,\ell)$ can be recovered.

Skill prices paid by firms in the model, $\omega(e,\ell_w)$, are not directly observable in the data. Earnings observable in the data are given by $\omega(e,\ell_w)h$ in the model. However, I use yearly migration decisions of college- and non-college educated workers in the Panel Study of Income Dynamics (PSID) to infer skill prices. That is, assuming small yearly changes in human capital accumulation (which is empirically verifiable), any difference in income post-migration must be due to differences in skill prices. Hence, I regress ln-differences in yearly earning, $y_{i,s,t}$, for those who migrated across states, on time effects δ_t, state fixed-effects λ_s, and a vector of demographic controls $X_{i,s,t}$,

$$y_{i,s,t} = \beta X_{i,s,t} + \delta_t + \lambda_s + \epsilon_{i,s,t} \tag{44}$$

where i is individual, s is state, and t is time. The state fixed-effects then capture difference in state skill-prices. I run the regression separately for college and non-college educated workers.

I set California be the omitted dummy, so all skill prices are measured relative to California. On average, relative differences across states are larger for college-educated workers

than for non-college-educated workers. For non-college educated workers (conditional on age and sex), skill prices in Alabama are 62% lower than in California, whereas they are 7% higher in Washington. For college-educated workers, skill prices in West Virginia are 81% lower than in California, whereas they are only 12% lower in New York.

5.3.1 College Quality Spillovers

I extend the benchmark model to allow for a broad set of positive college externalities by calibrating $\lambda > 0$. I do so because the federal government may have incentives to spend in a given location for reasons beyond those captured in the benchmark economy.

To estimate the college spillover parameter λ, I use publicly available data from a series of papers: Andrews (2023), Russell et al. (2024), and Andrews and Russell (2025). This dataset compiles natural experiments in which college location assignment was nearly random. These papers study the effects of college openings on local attainment rates, invention, economic mobility, and inequality. I use the same methods to estimate the effect of college openings on long-run firm productivity.

Given this empirical estimate, I then run an experiment in the model in which a college is opened in a new location. I choose fixed productivity parameters $\bar{A}(e_{nc}, \ell)$, $\bar{A}(e_c, \ell)$, and λ to replicate the observed productivity effect in the data.

5.4 Housing

5.4.1 Expenditure Share

The housing expenditure share, ξ, is estimated from the BLS CEX Top Means Tables. It is calculated as the ratio of total housing expenditures to average annual expenditures. An average value is calculated for the years 2015-2019. This gives $^{19{,}590}/_{59{,}526} = 33\%$.

5.4.2 Rent Prices

To calculate average rental prices for each state I follow a method similar to Eeckhout et al. (2014) and run a hedonic regression for census tract i and state s,

$$rent_{i,s} = bedrooms_{i,s} + units_{i,s} + age_{i,s} + \theta_s + \epsilon_{i,s} \tag{45}$$

where $rent_{i,s}$ is median rent, $bedrooms_{i,s}$ is the median number of bedrooms, $units_{i,s}$ is the median number of units in the structure, $age_{i,s}$ is the median age, and θ_s are state fixed effects. Using the fixed effects from this regression, average rent in a state is then assumed to be given by,

$$\overline{rent}_s = bedroom + units + age + \hat{\theta}_s \tag{46}$$

where *bedrooms*, *units*, and *age* are now national median levels. Data is taken from the 5-year ACS sample for the years 2015-2019, and is accessible through IPUMS National Historical GIS. Average rent is highest in California and lowest in South Dakota, with a ratio of 2.33 between the two states.

5.4.3 Housing Stock

Housing stocks $D(\ell_w)$ can be derived from the market clearing condition given by equation (27) and summarized as,

$$p(\ell_w)D(\ell_w) = \xi \cdot \Big(\text{Disposable Earnings}(\ell_w)\Big) \tag{47}$$

Observable data for housing rental rates and total disposable earnings implied by choice probabilities is then used to infer housing stocks by states.

5.5 Government

State and federal college expenditures are taken from IPEDS for the sample of colleges defined above in *Section 5.2*. Total expenditures are defined as the sum of university revenues from state and federal appropriations, grants, and contracts. Recall from *Section 5.2* that I allow federal Title IV aid to affect net tuition, but not state revenues. Data on disposable income are taken from the National Income and Production Accounts (NIPA) tables. Given these two data points, equations (24) and (25) then imply state and federal taxes.

5.6 Model Fit

Prior to conducting the counterfactual exercises and policy experiments, I verify that the model replicates key features of the U.S. economy. The model is able to match targeted moments well and predict key elasticities out of sample.

5.6.1 Targeted Moments

The internally calibrated parameters not estimated via model inversion are estimated using the generalized method of moments (GMM). These parameters include the preference parameters $[\gamma_1^e, \gamma_2^e, \gamma_1^w, \gamma_2^w, \gamma_3^w, \sigma_e, \sigma_{cq}, \sigma_{cw}, \sigma_{ncw}]$, initial distribution parameters $[\mu_a, \sigma_a, \mu_p, \sigma_p]$, college fixed productivity for each state $\bar{q}_\ell$, and the elasticity of college quality with respect to spending per student, θ. See *Appendix B* for formal details and the computational procedure.

Figure D.8 plots the model fit for each of the four discrete choices: the college vs. no-college education decision, the college location choice, and the working location choices for both college and non-college graduates. The over-identified model performs well in generating accurate choice probabilities, with an R^2 above 0.96 for each discrete choice and a regression line coefficient within 1.0% of unity.

Figure D.9 compares wage growth rates by college quality in the model and the data. The

Table 9: **Model Fit – Targeted Moments**

Moment	Model	Data	Source
Average earnings growth	1.79	1.79	ACS
Average earnings s.d.	1.01	1.01	ACS
College attendance, parent income Q1	0.31	0.31	CPS
College attendance, parent income Q2	0.36	0.36	CPS
College attendance, parent income Q3	0.41	0.41	CPS
College attendance, parent income Q4	0.44	0.48	CPS
College attendance, parent income Q5	0.47	0.53	CPS
Earnings to college spending elasticity	0.18	0.18	See *Section 5.2.5*

Notes: This table compares the model generated moments to the data counterparts for selected targeted moments.

model nearly exactly replicates these growth rates. *Figure D.9* also compares the qualities implied by $\bar{q}$ in the first stage of calibration to those generated in the full general equilibrium, again showing a near-exact fit.

Table 9 reports the remaining moments matched during the first stage of the calibration procedure. The model precisely matches average earnings growth rates and variance over the lifecycle. College attendance is accurately fit for parental income quintiles one through three, though it slightly undershoots for the top two income quintiles. Finally, the model perfectly matches the elasticity of earnings with respect to college expenditures per student.

Figure D.10 shows that skill prices and housing rental prices obtained through model inversion exhibit a near-exact fit. State taxes are less well matched; however, this is due to the fact that no parameter directly affects taxes. I am simply comparing model-generated data to the empirical counterpart, and hence the fit is quite reasonable.

5.6.2 External Validation

I now verify several additional elasticities that are central to the government's optimal policy problems but were not targeted when calibrating the model. First, I compare the elasticity

of in-state college attendance to spending per student, in-/out-of-state tuition, and college-educated wages, both in the model and the data. Second, I compare the elasticity of alumni remaining in-state to work, with respect to college-educated wages.

To compute the in-state college attendance elasticities in the data, I estimate the following panel regression for state s in year t, via OLS,

$$\ln(L_{s,t}) = \beta_1 \ln(T^{in}_{s,t}) + \beta_2 \ln(T^{out}_{s,t-5}) + \beta_3 \ln(M_{s,t-5}) + \beta_4 \ln(E_{s,t}) + \gamma X_{s,t} + \lambda_s + \epsilon_{s,t} \quad (48)$$

$L_{s,t}$ is the percentage of a given state's college students studying outside their home state. $T^{in}_{s,t}$ and $T^{out}_{s,t-5}$ are in- and out-of-state tuition, respectively. $M_{s,t-5}$ is total instructional spending per student, and $E_{s,t}$ is the wage of college graduates. $X_{s,t}$ is a vector of controls: admission rate, average SAT score, percentage of the student body from out of state, and the unemployment rate. Finally, λ_s is a state fixed effect. Out-of-state tuition and spending per student are lagged by five years to account for delayed student expectations over changes in college quality. The coefficients of interest, which will be compared to the model counterparts, are β_1–β_4. *Appendix B* reports the full regression results and additional specifications. *Appendix B* also includes a robustness check where the elasticities for spending per student and wages are estimated using an IV regression strategy. The coefficients are of similar magnitude across all robustness exercises.

To compute the elasticity of alumni working in-state with respect to college-educated wages in the data, I estimate the following cross-sectional regression for state s by OLS,

$$\ln(L_s) = \beta_5 \ln(E_s) + \gamma X_s + \epsilon_s \quad (49)$$

L_s is the percentage of a given state's alumni who move outside of the state for work after graduation. E_s is the wage of college graduates. X_s is a vector of controls, including the unemployment rate and the percentage of the student body from out of state. The coefficient of interest, which will be compared to the model counterpart, is β_5. Due to data limitations,

I am unable to estimate this regression using panel data, as was done for the first set of elasticities. This cross-sectional regression covers the time period 2011–2017. *Appendix B* reports the full regression results.

To estimate these five elasticities in the model, I exogenously shock each state by changing in-/out-of-state tuition, high-skilled productivity, and spending per student. I shock in-/out-of-state tuition and spending per student by one percentage point, and I shock high-skilled productivity by an amount that induces a one percentage point increase in college earnings. I then average the responses across states and compare them to the data. This requires an additional 4×50 solves of the model. For this reason, these moments are used for external validation, as it would be computationally infeasible to solve the model an additional 200 times at each step of the calibration routine.

Table 10 compares the data estimates β_1–β_5 to the model-generated moments. The model performs well with out-of-sample predictions for the effects of wages on in-state college attendance and in-state alumni employment. In both cases, higher college wages cause fewer high school graduates and alumni to leave the state. While in-state tuition raises college expenditures (and hence quality), the increased financial burden dominates, and in both the model and the data, the percentage of high school graduates leaving the state rises. In contrast, out-of-state tuition and instructional spending more directly raise college quality, and in both cases, the percentage of students leaving the state falls. The model does quite well at predicting the elasticity for instructional spending, although it slightly overestimates the elasticity for out-of-state tuition compared to the data. Overall, these elasticities are central to the state government's optimal policy decisions, and the model's performance lends additional confidence to the counterfactual exercises.

Table 10: **Model Fit – External Validation**

Moment	Model	Data	Description
β_1	0.059	0.024	in-state studying to in-state tuition
β_2	-0.13	-0.031	in-state studying to out-of-state tuition
β_3	-0.21	-0.18	in-state studying to spending per student
β_4	-0.17	-0.22	in-state studying to college wages
β_5	-1.34	-1.23	in-state working to college wages

Notes: This table compares the model generated moments to the data counterparts for out-of-sample moments. Data for β_1-β_4 is taken from IPEDS and the CPS for the years 1990-2017. Data for β_5 is taken from IPEDS, the CPS, and B&B for the years 2011-2017.

6 Results

This section reports the results of all counterfactual exercises. All results are reported for the benchmark model without spillovers from college qualities to firm productivities.

6.1 Mechanisms of the Government Choices

Before conducting the full welfare and optimal policy analysis in *Section 6.2*, I first present intuition for the main trade-offs and mechanisms faced by a state or the federal government when solving problems (29) and (30). For both the state and federal problems, I use a five-state example economy to simplify the exposition. The welfare analysis in *Section 6.2* then employs the full 50-state model.

6.1.1 State Government

The state government can allocate revenues toward: (1) reducing in-state tuition, (2) increasing in-state capacity, (3) reducing out-of-state tuition, (4) increasing out-of-state capacity, and (5) spending to directly increase college quality. In this section, I provide the economic intuition behind the optimal choice over these five objects.

