UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

Higher Education Dynamics: Assessing the Nexus of College Costs, Sports, and Amenities

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

AHMED EL FATMAOUI

Norman, Oklahoma

2025

Higher Education Dynamics: Assessing the Nexus of College Costs, Sports, and Amenities

A DISSERTATION APPROVED FOR THE

DEPARTMENT OF ECONOMICS

BY THE COMMITTEE CONSISTING OF

Dr. Tyler Ransom, Chair

Dr. Pallab Ghosh, Co-chair

Dr. Greg Burge

Dr. Anni Yang

DEDICATED

To my beloved mother,

FATIMA MANSOUR,

whose love, wisdom, and guidance have been my greatest strength.

Your sacrifices and unwavering support have made this journey possible.

Your being is unparalleled in my universe.

Acknowledgments

The completion of this dissertation represents not only years of research and academic pursuit, but also the culmination of support and encouragement from many individuals to whom I am deeply indebted.

I would like to express my sincere gratitude to my dissertation committee co-chairs, Dr. Tyler Ransom and Dr. Pallab Ghosh, for their exceptional guidance, rigorous feedback, and steadfast support throughout this journey. Their expertise and mentorship have been invaluable in shaping both this research and my development as an economist. I am equally grateful to my committee members, Dr. Greg Burge and Dr. Anni Yang, whose insightful comments significantly strengthened the analytical foundation and empirical approach of this work.

My deepest appreciation extends to the Department of Economics at the University of Oklahoma for providing an intellectually stimulating environment and the resources necessary to conduct this research.

I owe an immeasurable debt of gratitude to my family, whose unwavering belief in my abilities has sustained me throughout this endeavor. To my parents, thank you for instilling in me the value of education and perseverance. To my sisters and brothers, your encouragement has been a constant source of motivation. I am also grateful to my aunts, uncles, cousins, and extended family members who have supported me in countless ways during this academic pursuit.

I would also like to acknowledge my fellow doctoral colleagues whose camaraderie and intellectual discussions have enriched both my research and personal growth. Our conversations, whether focused on econometric methods or much-needed distractions, have made this journey more meaningful.

This academic achievement stands as testament to the collective support of all those mentioned and many others who have contributed in various ways to my intellectual and professional development. Any shortcomings or errors in this work remain entirely my own.

Contents

List of Tables

List of Figures

Abstract

This dissertation examines three distinct aspects of higher education and public policy. The first chapter investigates the differential impact of state appropriations on STEM and non-STEM degree completion at U.S. public institutions. Using a Bartik-style instrumental variables approach with panel data from 2003-2019, I find that a 10% increase in state funding leads to a 3.4% increase in STEM degrees, primarily affecting male students, science majors, and non-selective institutions, while having minimal impact on non-STEM degrees. The second chapter explores how recreational marijuana legalization (RML) affects college enrollment, finding that RML increases first-time undergraduate enrollments by 4.6-9% without compromising degree completion or graduation rates. This effect is driven by out-of-state enrollments in non-selective public colleges, suggesting marijuana functions as a consumption amenity that enhances college competitiveness. The third chapter analyzes the prevalence of student-athletes at elite private universities, documenting that these institutions allocate a disproportionate share of enrollment capacity to varsity athletes compared to larger public universities, with athletes often coming from more advantaged socioeconomic backgrounds. Together, these chapters provide insights into how public funding, state policies, and institutional priorities shape access to and outcomes in higher education.

Chapter 1

Dollars and Degrees: The Asymmetric Impact of State Appropriations on STEM and Non-STEM Fields

To meet the demand for STEM professionals and harness the social and economic spillovers of STEM education, the President's Council of Advisors on Science and Technology (PCAST) recommended a 34% increase in STEM graduates over current rates (PCAST, 2012). This is due to the significant economic benefits of STEM education, which are crucial for economic prosperity and social mobility (Bacovic, Andrijasevic, & Pejovic, 2022; Carnevale, Smith, & Melton, 2011; Wolniak, Seifert, Reed, & Pascarella, 2008). Approximately 44% of US college students enroll in public 4-year universities, with a higher proportion of these students coming from low-income backgrounds compared to those at private institutions (National Student Clearinghouse Research Center, 2024; Pew Research Center, 2019).

Public colleges and universities have experienced a decline in state appropriations, particularly following the Great Recession. Between 2001 and 2019, state appropriations for public higher education institutions in constant 2019 dollars decreased by 2.2%, declining from $82.6 billion to $80.8 billion (Cummings, Laderman, Lee, Tandberg, & Weeden, 2021). The literature examines how state funding for higher education affects various aspects of student and institutional outcomes, including faculty composition, enrollment, completion rates, student debt, and credit scores (Chakrabarti, Gorton, & Lovenheim, 2020; Deming & Walters, 2018; Hinrichs, 2022). This paper contributes by highlighting the asymmetric effects of state appropriations on STEM and non-STEM degree completion.[1]

Given the resource-intensive nature of STEM fields (Altonji & Zimmerman, 2019; Hemelt, Stange, Furquim, Simon, & Sawyer, 2021), I hypothesize that state appropriations have a significantly higher effect on STEM degree completion through two primary mechanisms. First, state funding enables colleges to adapt to labor market demands by introducing new programs, which are particularly pertinent for STEM majors. Second, the allocation of resources may have a more pronounced impact on STEM fields due to their greater reliance on production inputs.[2]

To test these hypotheses, I construct a unique panel dataset from various sources and employ a Bartik-like instrumental variable (IV) approach commonly used in related literature (Chakrabarti et al., 2020; Deming & Walters, 2018; Hinrichs, 2022). The panel data includes the number of degrees conferred in specific majors at each public 4-year institution, as well as these institutions' expenditures and tuition from the Integrated Postsecondary Education Data System (IPEDS) from 2003 to 2019 (National Center for Education Statistics, 2019, 2022a). The dataset also incorporates state-level higher education spending from the State Higher Education Executive Officers Association (2023), and economic and demographic variables from the University of Kentucky Center for Poverty Research (2023). Additionally, the panel includes institution-specific county-level economic and demographic variables from the Bureau of Economic Analysis and the U.S. Census Bureau.[3]

The analysis addresses endogeneity concerns stemming from reverse causality, selection of high-achieving students into well-funded institutions, and potential omitted variable bias through the implementation of a Bartik-like instrument. This instrument sidesteps the endogeneity issue

[1]On average, public universities received approximately \$15.4 million less in state funding from 2015 to 2019 compared to 2003 to 2007 (see summary statistics in Table 1.1). From 2001 to 2019, full-time equivalent (FTE) student enrollment in public higher education institutions increased by 25.3%, growing from 8.7 million to 10.9 million students (Cummings et al., 2021).