For simplicity and clarity of exposition, consider a stylized five-state model with two

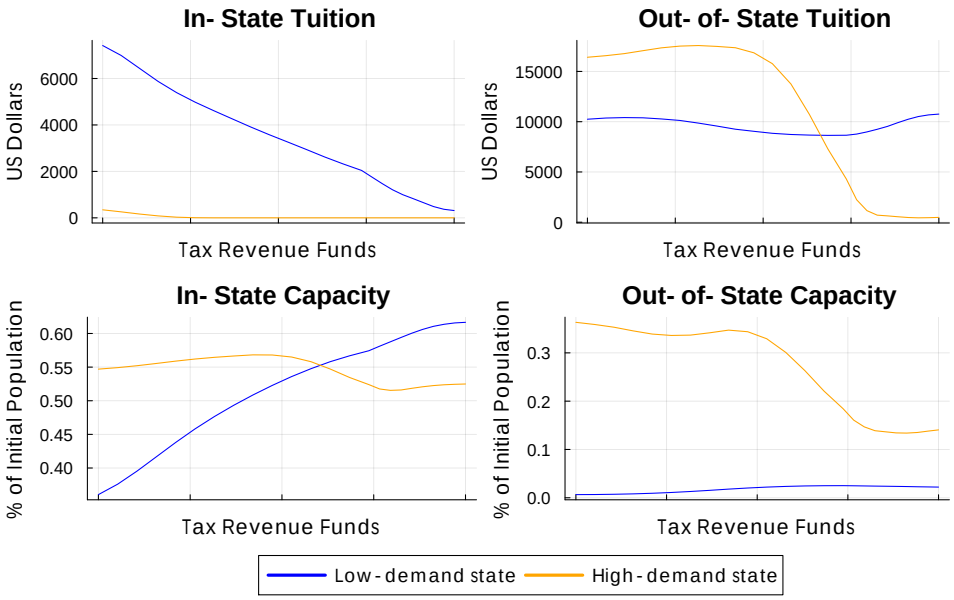
states of interest: one with high, and the other with low, out-of-state demand. That is, one with a high admissions cutoff and the other with a near open admissions cutoff. Call these high- and low-demand states, respectively. I choose these state types as they best highlight differences across states. *Figures 10* and *11* plot the optimal decisions for the state government as funding through state labor tax appropriations is increased—that is, as an extra dollar becomes available, how the state chooses to optimally allocate it.

For low levels of state appropriations, the low-demand state sets a relatively high in-state tuition and a small out-of-state capacity. Spending per student is also quite low, and quality is relatively low as well. Conversely, the high-demand state is able to set out-of-state tuition and capacity much higher. This allows for an almost full subsidization of in-state tuition, and spending per student that is significantly higher than in the low-demand state. Mean ability is lower in the high-demand state than in the low-demand state, as the higher capacity at the high-demand state results in a lower average ability of the student body. However, overall, the higher spending per student dominates, and quality is higher in the high-demand state.

As state appropriations increase, the low-demand state steadily decreases in-state tuition, and increases in-state capacity and spending per student. Out-of-state tuition remains relatively flat, and out-of-state capacity rises slightly. As spending per student and quality increase, so does demand, and hence, admissions criteria and average ability also increase.

For the high-demand state, as appropriations rise, tuition is first reduced to zero, and in-state capacity flattens out at "full capacity". That is, the level at which all students who wish to attend are able to do so. Out-of-state tuition and out-of-state capacity remain relatively flat. Where behavior differs from the low-demand state is when state appropriations become sufficiently high, it becomes more beneficial to restrict out-of-state capacity and lower out-of-state tuition. This causes a large increase in admissions criteria and hence peer effects, which dominate the decrease in spending per student, so quality continues to rise. Out-of-state tuition and capacity then remain at this new level, and all further increases in state

Figure 10: **State Government Solution – Tuition and Capacity**



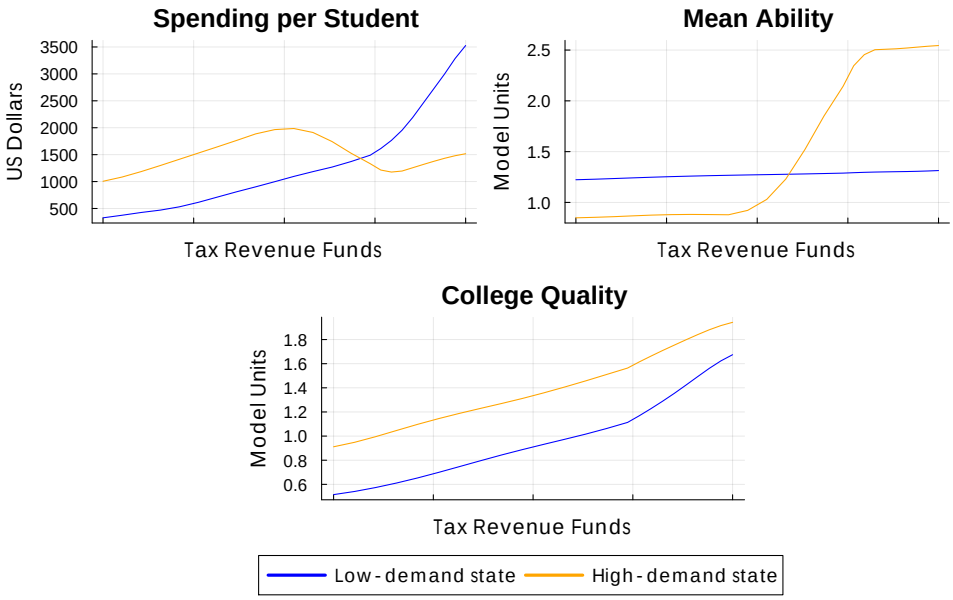
Notes: This figure plots the solution for the optimal policy problem of two example states when college revenues financed through labor taxes are increased.

appropriations are directed toward increasing spending per student.

For this calibrated model, both types of state governments find it optimal to simultaneously increase spending per student and capacity, and decrease tuition as revenues increase. However, this is not true in all regions of the parameter space. For example, under a different parameterization, it may be optimal for a policymaker to keep spending per student constant until tuition has been brought to zero. I now discuss in more detail the key trade-offs and parameters involved in allocating an additional dollar of college funding. This determines the relative rate at which these objects change. I largely ignore a discussion of general equilibrium changes in wages and housing rental rates, as these significantly complicate the analysis.

Consider directing a marginal amount of additional revenue toward spending per student, m_ℓ. The first-order effect is to increase quality. Increased college quality has several effects. First, it increases earnings for those who attend the given college. Increased earnings directly affect the welfare of those attending the college. It also increases output (and reduces taxes)

Figure 11: **State Government Solution – Spending, Peer effects, and Quality**



Notes: This figure plots the solution for the optimal policy problem of two example states when college revenues financed through labor taxes are increased.

in the given state, improving welfare for those not attending the college as well. The key model parameter here is θ, the elasticity of earnings with respect to spending per student. A higher θ increases the relative benefit of directing funds toward spending per student.

A second effect of increased spending per student is increased demand for the college. This raises admissions cutoffs and, hence, average ability—further increasing quality. Importantly, increased demand also occurs for out-of-state students. Given complementarities between ability and college quality in the human capital production function, this attracts higher-ability students from out of state to attend the given college. Many of these high-ability students then remain in-state to work after graduating. *Figure D.20* shows that the average ability of college-educated laborers in a given state increases monotonically as spending per student is exogenously increased. Importantly, however, *Figure D.20* reflects an exogenous increase in spending per student, without respecting the state budget constraint. When spending per student increases, it must come at the cost of either higher in-/out-of-state tuition or lower in-/out-of-state capacity.

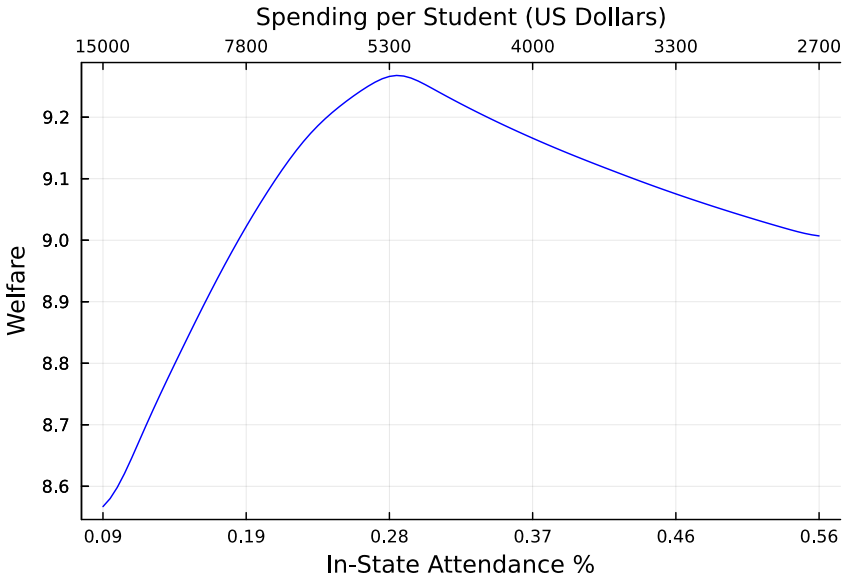
Suppose that the higher spending per student comes at the cost of a *ceteris paribus* decrease in in-state capacity. The first-order effect is simple, "one" fewer student is educated, but all students accumulate more human capital over their lifecycles. *Figure D.21* shows that decreasing capacity also raises the average ability of the student body, due to increased admissions cutoffs.

The key parameters here are those that control the convexity of the college cost function (κ_2) relative to the concavity of the human capital production function (ν). For high (low) capacity and low (high) spending per student, both the cost function and the human capital production function are steep (flat). Hence, (in partial equilibrium) it is more beneficial to direct existing resources toward spending per student. *Figure 12* shows welfare as a function of in-state capacity (bottom x-axis) and the spending per student implied by the budget constraint (top x-axis). Welfare increases monotonically until it reaches a peak, at which point the state policymaker is indifferent between increasing spending or increasing capacity. The welfare function is plotted for the model in general equilibrium, but the same pattern is also observed in partial equilibrium.

Suppose instead that the higher spending per student comes at the cost of a *ceteris paribus* increase in in-state tuition. This causes demand to fall for all types of students, and some students to be "priced out" of college. In particular, there will be high-ability but low-income students who are no longer able to attend college. *Figure D.22* plots average ability at a college as in-state tuition prices are exogenously increased. For sufficiently low demand (i.e., open admissions) and high tuition prices, the mass of those attending college also begins to fall.

Important parameters here are the mean and variance of the distribution of parental resources, μ_p and σ_p. These parameters determine the rate at which students are no longer able to afford college as tuition increases. *Figure 13* plots welfare as a function of in-state tuition (bottom x-axis) and the spending per student implied by the budget constraint (top x-axis). Again, the function increases monotonically until an indifference point is reached at

Figure 12: **State Welfare – In-State Capacity and Spending per Student**

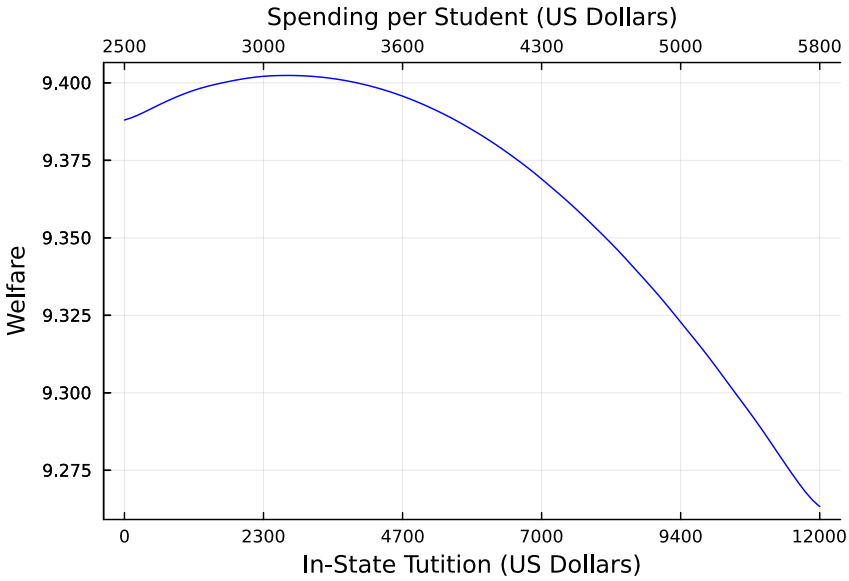


Notes: This figure plots state welfare as in-state capacity is increased. From the budget constraint this immediately implies a level of spending per student. All else is held fixed. The bottom x-axis shows increases in capacity, and the top x-axis shows corresponding decreases in spending per student.

the peak.

Finally, the policy maker could instead, reduce out-of-state capacity or increase out-of-state tuition. The welfare increases of an out-of-state student is twofold. First, higher average ability of the student body and higher tuition increase college quality. Second, a portion of the out-of-state student remain in-state to work. As show in *Section 4*, for a high-productivity state, this second benefit is small. For this type of state, they choose tuition and average ability of out-of-state students such that the marginal effect on college quality is equal. Moreover, as just discussed, it must be that this marginal effect on quality must be equal to the marginal effect of increasing in-state capacity or reducing in-state tuition. For a state with low-productivity there is the additional benefit of a relatively large increase of in-state college workers. This type of state will then set the marginal benefit of an out-of-state student on quality below the marginal benefit of increasing in-state capacity of reducing in-state tuition.

Figure 13: **State Welfare – In-State Tuition and Spending per Student**



Notes: This figure plots state welfare as in-state tuition is increased. From the budget constraint this immediately implies a level of spending per student. All else is held fixed. The bottom x-axis shows increases in tuition, and the top x-axis shows corresponding increases in spending per student.

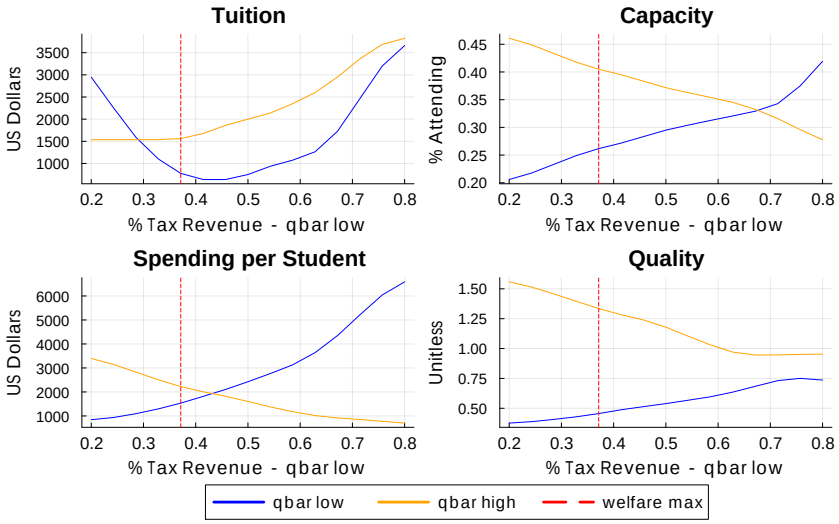
6.1.2 Federal Government

I now discuss the main trade-offs present for a federal policymaker simultaneously choose tuition, capacity, and spending per student in each state. When the federal government chooses polices the forces determining optimality "inside" a given state are the same as described above for the state problem. The key addition here is how the federal government decides between allocating another dollar to North Dakota or another dollar to California.

Consider an example with two states that are ex-ante identical except for, one with a high, and the other with a low, historic level of college quality (i.e., $\bar{q}$). Also, for simplicity of exposition assume that there is no separate out-of-state tuition or capacity.

Figure 14 plots this example economy, where both states have equal firm productivity. Each panel plots optimal policy as the federal government alters the fraction of total revenue allocated to the state (π_ℓ). The federal government finds it optimal to provide the majority of revenues to the high-$\bar{q}$ state. However, spending per student between the two states is similar, and tuition is set to zero in the low-$\bar{q}$ but not in the high-$\bar{q}$ state. The government

Figure 14: **Federal Government Solution, Same Firm Productivity**



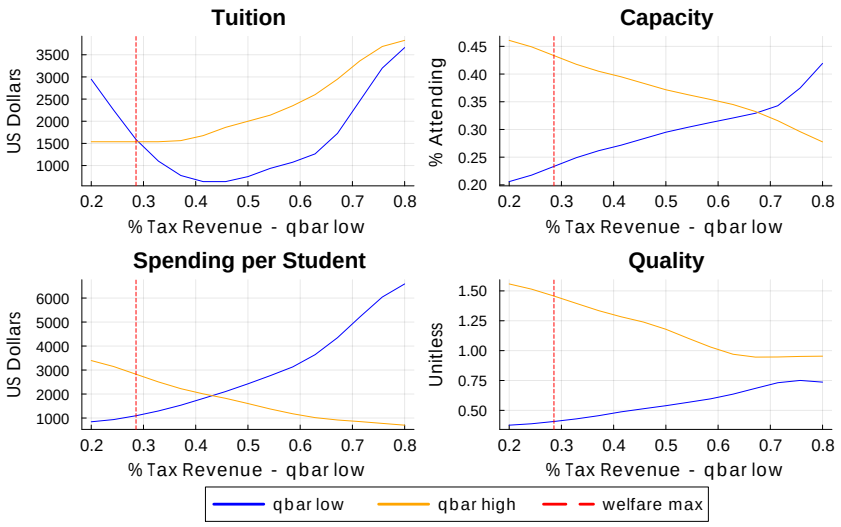
Notes: This figure plots the solution to the federal governments problem for two different states with the same productivity.

spends the additional revenue at the high-quality college, significantly expanding capacity. *Figure 15* shows that this same pattern is even starker when the high-productivity firms are also located near the high-quality colleges.

However, *Figure 16* shows that when the high-quality colleges are located near the (relatively) low-productivity firms (e.g., Wisconsin), the federal policymaker now finds it optimal to spend roughly half of revenues on each location and set capacity close to equal. That is, any mismatch between historical college locations and present firm activity significantly alters decisions by a federal policy maker. In this situation, the federal government would "like to" move the location of the high-quality colleges, however they are constrained in doing so.

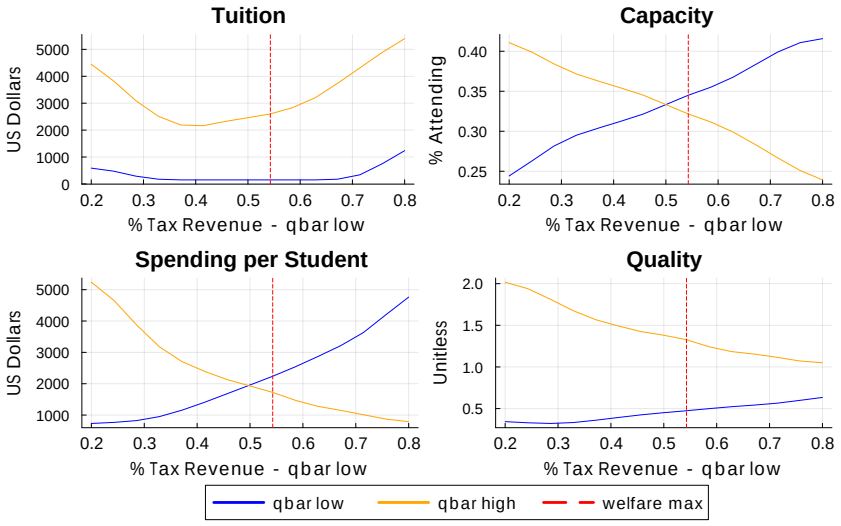
Next, consider the same economy with the only source of ex-ante heterogeneity being historical college quality or firm productivity. *Figures* **??** and **??** plot the welfare maximizing solution to the federal government problem as the available revenues generated through aggregate tax appropriations are increased from a low to high level. ***TBD***

Figure 15: **Federal Government Solution, Matched Firm-College Productivity**



Notes: This figure plots the solution to the federal governments problem for two different states with different productivity. Firm productivity is "matched" with colleges in the sense that high productivity firms are located in the same location as high fixed $\bar{q}$ colleges.

Figure 16: **Federal Government Solution, Unmatched Firm-College Productivity**



Notes: This figure plots the solution to the federal governments problem for two different states with different productivity. Firm productivity is "unmatched" with colleges in the sense that high productivity firms are not located in the same location as high fixed $\bar{q}$ colleges.

Finally, the federal government must also consider differences in housing rental rates across states, which here serve as a proxy for land prices. This force is simple, it is more expensive to build additional college capacity in California than in North Dakota. Hence, the federal government will allocate all additional funding devoted to increased capacity, relative to the marginal cost.

6.2 Welfare Analysis

In this section, I report welfare changes associated with the optimal state and federal policies. Welfare is measured using utilitarian weights on each agent and the consumption-equivalence definition.

6.2.1 Optimal State Policy

Next, I solve the main counterfactual exercise. Each state takes as given the optimal policy of another state; however, it must be that in equilibrium no state has a profitable deviation from its optimal strategy.[16] While I find that most states are roughly optimizing in the current benchmark U.S. economy, there are still significant welfare gains of 3.1%. On average, it is optimal for states to increase capacity and lower tuition, at the expense of lowering spending per student. This causes tax rates generating a fixed revenue to also fall on average, as more agents are working with a college degree and earning higher lifetime wages.

This is not the case for every state. I find that more rural states with low-quality and relatively high capacity (e.g., North Dakota) see welfare improvements by shrinking capacity and raising quality. As a result, fewer high-ability high school graduates leave the state to seek a college education elsewhere (e.g., California). This also results in fewer overall college graduates, but on net it is a positive increase in welfare for the state.

[16] An interesting extension to the current counterfactual would be to consider the possibility that states internalize the effect that their policy choices have on other states' policy choices.

I also find that when maximizing the welfare of those born in the state, it is indeed optimal for most states to set large tuition premiums for out-of-state students. On average, there is less movement across states under the optimal regime. This follows from states with low-quality education increasing quality and retaining more domestic high school graduates. However, quality does not rise enough in such locations to attract more out-of-state students on net.

6.2.2 Optimal Federal Policy

I now consider optimal federal education policy, where a federal policymaker simultaneously chooses policy for each state in order to maximize aggregate welfare. This solution does not coincide with each individual state's solution, and each state would find profitable deviations from this solution when maximizing the welfare of its domestic agents.

For the general case, there are large aggregate welfare gains of 7.2% when the federal government chooses all policy parameters optimally. Taxes needed to generate the fixed expenditures on higher education fall significantly. I also consider an experiment where total expenditures on higher education are reduced, yet aggregate welfare still rises when choosing a more efficient national education system.

The most notable difference from the state solution is out-of-state tuition policy. The federal government finds it optimal to set out-of-state tuition below in-state tuition for many high-quality colleges. This incentivizes high-ability agents to attend the high-quality colleges. These agents then experience a higher degree of complementarity with college quality than low-ability agents, and hence larger increases in lifetime human capital. Moreover, on average, firms with high productivity for skilled workers are located near high-quality colleges. These agents are now also more likely to work for a high-productivity firm, again with a higher degree of complementarity than low-ability agents.

Table 11: **Welfare Changes of Policy Experiments Relative to Benchmark Economy**

No OS Tuition, Diff. τ_ℓ^s (1)	No OS Tuition, Same τ_ℓ^s (2)	Free Tuition (3)	Moving Subs. (4)
+1.17	+0.73	+4.33	+2.14

Notes: This table plots the changes in welfare for selected policy experiments. Numbers are percentage point changes relative to benchmark U.S. economy.

6.3 Spatial Policies and Higher Education

I now consider several additional counterfactuals that have been proposed as policy reforms for the American college system. *Table 11* provides a summary of these policies. All policies are considered for the benchmark economy's level of spending on higher education across states and at the federal level.

6.3.1 Eliminating Out-of-State Tuition

For the first policy experiment, I consider the elimination of cross-state pricing frictions. First, out-of-state tuition is eliminated, and in-state tuition is held fixed at the calibrated level. I then consider a second version of the policy in which in-state tuition adjusts upward so that tax levels return to their benchmark values. This policy yields considerably smaller welfare increases than the optimal federal government solution.

Table 11 shows that welfare gains are modest for both experiments, with welfare rising by 1.17 and 0.73 percentage points, respectively. Under both policies, demand for low-quality colleges—which originally charge relatively low out-of-state tuition—falls. As a result, average quality falls at low-quality colleges and rises at high-quality colleges.

6.3.2 Free Tuition

Consider now a free national tuition policy. It is important to note that free national tuition was not the solution to the federal problem discussed above. The optimal solution to

this problem included free tuition at most low-quality colleges, but positive (albeit modest) tuition at high-quality colleges. Nevertheless, a free national tuition policy does increase aggregate welfare.

A free national tuition policy leads to a large increase in aggregate welfare of 4.33 percentage points. The quality of colleges falls everywhere due to the significantly reduced spending per student now available. However, as seen when eliminating out-of-state tuition, admissions cutoffs rise at high-quality colleges and fall at low-quality colleges. Taxes fall slightly, as higher-ability agents (on average) now attend high-quality colleges. In the long run, the policymaker might also consider expanding capacity to meet increased demand under free tuition. This would also likely cause taxes to fall for a fixed level of spending on higher education.

6.3.3 College Moving Subsidies

In addition to out-of-state tuition, moving costs are a key friction preventing the optimal spatial distribution of agents. In this experiment, I eliminate moving frictions by providing moving subsidies equivalent to the utility cost of migration. I assume the policy is funded by the federal government and requires increased revenues collected through labor taxes.

This policy also results in a large increase in aggregate welfare of 2.14 percentage points. Quality and demand rise (fall) at high- (low-) quality colleges. The changes in migration dynamics are qualitatively similar to eliminating out-of-state tuition. However, the aggregate welfare gains are much larger, even though taxes must adjust upward to fund the policy. In fact, the policy is equivalent to a subsidy on out-of-state tuition, as prescribed by the optimal solution to the federal government's problem.

7 Conclusion

This paper investigates the sources of large cross-state heterogeneity in college systems. I document several empirical facts showing that state size and productivity are positively correlated with in-/out-of-state tuition, spending per student, and the percentage of the student body that is in-state. I then develop a novel quantitative spatial model incorporating an endogenous distribution of college qualities, lifecycle human capital accumulation, and differentiated firms and wages across locations. The model is able to rationalize the observed differences in higher education across the United States. I show that the positive correlations between state size and wages are driven by scale economies and migration patterns.

At the national level, I empirically document an important source of friction for a counterfactual federal policymaker: the distribution of college quality is highly persistent over time and spatially concentrated. Unlike firms, where the average age is five years, 73% of students attend a public college whose location was fixed more than 100 years ago. I show that when high-productivity firms are "matched" with high-quality colleges, the federal government allocates a large majority of revenues to high-quality, high-productivity states. However, when high-productivity firms are mismatched with high-quality colleges, the policymaker distributes funds more evenly.