[2]In their analysis of the Florida State University System, Altonji and Zimmerman (2019) reveal substantial program-level cost disparities, with per-credit expenditures ranging from \$450 for engineering and health sciences to \$200-\$250 for social sciences, mathematics, business, and psychology. Reductions in state funding could also lead to higher tuition fees, potentially altering the student preference for STEM majors compared to non-STEM majors.

[3]For classification of programs as STEM (Science, Technology, Engineering, and Mathematics), I use the STEM Designated Degree Program list from the U.S. Department of Homeland Security (DHS).

by interacting the historical dependence of universities on state funding (shares) with the current state-level appropriations (shifts). My instrument differs from those in prior works in two key aspects: first, it extends 12 years before the first year of the data, and second, it constructs the share using five historical years (1987 to 1991) to provide a more accurate estimate of the historical shares.

The findings indicate that state appropriations have a significant effect on STEM degree completion and little effect on non-STEM field completion. The elasticities range between 0.24 and 0.34, depending on the cohort of students graduating in 4 (normal time), 5, or 6 years. The upper bound elasticity indicates that a 10% increase in state appropriations leads to a 3.4% increase in the number of students graduating with a bachelor's degree in each STEM major.

Next, I examine how the effect varies by institutional selectivity, STEM subfields, race, and gender. The observed effect is predominantly localized within non-selective institutions, attributable to their limited capacity to diversify revenue streams or access alternative funding sources (Bound, Braga, Khanna, & Turner, 2019). I also find that the effect is most relevant to science-related STEM fields, with a significant effect for minorities only in engineering majors. The gender analysis indicates that state appropriations have no significant effect on female STEM major completion.[4]

To investigate the driving mechanisms, I first examine the effect of state funding on two factors: spending (e.g., institutional grants and academic support), and the number of programs offered. I then analyze how these factors differentially influence STEM and non-STEM degree completion. The mechanisms analysis reveals that state appropriations have a significant effect on college spending and the number of fields offered. These factors, in turn, have a substantially greater effect on STEM field degree completion compared to non-STEM fields.[5]

The findings remain robust to various checks. Following Goldsmith-Pinkham, Sorkin, and

[4]This result corroborates the literature suggesting that the gender gap in STEM fields is due to factors such as social expectations and stereotypes, peer group preferences, professional expectations regarding the personality of STEM graduates, and females being more risk-averse towards low grades (Ahn, Arcidiacono, Hopson, & Thomas, 2024; Brenøe & Zölitz, 2020; Cheryan, Master, & Meltzoff, 2015; Crosnoe, Riegle-Crumb, Field, Frank, & Muller, 2008; Shapiro & Williams, 2012).

[5]The mechanisms involve examining the effect of the number of majors offered, and university expenditures on STEM and non-STEM degree completion. The Bartik-like instrument is not feasible here because the size of public universities remains nearly constant over time. For instance, flagship universities consistently maintain the largest share of total state-level university expenditure. I assess the omitted variable bias of the OLS estimate by presenting breakdown points following Diegert, Masten, and Poirier (2022) and Oster (2019)

Swift (2020), I demonstrate that the results are robust to the exclusion of institutions with the largest Rotemberg weights (a measure of relevance in generating the identifying variation in the instrument). The findings are also robust to the use of an alternative IV of aggregate state-level appropriations, following Bound et al. (2019); Bound, Braga, Khanna, and Turner (2020); Webber (2017). A placebo analysis, which assigns false STEM designations to non-STEM majors, shows that the observed effects are not random.

The remainder of this paper is organized as follows: Sections 2 and 3 introduce the literature and conceptual framework. Section 4 describes the data and presents descriptive statistics. Section 5 outlines the empirical strategy. Section 6 presents the main results and explores potential mechanisms. Section 7 examines heterogeneous effects across various dimensions. Section 8 presents robustness checks, and Section 9 concludes.

1.1 Literature Review

This paper contributes to several strands of literature examining the impacts of state funding on higher education outcomes. Numerous studies have documented the decline in state support for public higher education, particularly following the Great Recession of 2008 (Bound et al., 2019; Chakrabarti et al., 2020). In response to these funding cuts, public universities have employed various strategies to maintain their operations and educational quality.

Bound et al. (2020) find that public research universities responded to a 10% reduction in state funding with a 16% increase in foreign undergraduate enrollment. This finding highlights how institutions leverage out-of-state tuition revenue to offset state funding losses. Similarly, Webber (2017) shows that state divestment in higher education is associated with increases in in-state tuition at public universities. On the expenditure side, Deming and Walters (2018) and Bound et al. (2019) document reductions in spending at public universities in response to state funding cuts.

The literature has also examined how changes in state funding affect student outcomes. Chakrabarti et al. (2020) show that students take on more debt at public universities experiencing state funding cuts, with negative implications for their long-term financial outcomes. Importantly, the impacts of funding changes are not uniform across all students and institutions. Bound et al. (2019) and Chakrabarti et al. (2020) find that the negative effects of funding cuts tend to be larger for lower-income students and at less selective institutions.

While the literature has extensively documented the general impacts of state funding changes, less attention has been paid to how these effects might differ across academic disciplines. This gap is particularly relevant given the growing policy interest in STEM education and the resource-intensive nature of STEM fields. Altonji and Zimmerman (2019) and Hemelt et al. (2021) provide evidence on the differential costs of producing graduates across majors. They find that STEM fields, particularly engineering and health sciences, have substantially higher per-credit and per-graduate costs compared to fields like business or social sciences. This cost differential suggests that STEM programs may be more sensitive to changes in institutional resources. Hinrichs (2022) examines the effects of state appropriations on employment at higher education institutions, finding significant impacts on part-time faculty employment, but does not specifically address differential effects across academic fields.