Finally, I ask: is the current system of higher education welfare-maximizing at the state or federal level? I find that, for a fixed level of tax revenues spent on higher education, optimally setting policy parameters allows both state and federal governments to significantly improve aggregate welfare by 3.1% and 7.2%, respectively. This large welfare gain holds even in the absence of spillovers from university quality to firm productivity. Moreover, I find room for welfare improvements even when total expenditure on higher education is reduced.

Existing quantitative spatial models examining firm-side policies and redistribution typically assume an exogenous labor supply in each location, often of one type. Future work using this framework plans to analyze optimal firm policy in the presence of an endogenous

spatial distribution of university quality and skilled versus non-skilled human capital supply. Firm-side policies studied in the existing literature may prove ineffective in the absence of a nearby college capable of supplying the skilled labor necessary for production.

References

Abbott, Brant, Giovanni Gallipoli, Costas Meghir, and Giovanni L Violante, "Education Policy and Intergenerational Transfers in Equilibrium", *Journal of Political Economy* 127 (2019), 2569–2624.

Andrews, Michael J, "How do Institutions of Higher Education Affect Local Invention? Evidence from the Establishment of US Colleges", *American Economic Journal: Economic Policy* 15 (2023), 1–41.

Andrews, Michael J and Lauren Russell, "Historical Place-Based Investments and Contemporary Economic Mobility and Inequality: Impacts of University Establishment", *Working Paper* (2025).

Autor, David H and David Dorn, "The Growth of Low-skill Service Jobs and the Polarization of the U.S. Labor Market", *American Economic Review* 103 (2013), 1553–1597.

Bartik, Timothy J, "Who benefits from state and local economic development policies?" (1991).

Berry, Steven T, "Estimating Discrete-choice Models of Product Differentiation", *The RAND Journal of Economics* (1994), 242–262.

Brotherhood, Luiz, Bernard Herskovic, and Joao Ramos, "Income-based Affirmative Action in College Admissions", *The Economic Journal* 133 (2023), 1810–1845.

Cellini, Stephanie Riegg and Nicholas Turner, "Gainfully Employed?: Assessing the Employment and Earnings of For-profit College Students Using Administrative Data", *Journal of Human Resources* 54 (2019), 342–370.

Chetty, Raj, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan, "Income Segregation and Intergenerational Mobility Across Colleges in the United States", *The Quarterly Journal of Economics* 135 (2020), 1567–1633.

Conzelmann, Johnathan G, Steven W Hemelt, Brad J Hershbein, Shawn Martin, Andrew Simon, and Kevin M Stange, "Grads On the Go: Measuring College-specific Labor Markets for Graduates", *Journal of Policy Analysis and Management* (2023).

Deming, David J and Christopher R Walters, "The Impact of State Budget Cuts on US Postsecondary Attainment", *Working Paper 23736* (2017).

Eeckhout, Jan, Roberto Pinheiro, and Kurt Schmidheiny, "Spatial Sorting", *Journal of Political Economy* 122 (2014), 554–620.

Epple, Dennis, Richard Romano, Sinan Sarpça, and Holger Sieg, "A General Equilibrium Analysis of State and Private Colleges and Access to Higher Education in the US", *Journal of Public Economics* 155 (2017), 164–178.

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

Fajgelbaum, Pablo D and Cecile Gaubert, "Optimal Spatial Policies, Geography, and Sorting", *The Quarterly Journal of Economics* 135 (2020), 959–1036.

Fu, Chao, Junjie Guo, Adam J Smith, and Alan Sorensen, "Students' Heterogeneous Preferences and the Uneven Spatial Distribution of Colleges", *Journal of Monetary Economics* 129 (2022), 49–64.

Guvenen, Fatih and Serdar Ozkan, *TikTak: A Multistart Global Optimization Algorithm*, 2021.

Heckman, James J, Lance Lochner, and Christopher Taber, "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents", *Review of Economic Dynamics* 1 (1998), 1–58.

Henderson, J Vernon, "Understanding Knowledge Spillovers", *Regional Science and Urban Economics* 37 (2007), 497–508.

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

Herrington, Christopher, "Public Education Financing, Earnings Inequality, and Intergenerational Mobility", *Review of Economic Dynamics* 18 (2015), 822–842.

Hillman, Nicholas W, "Geospatial Analysis in Higher Education Research", *Higher Education: Handbook of Theory and Research* (2017), 529–575.

Holter, Hans A, "Accounting for Cross-country Differences in Intergenerational Earnings Persistence: The Impact of Taxation and Public Education Expenditure", *Quantitative Economics* 6 (2015), 385–428.

Huggett, Mark, Gustavo Ventura, and Amir Yaron, "Sources of Lifetime Inequality", *American Economic Review* 101 (2011), 2923–2954.

Ishimaru, Shoya, "Geographic Mobility of Youth and Spatial Gaps in Local College and Labor Market Opportunities" (2023).

Jaffe, Adam B, "Characterizing the "Technological Position" of Firms, with Application to Quantifying Technological Opportunity and Research Spillovers", *Research policy* 18 (1989), 87–97.

Kane, Thomas J, "Public Intervention in Post-secondary Education", *Handbook of the Economics of Education* 2 (2006), 1369–1401.

Kantor, Shawn and Alexander Whalley, "Knowledge Spillovers from Research Universities: Evidence from Endowment Value Shocks", *Review of Economics and Statistics* 96 (2014), 171–188.

Katz, Lawrence F and Kevin M Murphy, "Changes in relative wages, 1963–1987: supply and demand factors", *The Quarterly Journal of Economics* 107 (1992), 35–78.

Kennan, John, "Spatial Variation in Higher Education Financing and the Supply of College Graduates", *National Bureau of Economic Research Working Paper* (2015).

Kirkeboen, Lars J, Edwin Leuven, and Magne Mogstad, "Field of Study, Earnings, and Self-selection", *The Quarterly Journal of Economics* 131 (2016), 1057–1111.

Kline, Patrick and Enrico Moretti, "People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs", *Annual Review of Economics* 6 (2014), 629–662.

Krueger, Dirk and Alexander Ludwig, "On the Optimal Provision of Social Insurance: Progressive Taxation Versus Education Subsidies in General Equilibrium", *Journal of Monetary Economics* 77 (2016), 72–98.

McFadden, Daniel, "Conditional Logit Analysis of Qualitative Choice Behavior" (1972).

Parmley, Kelli, Allison Bell, Hans L'Orange, and Paul Lingenfelter, "State Budgeting for Higher Education in the United States", *Report, State Higher Education Executive Officers* (2009).

Redding, Stephen J and Esteban Rossi-Hansberg, "Quantitative Spatial Economics", *Annual Review of Economics* 9 (2017), 21–58.

Rosenthal, Stuart S and William C Strange, "Evidence On the Nature and Sources of Agglomeration Economies", *Handbook of Regional and Urban Economics*, vol. 4, Elsevier, 2004, pp. 2119–2171.

Russell, Lauren C, Lei Yu, and Michael J Andrews, "Higher Education and Local Educational Attainment: Evidence from the Establishment of US Colleges", *Review of Economics and Statistics* 106 (2024), 1146–1156.

Turley, Ruth N López, "College Proximity: Mapping Access to Opportunity", *Sociology of Education* 82 (2009), 126–146.

Wright, Jacob, "Quality, Inequality, and Mobility: American vs. Canadian Systems of Higher Education", *Available at SSRN: https://ssrn.com/abstract=4953531* (2023).

Wright, Jacob and Alex Wurdinger, "Declining Teen Employment: Causes and Consequences", *Available at SSRN: https://ssrn.com/abstract=4953600* (2025).

Wright, Jacob and Angela Zheng, "From Kindergarten to College: The Impact of Education Policies over the Lifecycle", *IZA Discussion Papers, No. 17301, Institute of Labor Economics* (2025).

Appendix

"On the Spatial Distribution of Colleges"

Jacob Wright

Contents

A Data **77**

A.1 Data Sources 77

A.2 Higher Education Across States 82

B Calibration **84**

B.1 Colleges 84

B.1.1 Reciprocity Agreements 84

B.1.2 Cost Function 85

B.1.3 Quality Function 85

B.2 External Validation 85

C Computation **86**

C.1 Competitive Equilibrium 86

C.2 Calibration 86

C.3 State Government 87

C.4 Federal Government 87

D Additional Figures and Tables **88**

A Data

A.1 Data Sources

In this Appendix, I report details of each dataset used in *Section 2* for motivating statistics and *Section 5* for model parameters and moments.

Integrated Postsecondary Education Data System – The Integrated Postsecondary Education Data System is built from interrelated surveys conducted annually by the U.S. Department of Education's National Center for Education Statistics. The survey is one of the key datasets used for much of the analysis throughout the paper. The survey covers a large array of variables on admissions, enrollment, graduation rates, and college finances. Most variables cover the 1990 to 2022 time period.

College Scorecard – The College Scorecard dataset is maintained by the Department of Education. The key variable used from the College Scorecard is earnings by former students. This data is reported for 6, 7, 8, 9, and 10 years after a given pooled cohort entered college. Median earnings are given for individuals who are working and not enrolled in a college at the point of measurement. Earnings are defined as wages and deferred compensation from the IRS W-2 form.

National Association of College and University Business Officers – The National Association of College and University Business Officers is a membership organization representing more than 1,900 colleges and universities across the country. Among other activities, it conducts an annual study of college endowments in conjunction with the Teachers Insurance and Annuity Association of America. The total endowment variable is the one used here.

Grapevine – Grapevine is an annual data collection effort that reports on state funding for higher education, jointly produced by the State Higher Education Executive Officers (SHEEO) Association and the Center for the Study of Education Policy at Illinois State University. Consistent with the approach in Deming and Walters (2017), I use Grapevine's appropriation figures instead of those from IPEDS due to concerns about double-counting state support across different campuses within the same institution and potential inaccuracies in administrator responses to the IPEDS survey.

Baccalaureate and Beyond – The Baccalaureate and Beyond (B&B) is a nationally representative longitudinal study that tracks individuals who earn a bachelor's degree in a specific academic year, following their outcomes one, four, and ten years after graduation. This allows me to obtain migration rates from the state of college attendance to the subsequent state of employment. I use the B&B as opposed to the NLSY, as the sample size for college graduates is roughly 10 times larger, and I am able to use the 2016–2018 time period as opposed to the late 1990s or early 2000s as in the NLSY. B&B microdata is restricted access; however, the National Center for Education Statistics (NCES) Datalab provides the creation of custom, publicly available tables.

College Founding Dates – To the best of my knowledge, there does not exist a publicly available dataset covering college founding or establishment dates. I build this dataset by scraping Wikipedia. The overwhelming majority of college Wikipedia pages contain a box along the right-hand side of the page with a field labeled "Established" or "Founded," as seen in *Figure D.17*. This makes scraping simple and accurate using a list of college names from IPEDS. A few remaining colleges were then manually added. In total, I gather 1,610 founding dates for private and public colleges. I have made the dataset containing these compiled founding dates publicly available at https://doi.org/10.3886/E233781V1. College founding dates vary in accuracy in the sense that, as colleges grow, merge, split, etc. over

time, there is not a standardized way to determine how the founding date evolves. However, for the simple correlation motivating statistics, the officially reported college dates serve as a reasonable proxy on average.

Deming and Walters (2017) – For data on tuition freezes and caps by state, I take data from Deming and Walters (2017). This paper compiles the data for the years 1990–2013 using a combination of official sources and Lexis-Nexis searches of state newspapers. All caps or freezes were coded only under the condition that they could be independently verified.

Chetty et al. (2020) – Data for wage profiles for the ages 23–35 is taken from Chetty et al. (2020). Publicly available statistics are published using microdata taken directly from IRS tax records.

Conzelmann et al. (2023) – Conzelmann et al. (2023) construct a dataset on college alumni by scraping publicly available LinkedIn (LI) profiles from institutional webpages. The resulting data offer aggregated information on the geographic distribution of graduates from nearly all public and private nonprofit higher education institutions in the U.S. Through a series of validation checks, the authors show that the LinkedIn-based data closely align with official government sources in both graduate counts and location patterns. I use data from Conzelmann et al. (2023) as opposed to the B&B or other sources when I require alumni migration patterns at the individual college level instead of the aggregate state level.

Panel Study of Income Dynamics – The Panel Study of Income Dynamics (PSID) launched in 1968 with a nationally representative cohort of over 18,000 individuals from 5,000 U.S. families. It has since tracked these individuals and their descendants over time, collecting detailed information on income, mobility, and other socioeconomic outcomes. I use the PSID to calculate skill prices, as it is the largest dataset that allows me to do so.

While the ACS contains information on migration, it does not report previous years' earnings.

National Longitudinal Survey of Youths – The NLSY97 is a nationally representative longitudinal dataset administered by the U.S. Bureau of Labor Statistics, tracking a cohort of American youth born between 1980 and 1984. Participants were between 12 and 17 years old when they were first surveyed in 1997. The survey provides detailed information on a wide range of topics, including education, labor market outcomes, fertility, public assistance, health, family characteristics, and individual beliefs. For our analysis, we use data from the NLSY97 on student achievement and family background to discipline parameters related to intergenerational processes.

American Community Survey – The American Community Survey (ACS) is an ongoing, nationally representative survey conducted by the U.S. Census Bureau. It provides annual data on demographic, economic, housing, and social characteristics of the U.S. population. I use ACS microdata to measure cross-state migration patterns, population sizes, educational attainment, and earnings. The ACS is available annually from 2005 onward.

Current Population Survey – The Current Population Survey (CPS) is a monthly and annual, nationally representative survey conducted by the U.S. Census Bureau and the Bureau of Labor Statistics. It provides detailed information on earnings and educational attainment. I use the CPS over the ACS when I require annual data prior to 2005, and the significantly larger sample size of the ACS is not required.

United States Census – The U.S. Census, conducted every ten years by the U.S. Census Bureau, provides detailed demographic, social, housing, and economic data on the entire U.S. population. The decennial census offers comprehensive population counts and characteristics at various geographic levels, including states, counties, and census tracts.

I use U.S. Census data for several purposes. First, to calculate rent prices at the state level. I use the National Historical Geographic Information System (NHGIS), which is a comprehensive source of historical U.S. Census data and geographic boundary files, covering the period from 1790 onward. It delivers harmonized summary statistics on demographics, housing, agriculture, and economic indicators across multiple geographic levels—ranging from states to census blocks. The database is particularly valuable for conducting longitudinal and spatial research using consistent census geography. Second, I use Census data to calculate population-weighted centroids for each state. I do so using the population and coordinates reported by Block Groups from the U.S. Census. Lastly, I use the U.S. Census for historically comparable county population estimates dating back to 1900.

National Income and Product Accounts – The National Income and Product Accounts (NIPA), maintained by the U.S. Bureau of Economic Analysis (BEA), provide comprehensive data on the economic activity of the United States. These accounts include measures such as GDP, personal income, and government expenditures, offering a consistent framework to track the flow of income and production across sectors. In this paper, I use NIPA for data on disposable income after taxes by state. The above datasets all report pre-tax earnings or income.

Consumer Expenditure Survey – The Consumer Expenditure Survey (CEX), produced by the U.S. Bureau of Labor Statistics, provides detailed information on household spending across a wide range of categories. I use the published "Top Line Mean" tables, which report average annual expenditures by major category, to compute the share of total housing-related expenses—such as rent, mortgage payments, utilities, and maintenance—relative to average total annual household expenditures.

A.2 Higher Education Across States

In this appendix, I report additional details beyond those discussed in *Section 2* on cross-state differences in higher education. All figures are shown in *Appendix D*.

Financing – In addition to spending per student, state financing of colleges differs in several other important ways. *Figure D.11* shows that higher education accounts for a much larger share of total government outlays in Texas than in California or Pennsylvania, with the former spending three times as much as the latter. Another major dimension of variation is the proportion of total college expenditures financed by state tax revenues. *Figure D.12* plots the average share of a college's budget funded by state appropriations. In California and New York, nearly half of the college budget comes from state appropriations—a share that is even higher when including state grants and contracts. At the other end of the spectrum, colleges in Pennsylvania and Michigan fund less than 10% of expenditures through state appropriations. This implies that alternative revenue sources, such as tuition or donations, are much more important for college budgets in some states than in others.

Migration – Now consider student migration across states. *Figure D.13* plots the percentage of a state's college student body that originates from out of state. For example, New York and Texas have relatively few out-of-state students, with less than 10% of their student populations coming from elsewhere. In contrast, in Vermont, the majority of college students are from outside the state. Alternatively, consider student outflows. *Figure D.14* plots the percentage of students who attend college outside their home state. Nearly all students who graduate from high school in California remain in-state for college, whereas the majority of students from New Jersey leave to attend college elsewhere.

Measures of Ability – I now contrast different measures of student ability at public colleges across states. First, *Figure D.15* plots median ACT composite scores by state.[17] Average ACT composite scores fall around the 86th percentile of nationwide scores in states like New York and Michigan, whereas states such as South Dakota and Idaho have average scores near the 66th percentile.

A second common proxy for the academic composition of college students is the 6-year (or 150%) graduation rate, shown in *Figure D.16*. Once again, New York and Michigan have high rates, exceeding 65%, while Idaho and South Dakota have graduation rates below 50%. These graduation rates are important for understanding the cost of producing a graduate, as opposed to spending per enrolled student. A state may spend relatively little per student but have a high dropout rate, which results in a high cost per graduate. An interesting extension to this paper would be to endogenize graduation rates in this context.

Tuition and Spending per Student – The dependent variables in the motivating regressions presented in *Section 2* are also significantly correlated with each other. *Table D.1* reports the relationship between spending per student and in-/out-of-state tuition. Given a state and college budget constraint, we see that states with higher in- and out-of-state tuition spend significantly more per student on higher education.

[17] In the data, I observe the 25th and 75th percentiles and define the median as the midpoint of these two scores.

B Calibration

This section reports additional details and robustness exercises for the model calibration presented in *Section 5*. All figures and tables are shown in *Appendix D*.

B.1 Colleges

B.1.1 Reciprocity Agreements

Full Reciprocity – Reciprocity agreements are not automatic, and students must apply and prove they qualify. The only full statewide reciprocity agreements are between Minnesota and Wisconsin, and Minnesota and North Dakota, for all public colleges. Minnesota and South Dakota had a past agreement, but South Dakota exited the agreement in Fall 2024, and Minnesota will exit in Fall 2025.[18]

Regional Compacts – There are four large regional compacts: (1) Western Undergraduate Exchange, (2) Midwest Student Exchange Program, (3) Academic Common Market (Southern region), and (4) New England Regional Student Program.

These compacts differ from full state-to-state reciprocity, as they offer discounts for in-state tuition at a specific set of schools and for a specific set of majors. For instance, the Academic Common Market only offers tuition discounts for majors not offered in the home state, and the New England Regional Student Program provides a 25%–75% discount, which varies by campus at participating colleges. I exclude these four agreements from the analysis.

Other – There also exist reciprocity agreements exclusively for residents of specific border counties, campuses, or professional schools. I exclude these from the analysis. An example of a larger county-based agreement is that between Indiana and Ohio, which is active from 2023 to 2025. However, this agreement mostly includes community colleges, which are excluded

[18] Minnesota also has full tuition reciprocity with the Canadian province of Manitoba.

from the sample of colleges in this analysis. In general, most county-specific agreements are geared toward community colleges.

B.1.2 Cost Function

TBD

B.1.3 Quality Function

TBD

B.2 External Validation

Here I report regression details and robustness specifications for the elasticities reported for external validation in *Section 5.6.2*. *Table D.2* in *Appendix D* reports all results. The preferred specification used in *Section 5.6.2* is reported in column (1). The remaining columns show varying robustness and instrumental variable specifications.

The instrumental variable for spending per student is constructed as in Deming and Walters (2017) and described in *Section 5.6.2*. The instrumental variable for wages is a standard Bartik-style instrument, where lagged industrial shares by state are interacted with national wage changes by industry. The coefficients of interest are quite stable across all specifications.

C Computation

C.1 Competitive Equilibrium

The solution algorithm for solving the competitive equilibrium proceeds as follows,

1. Provide an initial guess for aggregate labor demand in each location, $Y(c, \ell)$ and $Y(nc, \ell)$, and for college qualities Q_ℓ.

    1.1 Solve for all other endogenous objects using the model's equilibrium equations.

2. Solve the household's lifecycle problem backwards and obtain values $W^{nc}(a, h, \ell_w)$ and $W^c(a, h, \ell_c, \ell_w)$ for non-college and college workers, respectively.
3. Solve discrete choice problems sequentially, working backwards.
4. Solve for admissions cutoffs that clear in-/out-of-state capacity at each college. This requires a modified solution to the sequence of discrete choice problems.
5. Update guesses using a small $d \in (0, 1)$ weighting on the newly computed objects.
6. Repeat steps 2 through 5 until convergence is achieved in the sup-norm.

C.2 Calibration

The parameters to be internally estimated consist of the preference parameters, $[\gamma_1^e, \gamma_2^e, \gamma_1^w, \gamma_2^w, \gamma_3^w, \sigma_e, \sigma_{cq}, \sigma_{cw}, \sigma_{ncw}]$, initial distribution parameters $[\mu_a, \sigma_a, \mu_p, \sigma_p]$, college fixed productivity for each state $\bar{q}_\ell$, and the elasticity of college quality with respect to spending per student, θ. Call this vector of parameters Θ. Define the loss function to minimize as $\mathcal{L}(\Theta) \equiv \big(m(\Theta) - \hat{m}\big)' \mathbf{W} \big(m(\Theta) - \hat{m}\big)$, where $m(\Theta)$ is the vector of model-simulated moments, $\hat{m}$ is the vector of data-equivalent moments, and $\mathbf{W}$ is a weighting matrix. I set the weighting matrix to be diagonal and inversely proportional to the standard deviation of the estimated values of $\hat{m}$.

The internal calibration procedure then involves a nested fixed point problem and proceeds as follows,

1. Guess θ and the vector $\bar{q}_\ell$.
2. Solve for the set of parameters $[\gamma_1^e, \gamma_2^e, \gamma_1^w, \gamma_2^w, \gamma_3^w, \sigma_e, \sigma_{cq}, \sigma_{cw}, \sigma_{ncw}, \mu_a, \sigma_a, \mu_p, \sigma_p]$ to minimize the distance between the model and data.[19]
3. Solve for $\bar{q}_\ell$ to match the earnings growth rate in the model and data using a bisection root-finding routine.
4. Repeat steps 2 and 3 until convergence on $\bar{q}_\ell$ is achieved.
5. Shock the model with an exogenous 1 percentage point increase in spending at each college. Compare the change in earnings to the elasticity estimated in the data. Using a bisection routine, find θ that matches the model elasticity to the data.
6. Repeat steps 2 through 5 until convergence in θ is achieved.

C.3 State Government

TBD

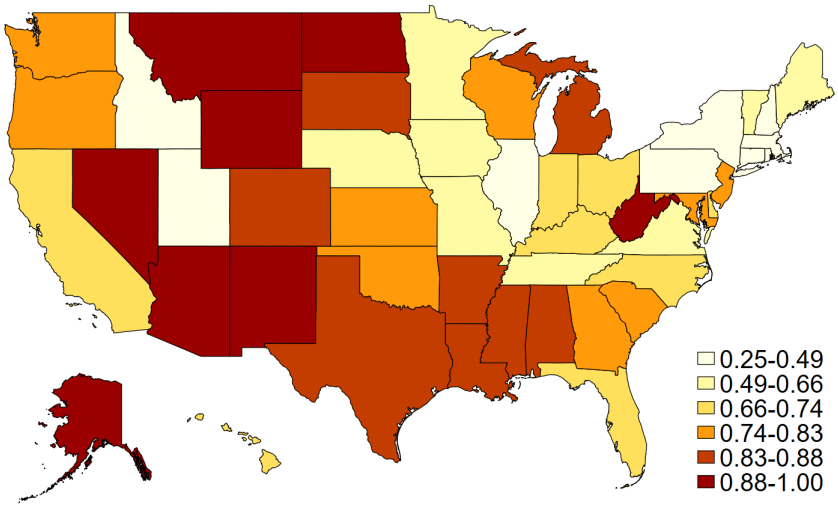
C.4 Federal Government

TBD

[19] I use the TikTak global optimization method from Guvenen and Ozkan (2021) with 500 starting points and Nelder-Mead as the local optimizer to solve for this set of parameters.