The importance of STEM education for economic growth and social mobility has been highlighted by several studies (Bacovic et al., 2022; Carnevale et al., 2011; Wolniak et al., 2008). However, the literature has not yet fully explored how state funding changes might differentially affect STEM vs. non-STEM degree production. The mechanisms through which state funding impacts degree production in different fields are not well understood. Finally, there is a need for more research on how the effects of state funding changes might vary across different types of institutions and student demographics within the context of STEM education.

This paper aims to address these gaps by examining the differential impact of state appropriations on STEM and non-STEM degree completion, exploring the mechanisms behind these effects, and investigating heterogeneity across institution types and student characteristics. By doing so, it contributes to the understanding of how public funding shapes the production of human capital in critical fields for economic growth and innovation. The study builds on the existing literature by leveraging a unique panel dataset and employing a Bartik-style instrumental variables approach, similar to methods used in recent studies (Bound et al., 2019; Chakrabarti et al., 2020; Deming & Walters, 2018; Hinrichs, 2022; Webber, 2017).

1.2 Conceptual Framework

To analyze how state appropriations affect degree completion in public universities, I develop a model of a college's optimization problem. The model, detailed in Appendix A.3, provides insights into the mechanisms through which funding impacts degree production, with a focus on differentiating between STEM and non-STEM fields.[6]

Consider a representative public university that aims to maximize degree production subject to a budget constraint. I hypothesize that the university's production function for degrees follows a Cobb-Douglas form, depending on instructional spending, financial aid, academic support, and the number of programs offered. The number of programs is assumed to be a function of instructional spending, capturing the idea that increased instructional resources enable universities to offer a wider array of programs.[7]

The model yields the following key prediction for the marginal effect of state appropriations on degree production:

$$\frac{\partial Y}{\partial R_s} = (\alpha + \beta + \gamma + \eta\theta)\frac{Y}{R_s + R_n} \tag{1.1}$$

where Y is the number of degrees produced, R_s is state appropriations, R_n is revenue from other sources, and α, β, γ, η, and θ are parameters representing output elasticities with respect to different inputs. This framework suggests two reasons why state appropriations may have a greater effect on STEM degree completion:

1. *Resource intensity*: STEM fields typically require more expensive core production inputs and more intensive tutoring and mentoring. For instance, instructional expenditures per student credit hour are 92% higher for electrical engineering compared to English-related courses, and per-credit expenditures range from \$450 for engineering and health sciences to \$200-\$250 for social sciences, mathematics, business, and psychology (Altonji & Zimmerman, 2019; Hemelt et al., 2021). Moreover, STEM courses typically demand more intensive tutoring and mentoring, as well as smaller class sizes (Bettinger & Baker,

[6]In fiscal year 2019-20, the top two revenue sources for 4-year public institutions were tuition and fees (20.29%) and state appropriations (16.69%). In contrast, in fiscal year 2007-08, state appropriations (23.83%) exceeded tuition and fees (17.93%) as the primary revenue source (National Center for Education Statistics, 2024).

[7]I employ the Cobb-Douglas function for its tractable properties and for consistency with my empirical approach and data (see Appendix A.2 for a background discussion on the objective function of public institutions).

2014; Kara, Tonin, & Vlassopoulos, 2021).This suggests that the elasticity of output with respect to instructional spending (α) and academic support spending (γ) may be larger for STEM fields.

2. *Program diversity*: The rapid development of new technologies may necessitate more frequent updates to STEM curricula (Autor & Dorn, 2013). This implies that the elasticity of programs with respect to instructional spending (θ) and the elasticity of output with respect to program offerings (η) may be larger for STEM fields.

If these hypotheses hold, I would expect $(\alpha+\beta+\gamma+\eta\theta)_{STEM} > (\alpha+\beta+\gamma+\eta\theta)_{non-STEM}$, implying a larger marginal effect of state appropriations on STEM degree production. This framework motivates my empirical strategy to estimate and compare the elasticities of STEM and non-STEM degree completion with respect to state appropriations. It also guides my investigation of the mechanisms through which state funding affects degree production, including its impact on different categories of university spending and program offerings.

The model also suggests potential sources of heterogeneity in the effects of state funding. Universities with different initial levels of resources (R_n) or different production technologies (as captured by the elasticity parameters) may respond differently to changes in state appropriations. This motivates my empirical analysis of heterogeneous effects across institution types.

1.3 Data and Descriptive Statistics

1.3.1 Data Sources

The Integrated Postsecondary Education Data System (IPEDS), administered by the National Center for Education Statistics (NCES), serves as a comprehensive source of data on various facets of postsecondary education in the United States. The system encompasses a series of surveys that collect information on enrollment trends, graduation rates, and the characteristics of institutions that are eligible to disburse federal student aid funds, including Pell Grants and Stafford Loans. As mandated, these institutions must submit their data to the NCES via the IPEDS surveys (National Center for Education Statistics, 2019, 2022a). A survey of particular significance within this collection is the Completions survey, which meticulously records the number of academic awards and degrees conferredâĂŤcategorized by program using a 6-digit Classification of Instructional Programs (CIP) codeâĂŤat each public four-year institution over a period extending from 2003 to 2019

I incorporate a comprehensive set of control variables derived from the University of Kentucky Center for Poverty Research (2023) to account for the multifaceted economic, political, and demographic characteristics at the state level. These variables encompass the state-mandated minimum wage, per capita personal income, and the poverty rate, alongside a binary indicator for the presence of a Democrat governor. Additionally, I include metrics indicative of social welfare engagement, quantified as the number of recipients per 100,000 population for programs such as the Supplemental Nutrition Assistance Program (SNAP), Medicaid, and Temporary Assistance for Needy Families (TANF). To further refine the demographic controls, I integrate data on the state population aged 18 to 24, sourced from the U.S. Census Bureau.