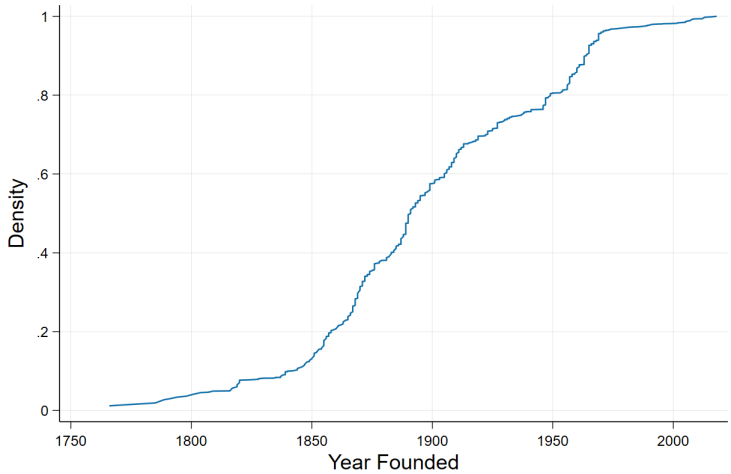
D Additional Figures and Tables

Figure D.1: **Percentage of Total Students Attending Public College**



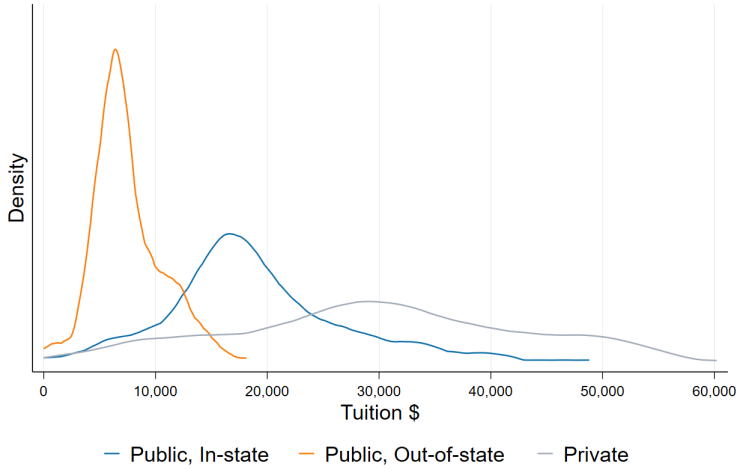
Notes: This figure plots the percentage of total college students enrolled at a public college across states. Data is taken from the Integrated Postsecondary Education Data System (IPEDS) for the 2017-2018 academic year.

Figure D.2: **Cumulative Density Function of College Founding Dates**



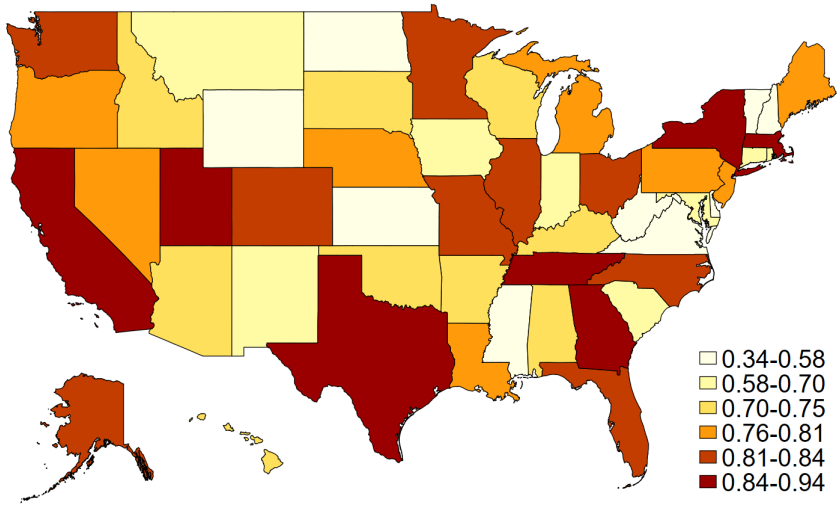
Notes: This figure cumulative percentage of college attendance by founding data. Data is taken from the Integrated Postsecondary Education Data System (IPEDS) for the 2017-2018 academic year for enrolled students and own source for founding dates.

Figure D.3: **Distribution of Tuition by Type**



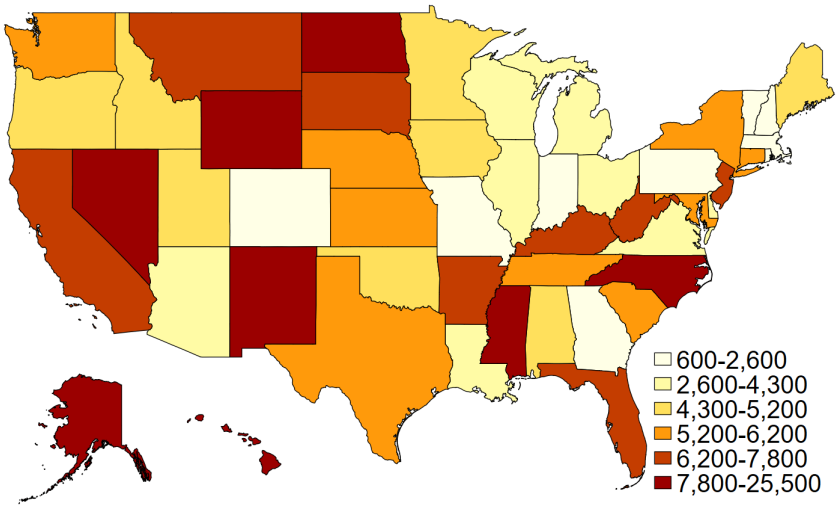
Notes: This figure plots the distribution of tuition by in-state public, out-of-state public, and private. Data is taken from the Integrated Postsecondary Education Data System (IPEDS) for the 2017-2018 academic year.

Figure D.4: **Percentage of Alum Living in State**



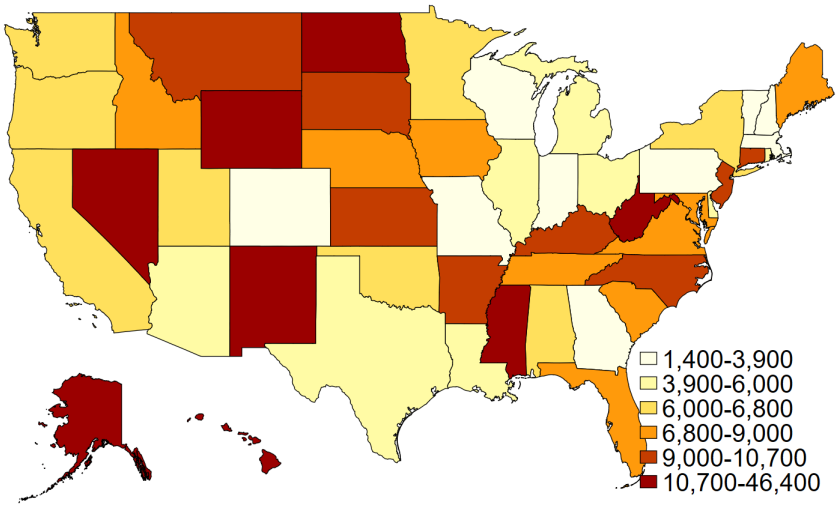
Notes: This figure plots the percentage of college grads working in the same state in which they attended college. Data is from Baccalaureate and Beyond.

Figure D.5: **Cost of Produced College Graduates**



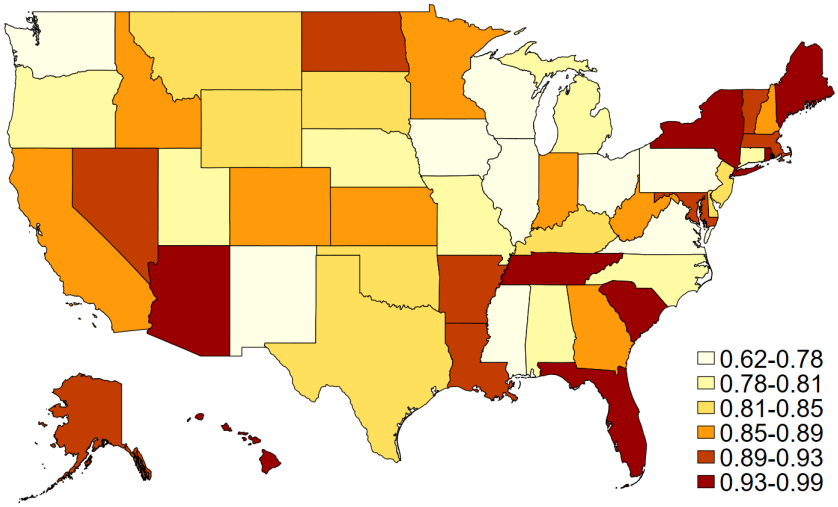
Notes: This figure plots the cost of producing a college graduate. Data open access from Conzelmann et al. (2023).

Figure D.6: **Cost of Retained College Graduates**



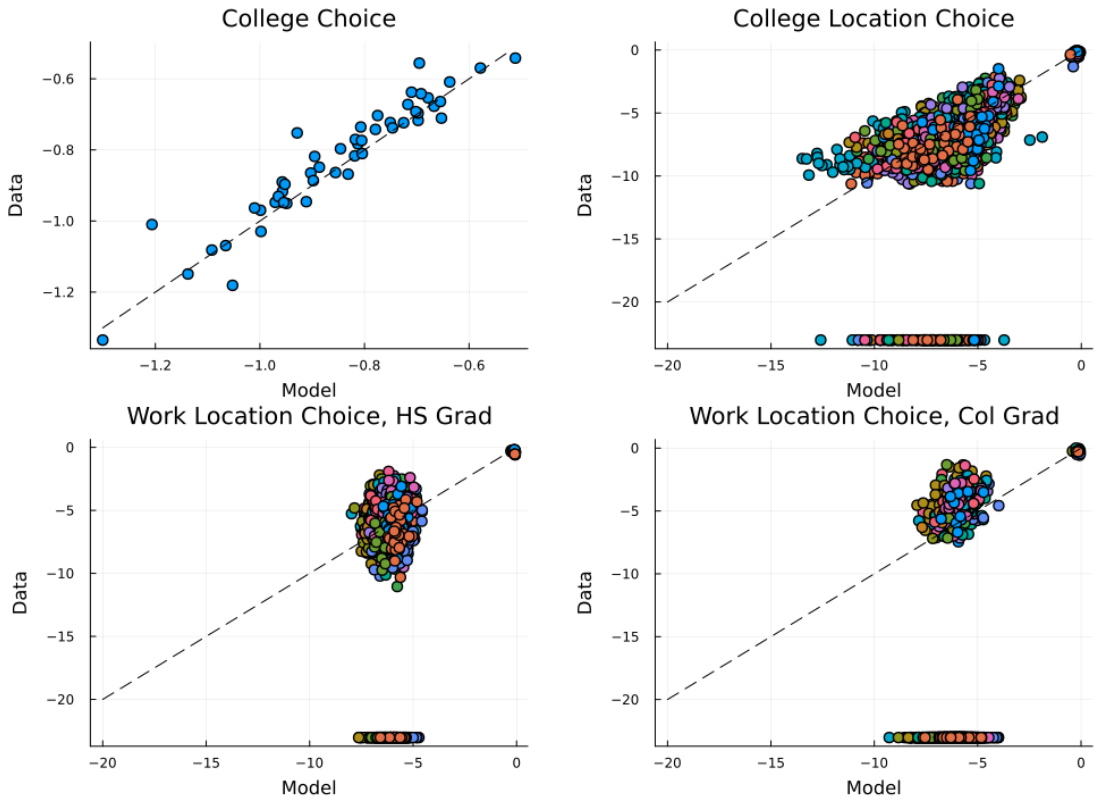
Notes: This figure plots the cost of retaining a college graduate. That is, the cost of educating a student that then remains to work in state once graduating. Data open access from Conzelmann et al. (2023).

Figure D.7: **Percentage of Total Revenues from Tuition or Government Sources**



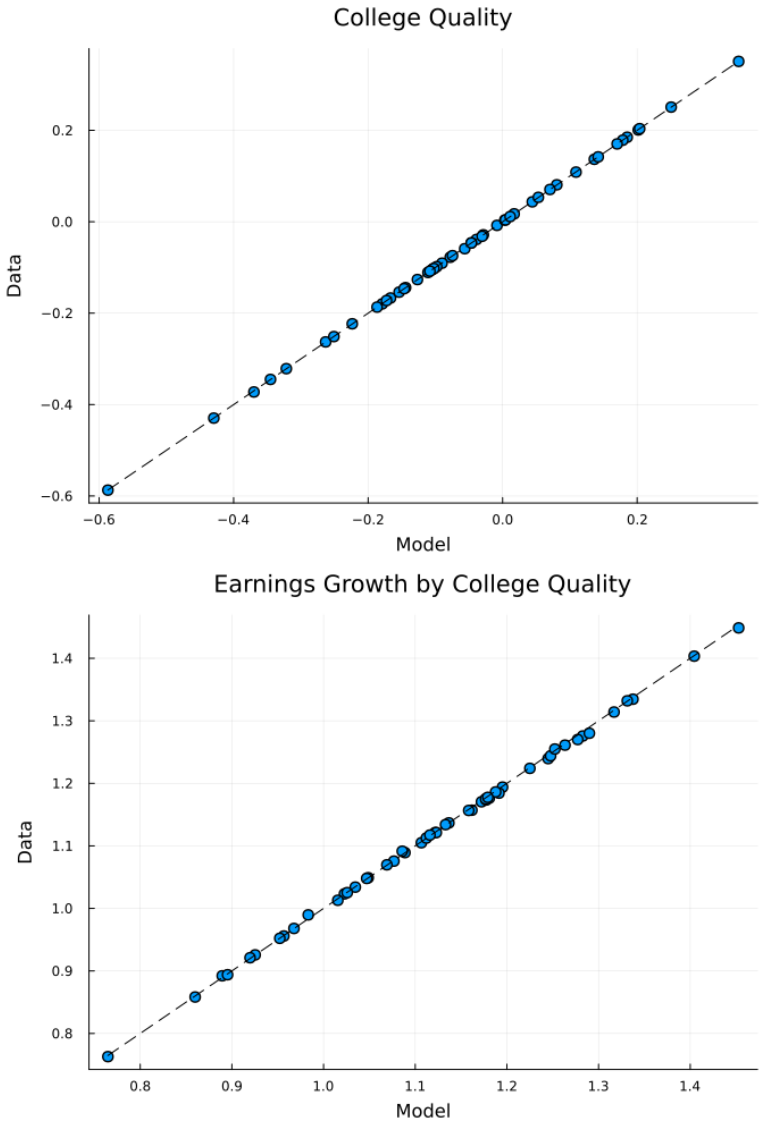
Notes: This figure plots the percentage of total college revenues received from tuition, or government sources (excluding tuition subsidies and grants). Data is taken from the Integrated Postsecondary Education Data System (IPEDS) for the 2017-2018 academic year.