Complementing the state-level controls, I also introduce a series of county-level variables that may exert influence on enrollment decisions (Betts & McFarland, 1995; Card, 1993; Manski & Wise, 1983). These include the young adult population (aged 18 to 24), local per-capita personal income, and the unemployment rate. Moreover, I consider vital statistics such as births and deaths per 100,000 population, as well as the total population and its breakdown by gender and ethnicity, including male, white, black, and Hispanic demographics. The data for these county-level variables are procured from the Bureau of Economic Analysis and the U.S. Census Bureau.

Lastly, one important data source is the state budget expenditures to higher education or SHEEO Grapevine data from the State Higher Education Executive Officers Association (2023). This source provides the state and year level data on total expenditures allocated to higher education. The finance survey from IPEDS provides the college-level appropriations to each college over time, but SHEEO Grapevine total values are more complete and precise.

I use the STEM Designated Degree Program list from the U.S. Department of Homeland Security (DHS) to classify programs as STEM (Science, Technology, Engineering, and Mathematics) or non-STEM. This list includes the CIP (Classification of Instructional Programs) codes for majors that the department considers part of STEM fields (Immigration & Enforcement, 2022). I then utilize the National Center for Education Statistics CIP code crosswalks to harmonize all codes to the 2010 classification.[8]

1.3.2 Financial Trends

Table 1.1 presents summary statistics for college spending, state appropriations, number of STEM and non-STEM degrees, and number of programs offered by major type across two periods: the first and last five years of the dataset. On average, public 4-year institutions received approximately \$15.4 million less in state support from 2015 to 2019 compared to the period from 2003 to 2007, representing a decrease of 15.3%. The data indicates a significant increase in the cost of college attendance for the average university, with an approximate rise of \$2,113 (31.94%) in sticker price (published tuition and fees) and \$932 (7.56%) in net price (total cost minus all grants and aid). Concurrently, there has been an accompanying growth in institutional expenditures. This spending growth is expected due to increasing population and enrollments over time.

Table 1.1: Summary Statistics–College Spending and Prices

	2003 to 2007 (N=2765)		2015 to 2019 (N=3120)		
	Mean	SD	Mean	SD	Diff. in Means
State Appropriation ($M)	99.891	122.386	84.512	109.865	-15.380***
Total Expenditures ($M)	378.030	616.594	472.734	903.632	94.705***
Instructional Expenditures ($M)	102.937	138.909	138.006	210.066	35.070***
Academic Support Expenditures ($M)	25.907	40.942	39.786	74.053	13.879***
Institutional Grants ($M)	10.345	19.761	23.543	41.643	13.199***
N Non-STEM Degrees	1340.153	1323.535	1494.153	1597.454	154.000***
N STEM Degrees	319.187	441.898	533.467	789.146	214.279***
N of Non-STEM Programs	35.042	21.883	36.105	25.804	1.062
N of STEM Programs	12.551	9.194	14.258	11.108	1.706***

SOURCES.—All data are sourced from the Integrated Postsecondary Education Data System (IPEDS) as reported in U.S. Department of Education, National Center for Education Statistics (2024).

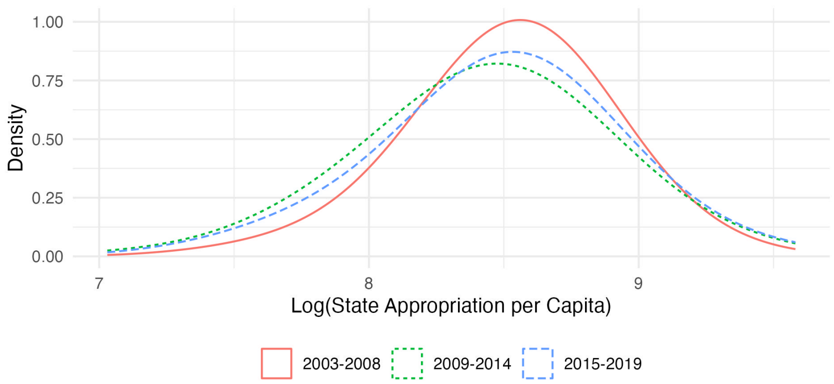
NOTES.—The unit of analysis is institution-year. All amounts are adjusted for inflation using the Consumer Price Index (CPI) and are presented in 2019 dollars. Table A.1 shows summary statistics for all variables, including the state and county covariates.

[8]To align education with industry needs, the Department of Homeland Security (DHS) designates certain degree programs as fields of study in science, technology, engineering, and mathematics (STEM) for the 24-month Optional Practical Training extension available to foreign student graduates seeking employment in the United States (U.S. Department of Homeland Security, 2023). Appendix Table A.25 shows the list of DHS STEM-designated programs.

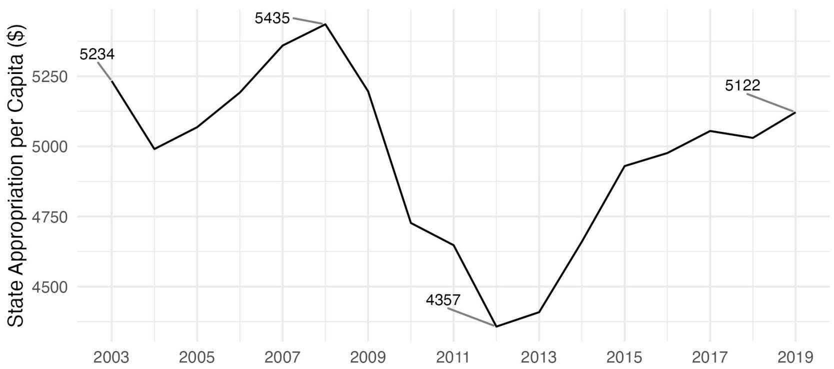
Figure 1 illustrates the temporal changes in real state appropriations per capita. Subfigure (a) demonstrates that the distribution, as time progresses, exhibits a more pronounced left tail, indicating an increased probability of lower funding allocations over time. Subfigure (b) reveals that pre-Great Recession state funding per student in 2008 was $5,435. However, post-2008, state funding continued to decline, reaching its nadir of $4,357 in 2012. Although state appropriations increased after 2012, they did not return to pre-Great Recession levels. These raw appropriation amounts do not explicitly demonstrate the covariation between state funding and college spending or prices over time.

Figure 1: Trends in State Appropriations

(a) Distribution by Periods



(b) Average over Time

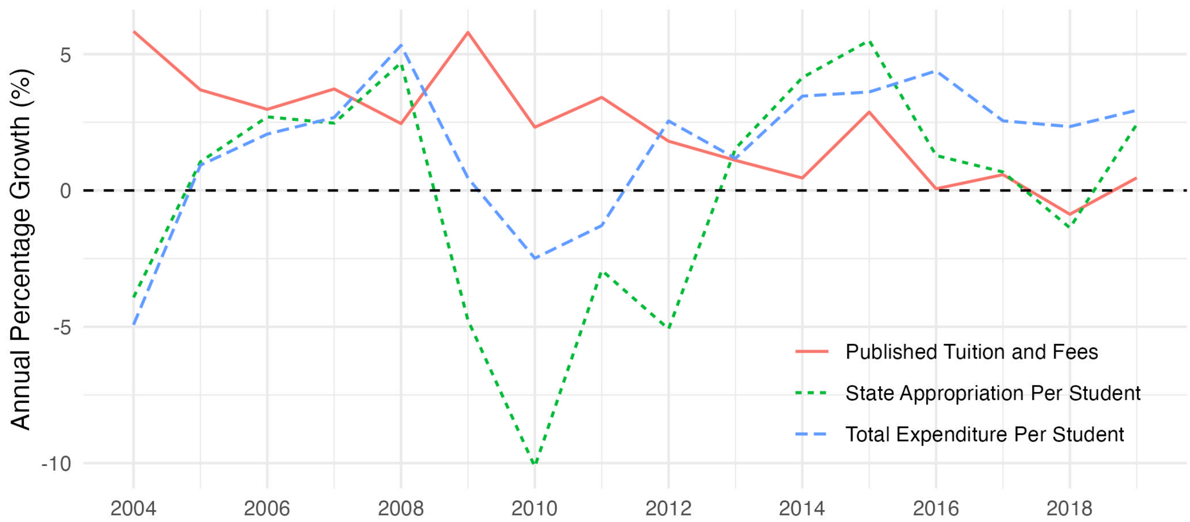


SOURCES.— State Appropriation is extracted from the Grapevine report as documented in State Higher Education Executive Officers Association (2023). The population data for individuals aged 18 to 21 years within the state is acquired from U.S. Census Bureau (2024). All monetary values are adjusted to 2019 dollar terms utilizing the Consumer Price Index (CPI) data provided by the Bureau of Labor Statistics (BLS).

NOTES.— Density plot (a) presents the distribution of log-transformed state appropriations per college-age population. The data spans equally spaced intervals from 2003 to 2019. A kernel density estimate (KDE) is utilized to offer a smooth approximation of the probability density function (PDF), providing insight into the underlying distribution. Figure (b) chronicles the evolution of average state expenditure on higher education per capita, focusing on the college-age demographic (ages 18 to 21), and tracks its progression over the specified time frame. The average is weighted by the state college-age population.

Figure 2 compares the percentage annual changes in college spending and tuition with state appropriations over time. Prior to the 2008 Great Recession, average growth in college spending aligned closely with state funding, while tuition decreased. Post-2008, average per-student state appropriations declined sharply, reaching -10% in 2010. In response to this reduction in state funding, per-student total spending decreased to approximately -2.5% in the same year. Concurrently, tuition growth increased to slightly above 5% in 2009. After 2010, spending and state funding exhibited a positive correlation with fluctuating growth, although state appropriations growth post-2015 remained below pre-Great Recession levels. From 2012 onward, tuition growth remained relatively stable near zero, largely due to the implementation of tuition caps and freezes by 24 states (Deming & Walters, 2018; Miller & Park, 2022). In sum, college spending demonstrates a positive correlation with the level of state support received, suggesting that state funding may influence degree completion through the spending channel.