Figure D.8: **Model Fit – Choice Probabilities**



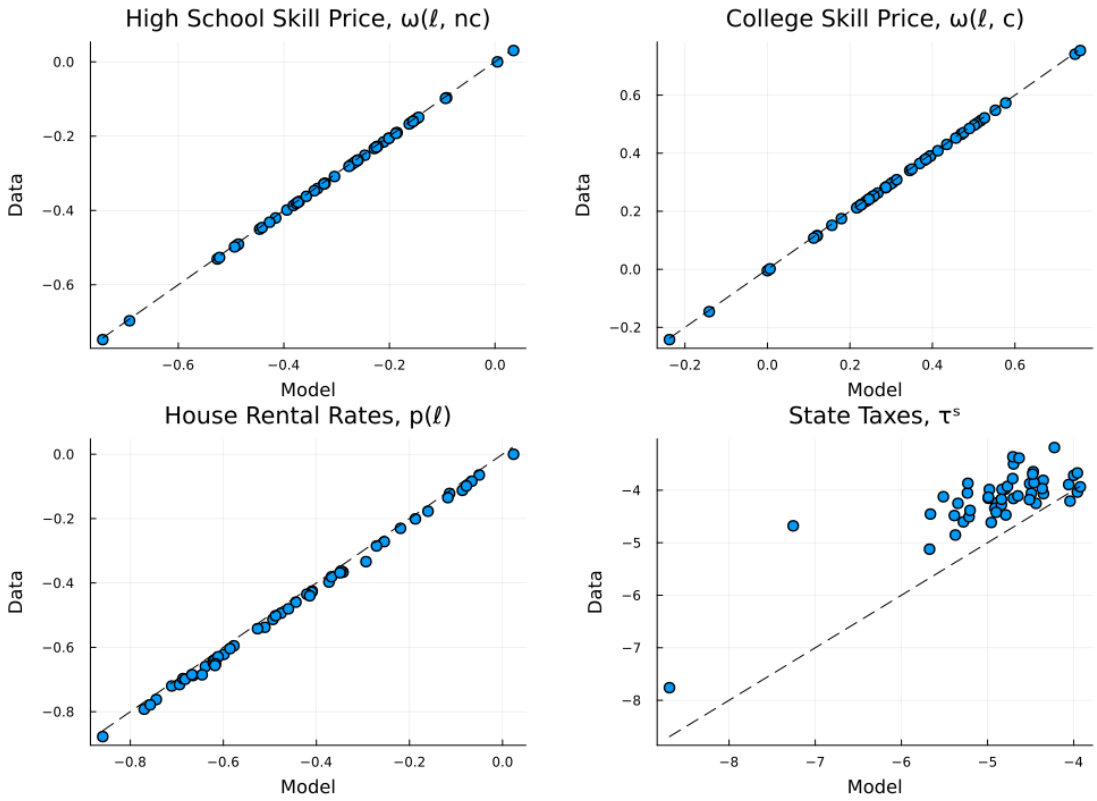
Notes: The model with data choice probabilities for each discrete choice. The 45-degree line represents a perfect fit between the model and data.

Figure D.9: **Model Fit – Earnings Growth and College Qualities**



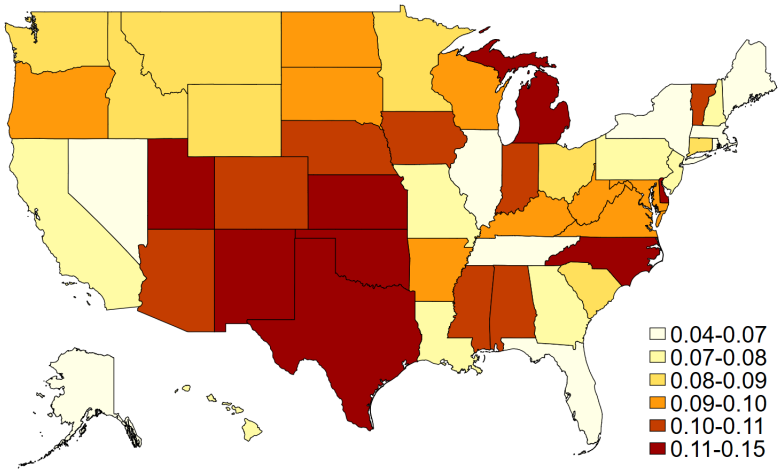
Notes: The model with data choice probabilities for each discrete choice. The 45-degree line represents a perfect fit between the model and data.

Figure D.10: **Model Fit – Skill Prices, Rental Prices, and Tax Rates**



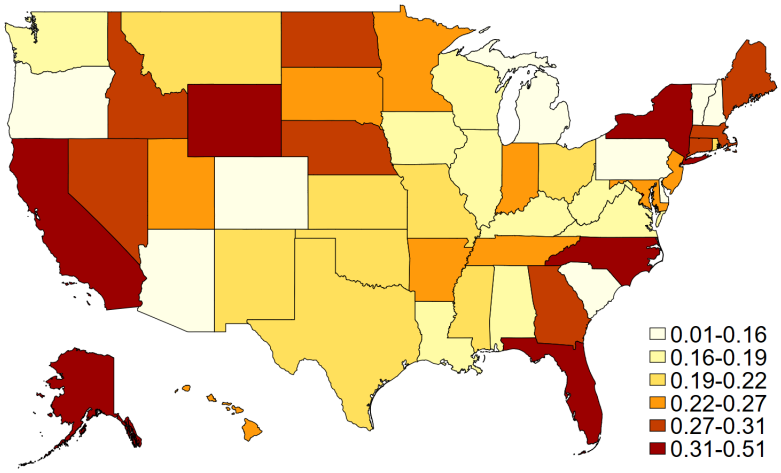
Notes: The model with data choice probabilities for each discrete choice. The 45-degree line represents a perfect fit between the model and data.

Figure D.11: **Higher Education Expenditures as Proportion of Total State Outlays**



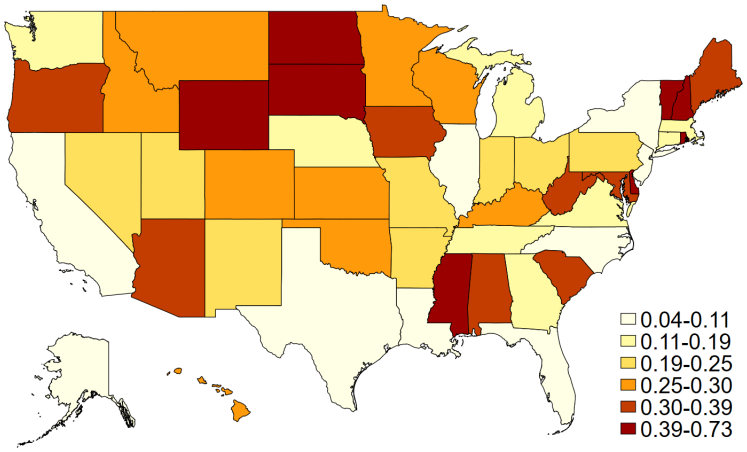
Notes: This figure plots the proportion of total state outlays allocated to higher education across the 50 U.S. states. Data are taken from the U.S. Census Bureau's State and Local Government Finances. Average values are reported for fiscal years 2016-2018 to account for short-run fluctuations.

Figure D.12: **Proportion of College Budget Financed by State Appropriations**



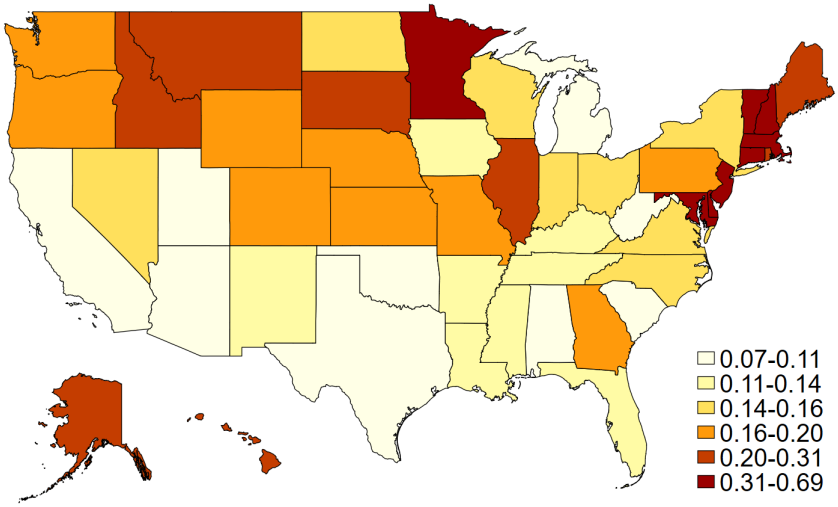
Notes: This figure plots the proportion of total college expenditures financed by state government revenues across the 50 U.S. states. Data are taken from the Integrated Postsecondary Education Data System (IPEDS) for the 2017-2018 academic year.

Figure D.13: **Percentage of Student Body From Out-of-State**



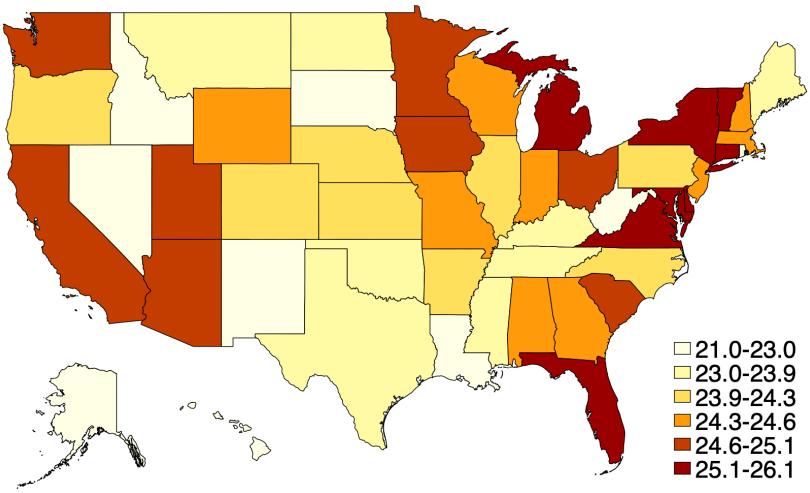
Notes: This figure plots the percentage of out-of-state students across the 50 U.S. states. Data are taken from the Integrated Postsecondary Education Data System (IPEDS) for the 2017-2018 academic year.

Figure D.14: **Percentage of Students Leaving Home State for College**



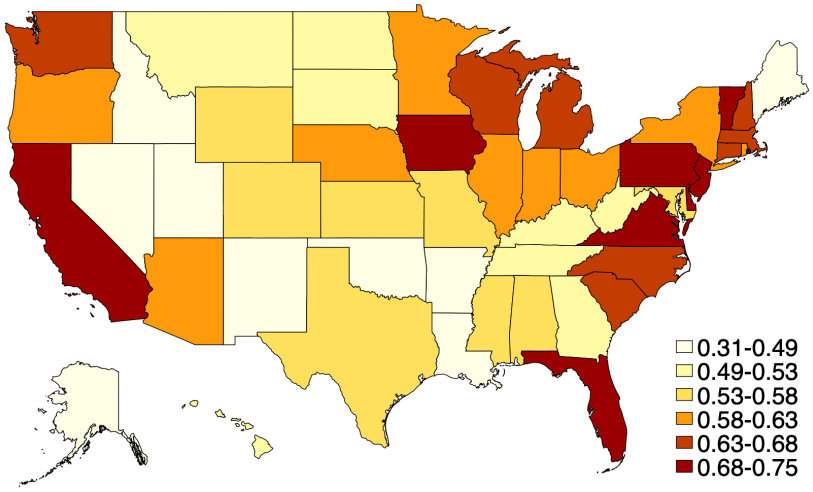
Notes: This figure plots the percentage of high students that study outside state of high school graduation, across states. Data is taken from the Integrated Postsecondary Education Data System (IPEDS) for the 2017-2018 academic year.