Figure 2: University Price, Appropriation and Expenditure Annual Changes

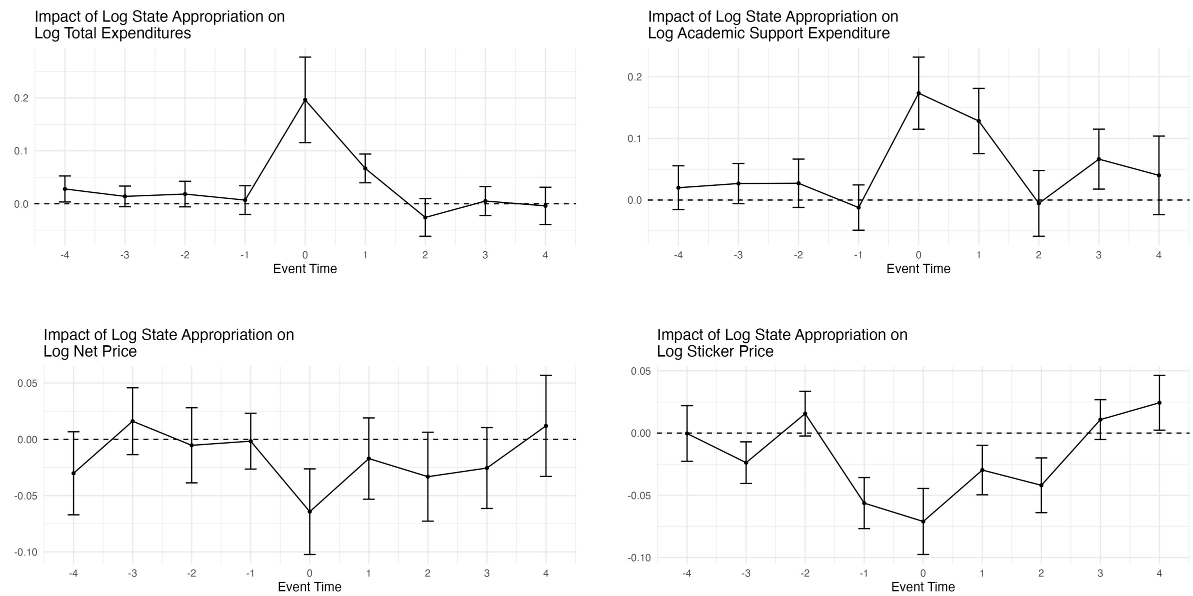


SOURCES.— Data on Published Tuition and Fees, as well as State Appropriation and Expenditure, are sourced from the Integrated Postsecondary Education Data System (IPEDS) as reported in U.S. Department of Education, National Center for Education Statistics (2024). Specifically, Published Tuition and Fees are derived from the Institutional Characteristics Survey (Student charges for academic year programs), while State Appropriation and Expenditure information is obtained from the Finance Survey. The per-student basis for state appropriations and expenditures is calculated using data from the 12-Month Enrollment Survey of IPEDS.
All monetary values are adjusted to 2019 dollar terms utilizing the Consumer Price Index (CPI) data provided by the Bureau of Labor Statistics (BLS)

NOTES.— The Figure illustrates the annual percentage change in the average sticker tuition and fees, state appropriations, and total expenditures per student at all public 4-year institutions.

The relationship between state funding and college spending and pricing decisions is illustrated in Figure 3. By regressing spending and pricing variables on leads and lags of state appropriations, while controlling for potential confounding factors, one can observe the dependence of public colleges on state funding. The elasticities of total expenditures and academic support with respect to state appropriations are substantially larger in magnitude than those of sticker or net price. This finding suggests that state appropriations could influence student completion rates, particularly in majors that are most dependent on college resources (e.g., STEM fields).

Figure 3: State Appropriations Vs. Prices and Spending



SOURCES.— All the data are sourced from the Integrated Postsecondary Education Data System (IPEDS) as reported in U.S. Department of Education, National Center for Education Statistics (2024). Specifically, Sticker Price which refers to published tuition and fees is derived from the Institutional Characteristics Survey (Student charges for academic year programs), Net Price from the Student Financial Aid and Net Price, and other variables from the Finance Survey.

NOTES.— The figure presents estimates and 95 percent confidence intervals from regressions of different expenditures or prices on lags and leads of state appropriations for all public 4-year institutions. Models include state level controls (i.e., log population, minimum wage, log personal income, log number of SNAP recipients, democratic governor dummy), county level controls (i.e., log population, share of population by race, log personal income, unemployment rate, birth and death rate) and institution and year fixed effects. Standard errors are clustered at the institution level.

1.3.3 Degree Completion

Table A.2 also presents summary statistics for the number of degrees conferred by major typeâĂŤSTEM or non-STEM. There are significantly more non-STEM majors than STEM majors. Additionally, STEM majors have a smaller average number of degrees conferred. The literature shows that expected earnings, comparative advantage in a major, and subjective tastes are the main reasons for STEM major choice (Arcidiacono, 2004; Arcidiacono, Hotz, & Kang, 2012; Berger, 1988; Montmarquette, Cannings, & Mahseredjian, 2002; Wiswall & Zafar, 2015).[9]

The literature also identifies various drivers of STEM major choice, including preferences and non-pecuniary aspects, academic preparation, high school curriculum, labor market discrimination, and peer influence and support (Altonji, Blom, & Meghir, 2012; Berger, 1988; Card & Payne, 2021; Ketenci, Leroux, & Renken, 2020). This paper contributes to this body of research by examining how state funding influences the completion of STEM majors. Specifically, I investigate whether state appropriations have a greater effect on STEM majors compared to non-STEM majors.

Table 1.1 shows significant trends in higher education from 2003-2007 to 2015-2019. STEM degree production increases more sharply (67.1%) than non-STEM (11.5%), with STEM programs also expanding significantly by 13.6%. This suggests a growing emphasis on STEM education. Larger standard deviations in the later period, especially for degree counts, indicate increasing variability across institutions. This may reflect growing disparities due to a $15.3 million reduction in state appropriations, which non-selective universities might struggle to manage, potentially widening gaps in the higher education landscape.[10]