Figure D.15: **Median ACT Composite Scores**



Notes: This figure plots the median American College Testing (ACT) composite scores at public colleges across the 50 U.S. states. Composite scores are an average of English, Mathematics, Reading, and Science scores. Data are taken from the Integrated Postsecondary Education Data System (IPEDS) for the 2017-2018 academic year.

Figure D.16: **6-year Graduation Rate**



Notes: This figure plots the 6-year graduation rate at public colleges across the 50 U.S. states. Data is taken from the Integrated Postsecondary Education Data System (IPEDS) for the 2017-2018 academic year.

Figure D.17: **University of Minnesota–Twin Cities Wikipedia Page**



Notes: This figure shows the Wikipedia page of the University of Minnesota–Twin Cities used for scraping founding dates.

Table D.1: **Motivating OLS – Policy Variable Correlations**

	(1) ln(spend per)	(2) ln(spend per)
$\ln(T^{in})$	0.195	
	(0.084)	
$\ln(\%in-state)$		0.381
		(0.043)
constant	9.52	11.35
	(0.000)	(0.000)
R^2	0.06	0.08
observations	50	50

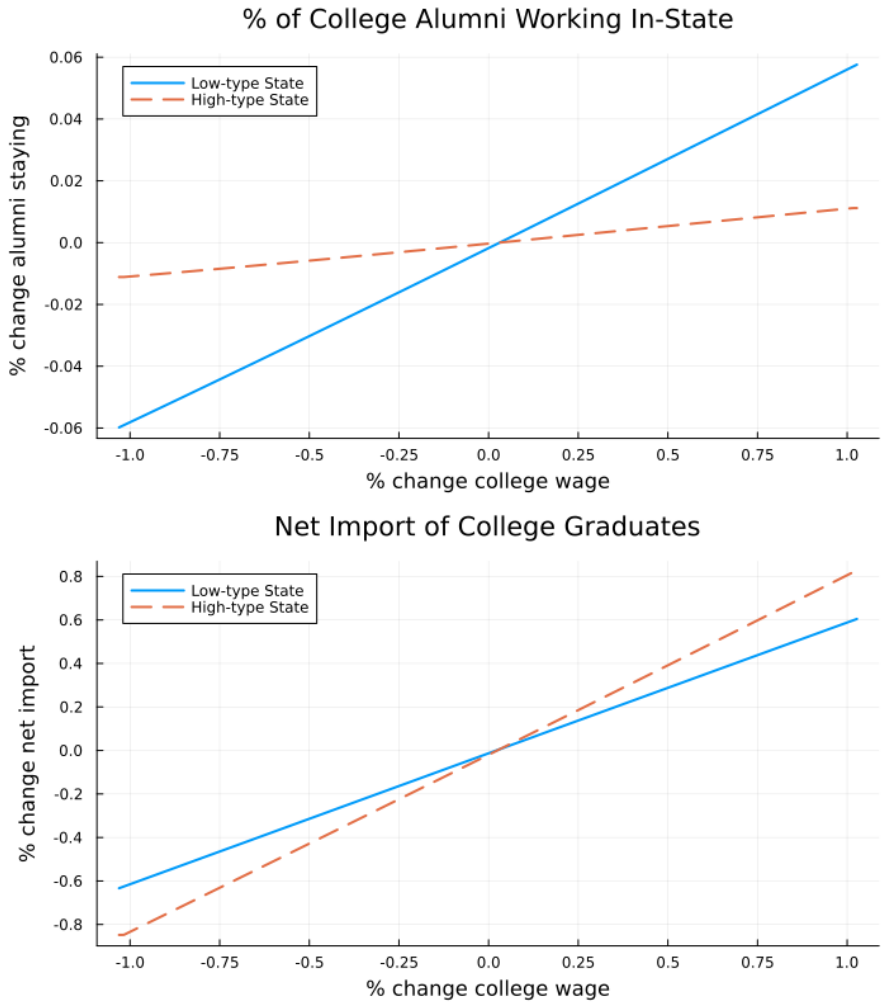
Notes: P-values show in brackets. Data are taken from the Integrated Postsecondary Education Data System (IPEDS) for the 2017-2018 academic year and the American Community Survey (ACS).

Table D.2: **External Validation Estimation and Robustness**

	(1)	(2)	(3)	(4)	(5)	(6)
	ln(% leave grad)	ln(% leave grad)	ln(% leave grad)	ln(% leave grad)	ln(% leave grad)	ln(% leave alum)
$\ln(T^{in})$	0.059	0.061	0.056			
	(0.007)	(0.077)	(0.038)			
$ln(col\ earn)$	-0.164	-0.243			-0.178	-1.36
	(0.146)		(0.016)		(0.004)	(0.000)
$L5.\ln(spend\ per)$	-0.216	-0.211	-0.244	-0.181		
	(0.141)	(0.031)	(0.021)	(0.023)		
$L5.\ln(T^{out})$	-0.134	-0.133				
	(0.150)	(0.053)				
$\%\ out-state$	2.01	2.01	2.08	2.10	2.09	-3.79
	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$admit\ rate$	-0.493	-0.503	-0.480	-0.525	-0.545	
	(0.095)	(0.022)	(0.102)	(0.005)	(0.003)	
$med\ act$	0.008	0.008	0.012	0.017	0.015	
	(0.642)	(0.567)	(0.497)	(0.154)	(0.171)	
$unemp$	0.026	0.025	0.484	0.622	0.436	0.082
	(0.967)	(0.958)	(0.507)	(0.218)	(0.355)	(0.651)
$constant$	1.74	-0.146	1.45	-1.45	-0.461	12.70
	(0.260)	(0.786)	(0.318)	(0.013)	(0.566)	(0.002)
IV	No	No	No	Yes	Yes	No
Observations	692	692	826	829	829	50

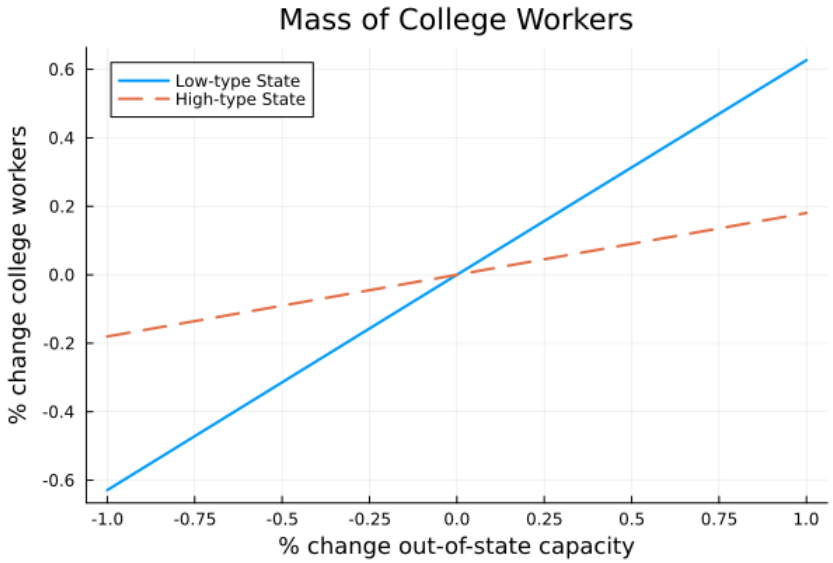
Table D.3: *Notes:* P-values show in brackets. Data is taken from the Current Population Survey, Integrated Postsecondary Education Data System, and Conzelmann et al. (2023). Specifications (1) to (5) feature clustered standard errors and fixed effects.

Figure D.18: **Changes in In-State Alumni and Net College Grad Imports**



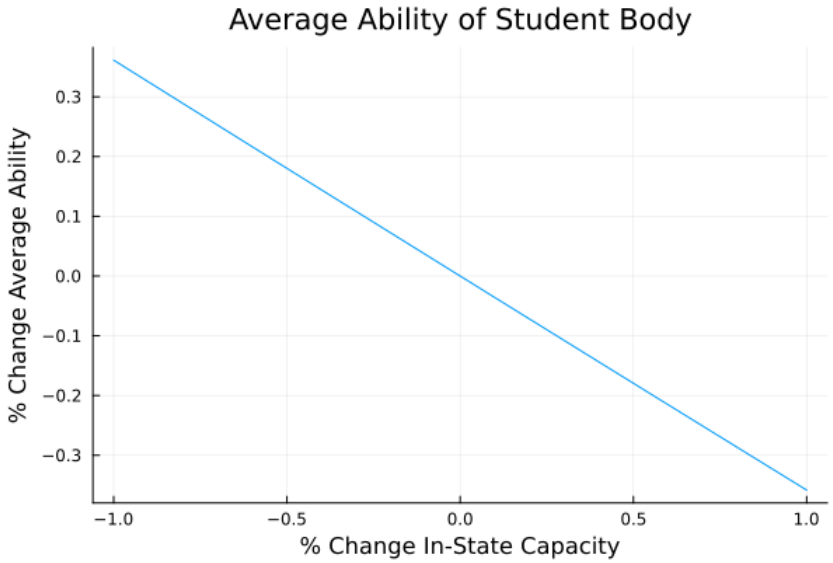
Notes: This figure plots the percentage point change in the percentage of alumni who stay to work in-state (top panel) and the net import of college graduates (bottom panel) by percentage point change in college skill prices.

Figure D.19: **Changes in College Workers and Out-of-State Capacity**



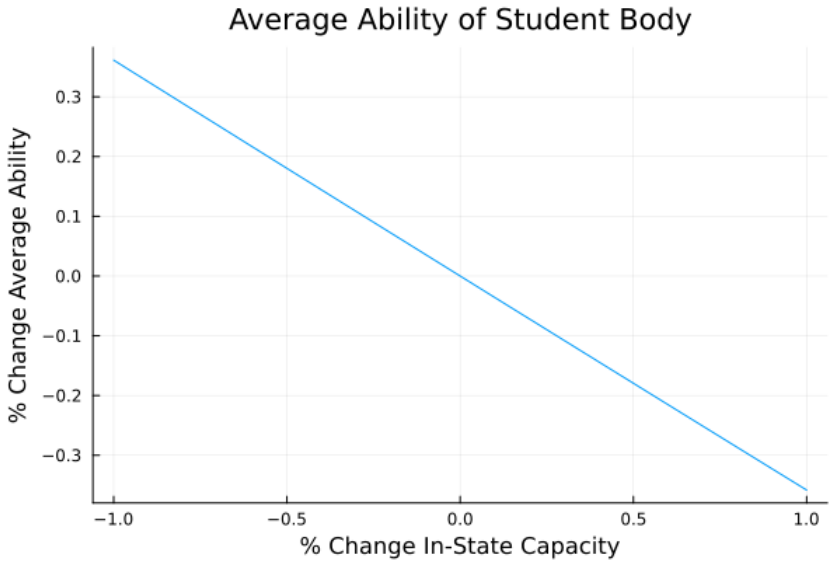
Notes: This figure plots the percentage point change in the percentage of college workers in a state by percentage point change in out-of-state capacity.

Figure D.20: **Changes in Peer Effects and Spending per Student**



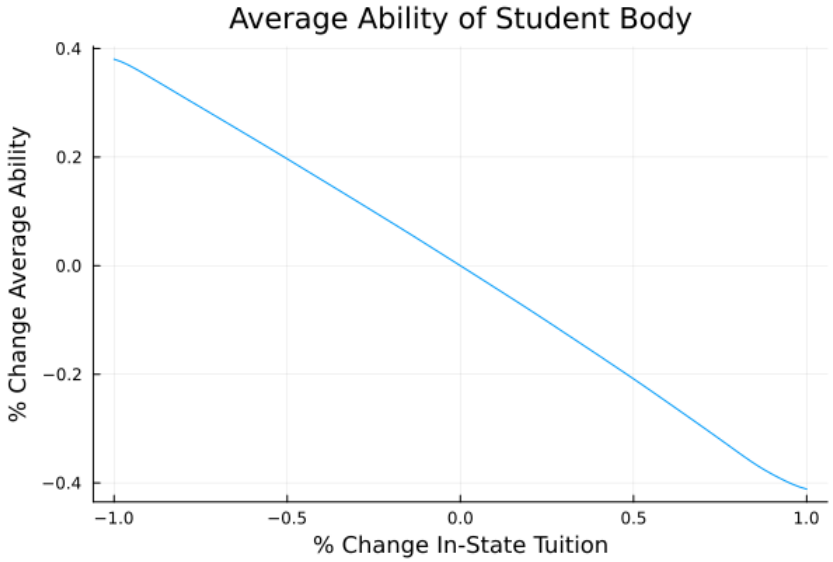
Notes: This figure plots the percentage point change in average ability of college body by percentage point change of spending per student. Spending per student is change exogenously and the government budget constraint in general does not hold here.

Figure D.21: **Changes in Peer Effects and In-State Capacity**



Notes: This figure plots the percentage point change in average ability of college body by percentage point change of in-state capacity.

Figure D.22: **Changes in Peer Effects and In-State Tuition**



Notes: This figure plots the percentage point change in average ability of college body by percentage point change of in-state tuition.