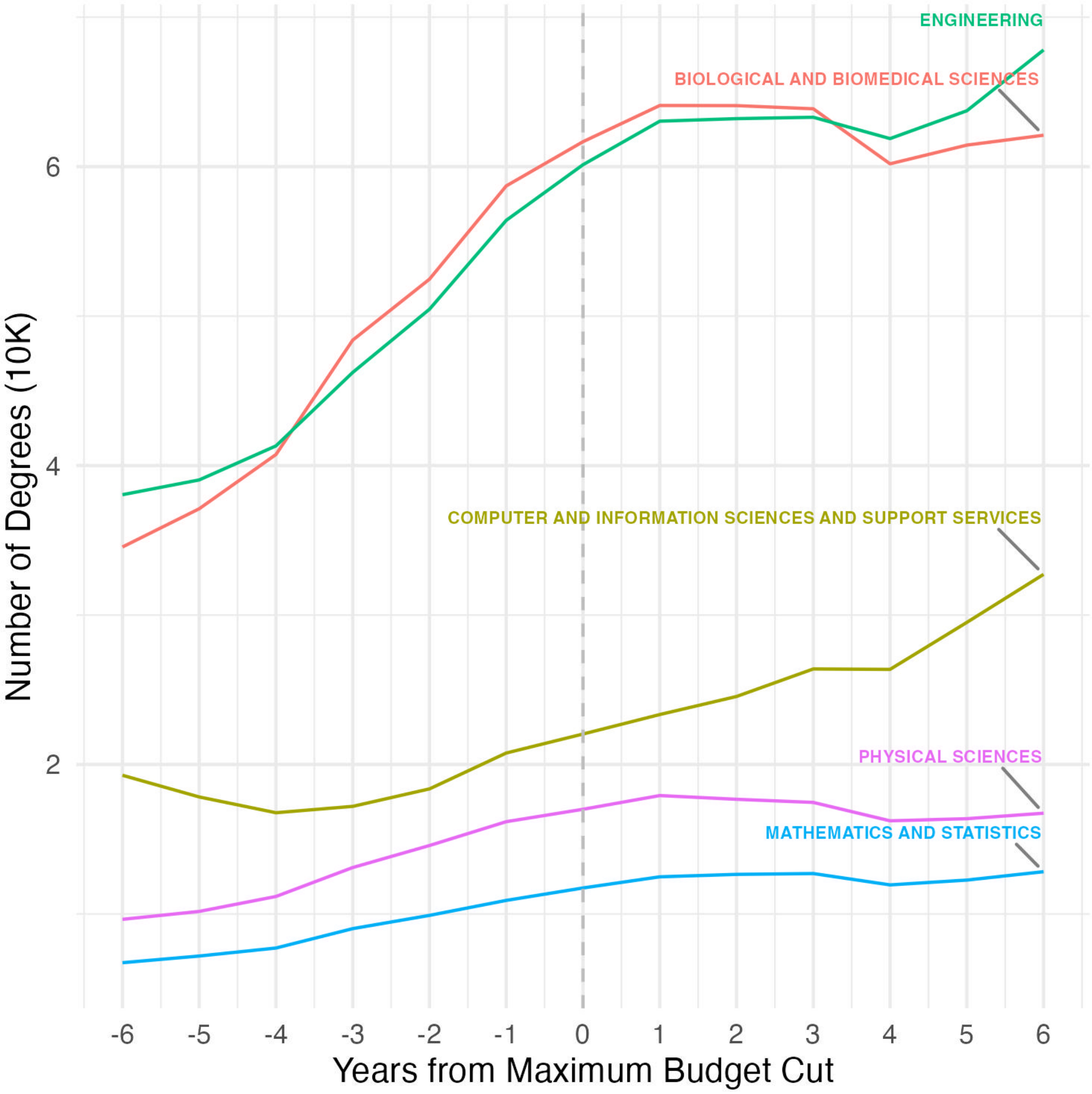
Figure 4 shows the relationship between the total number of different STEM degrees and the event of maximum reduction in state funding. To illustrate the relationship between state appropriations and STEM degree completion, I sum the number of STEM degrees conferred by all public 4-year institutions in the state before and after a maximum decrease in state funding for each institution and STEM major. Most institutions received the largest reduction in state funding between 2009 and 2012. With the exception of computer and information-related STEM majors, a reduction in state funding halts the growth of STEM degrees, with biological and biomedical fields experiencing the most significant drop. Figure A.1 shows the results for non-STEM majors. Business and management-related non-STEM majors appear to be the most responsive to state budget cuts. This simple correlation, however, does not take into account the

[9]Table A.1 provides a comprehensive list of variables, including county and state covariates, for the full sample across all majors. Appendix Table A.3 provides similar descriptive statistics disaggregated by gender and race. Appendix Table A.4 presents a breakdown of the number of degrees by major type (STEM or non-STEM) and examines the differences within each major type for two periods: the first and last five years of the dataset.

[10]Table A.2 presents the summary statistics for the number of degrees by gender and race at university-program-year level.

confounding factors that affect state funding and degree completions. The next section discusses the empirical strategy that addresses this endogeneity problem.

Figure 4: Number of STEM Bachelor Degrees vs. Budget Shock



SOURCES.— Data on Number of Degrees (Bachelor) and State Appropriation are sourced from the Integrated Postsecondary Education Data System (IPEDS) as reported in U.S. Department of Education, National Center for Education Statistics (2024). In particular, Number of Degrees is derived from the Completions–Awards/degrees conferred by program (6-digit CIP code), award level, race/ethnicity, and gender, while State Appropriation information is obtained from the Finance Survey. STEM classification is obtained from The U.S. Department of Homeland Security (2020).
All monetary values are adjusted to 2019 dollar terms utilizing the Consumer Price Index (CPI) data provided by the Bureau of Labor Statistics (BLS)

NOTES.— The unit of analysis is college-program type and year. The Figures displays yearly number of STEM bachelor's degrees awarded by all 4-year public institutions for four years before and after a maximum cut in state appropriations for each university in the period 2003 to 2019.

1.4 Empirical strategy

I use a panel data on public 4-year universities and regress institution-level numbers of degrees conferred for specific majors (i.e., at the 6-digit CIP level) on state appropriations, and state and county economic and demographic conditions. Observations on university (i) located in state s and county c, program j, and the year t as follows:

$$y_{ijsct} = \beta_0 + \beta_1 \text{Appr}_{ist} + X_{st}\phi + W_{ct}\rho + \gamma_i + \alpha_j + \delta_t + \epsilon_{ijsct} \qquad (2)$$

where y refers to the dependent variable of interest, Appr_{ist} is the institutional-level appropriations, X is state-level time-varying controls, W is county-level time-varying controls, and γ, α, and δ to institution-, program type-, and year-specific fixed effects. Hence, the model controls for time-invariant characteristics of institutions and program/major types (6-digit CIP) to account for labor demand for each field, as well as for macroeconomic and other temporal shocks affecting all institutions. I use the following log-log model so that coefficients can be directly interpreted as elasticities (Ω_{ijsct} refers to all controls and fixed effects):

$$\log(y_{ijsct}) = \beta_0 + \beta_1 \log(\text{Appr}_{ist}) + \Omega_{ijsct} + \epsilon_{ijsct} \qquad (3)$$

However, this model suffers from serious endogeneity problems due to either omitted variable bias or reverse causality. Omitted variable bias may result in biased estimates if confounders correlate with both state funding and degree completion. For instance, colleges may adjust their curricula or other unobserved aspects that attract both more state funding and higher degree completion rates. Additionally, higher degree completion could be part of the formula through which states allocate funding, leading to reverse causality. To address this challenge, I use the following Bartik-like instrumental variable $\log(Z_{ist})$ (Bartik, 1991):

$$Z_{ist} = \left(\frac{\sum_{\tau=87}^{91} \text{Appr}_{is\tau}}{\sum_{\tau=87}^{91} \text{Rev}_{is\tau}} \right) \times \left(\frac{\text{Appr}_{st}}{\text{Pop}_{st}} \right) \qquad (4)$$

Where $\text{Appr}_{\text{is}\tau}$ and Appr_{st} refer respectively to the institution and state-level appropriation,

$\mathrm{Rev}_{\mathrm{is}}$ to the institution's total revenue, and $\mathrm{Pop}_{\mathrm{st}}$ to the state college-aged population (i.e., population aged 18 to 24). The first term is the historical share, measuring how reliant each institution has been on state appropriations. The second term is the aggregate state-level appropriations per the state college-aged population. The share is useful for purging variation resulting from endogenous college factors driven by state funding. For example, the historical shares or reliance on state appropriations ensure the use of variation that is net of any state funding adjustment due to evolving college unobserved factors such as the ability to acquire income from other sources (e.g., increased donations or selectivity-driven out-of-state enrollment). The shift (per college-aged population state-level appropriations) is considered exogenous because it affects all state institutions uniformly, and because states determine their higher education budget in advance (Deming & Walters, 2018; Hinrichs, 2022; Parmley, Bell, L'Orange, & Lingenfelter, 2009).[11]

While related literature, notably Chakrabarti et al. (2020); Deming and Walters (2018); Hinrichs (2022), has employed a similar instrumental variable, my approach differs in two aspects. First, it uses historical shares that go back at least 12 years from the first year of the panel data (the first year of data is 2003, and the last year in the shift is 1991). Going back in time over a decade strengthens the exogeneity assumption of the shares. Second, I use a more reliable measure of the share by taking the average over five yearsâĂŤfrom 1987 to 1991. This avoids any cyclicality in state funding allocation to specific institutions over time and hence reduces any potential measurement error.[12]

A new strand of econometric literature has emerged recently, calling into question different parts of the identification assumptions of Bartik instruments. Whereas Goldsmith-Pinkham et al. (2020) argues that the identification comes from the historical shares, Borusyak, Hull, and Jaravel (2022) contends that the exogeneity comes from the aggregate shifts. In my context, the identification is most likely derived from the shares for at least two reasons.

First, most state legislatures consider the higher education budget a so-called "balance wheel",

[11]This instrument is widely used in the immigration literature (Autor, Dorn, & Hanson, 2013; Bartik, 1991; Blanchard, Katz, Hall, & Eichengreen, 1992; Card, 2001). The findings are robust to the instrument scaling; using Z instead of $\log(Z)$ produces similar results.

[12]For example, an institution could receive more or less state appropriations in a particular year in a cyclical fashion, so using only a single historical cross-section may not capture the true reliance on state funding.

implying the allocation of what remains after addressing other priorities such as Medicaid or K-12 education (Bell, 2008; Bound et al., 2019). This suggests that aggregate appropriations at the state level are likely uncorrelated with college-specific factors driving degree completion at the university level, especially given that most states have balanced budget requirements. In 1977, 33 states required balanced budgets, but by 2015, this number had grown to 46, with 37 states adopting constitutional balanced budget requirements (Rueben & Randall, 2017). By 2020, Vermont remained the only state without any balanced budget requirement (Tax Policy Center, 2020). Second, the dependence on state funding, especially for small colleges, may not change drastically over time. Hence, I implement Goldsmith-Pinkham et al. (2020) in my robustness checks.[13]

1.5 Results

1.5.1 The Effect of State Appropriations on Degree Completion, by STEM Type

Table 1.2 presents the baseline results of the impact of state appropriations on degree completions, comparing STEM fields to non-STEM fields. A comparison of Panels (a) and (b) reveals a pronounced effect on STEM degrees, particularly at $t + 4$, which aligns with the expected four-year graduation timeline. This temporal juncture marks a statistically significant influence solely for STEM degrees. The elasticity of 0.341 indicates that a 10% increase in state appropriations leads to a 3.41% increase in the number of STEM degrees conferred in each STEM major on average. Based on mean values, this implies that approximately $10 million in additional state funding corresponds to an additional degree in each STEM major for students who complete their studies within the standard four-year period.

Advancing to $t + 5$, the period indicative of a five-year graduation span, the STEM coefficient remains statistically significant at the 5% level, while the non-STEM one becomes weakly significant at the 10% level. The addition of county control variables reduces the magnitude of

[13]Given the presence of competing approaches, T. J. Wright (2022) recommends that researchers carefully select the most appropriate identification assumption based on the specific context of their study.

the non-STEM coefficient while keeping the STEM one intact. Hence, the non-STEM estimate is, at best, an upper bound. The magnitude of the STEM coefficient is approximately 1.6 times larger. The final coefficient for the cohort of students who graduate in six years $t + 6$ remains significant only for STEM degrees.

Table 1.2: The Effect of State Appropriations on Bachelor's Degrees, by STEM Type

	Dependent Variable: Log Number of Bachelor Degrees							
	t+3		t+4		t+5		t+6	
Panel (a): STEM								
Log State Appropriations	0.181	0.172	0.326**	0.341**	0.297**	0.297**	0.251**	0.235**
	(0.172)	(0.176)	(0.160)	(0.169)	(0.130)	(0.130)	(0.111)	(0.108)
N Institutions	617	617	617	617	617	617	617	617
N Programs	339	339	339	339	339	339	339	339
Observations	94,249	92,804	85,574	84,262	77,206	76,019	69,228	68,157
F-test (1st stage)	110.51	99.742	129.83	114.24	205.53	182.92	193.32	171.02
Panel (b): Non-STEM								
Log State Appropriations	0.183	0.171	0.223	0.206	0.216*	0.188*	0.182*	0.142
	(0.198)	(0.219)	(0.158)	(0.170)	(0.113)	(0.111)	(0.098)	(0.091)
N Institutions	663	663	663	663	663	663	663	663
N Programs	783	783	783	783	783	783	783	783
Observations	263,276	259,969	238,383	235,340	214,454	211,681	191,533	189,026
F-test (1st stage)	298.97	260.50	405.29	363.31	678.11	632.93	608.29	583.50
Institution FE	✓	✓	✓	✓	✓	✓	✓	✓
Program FE (6-Digit CIP)	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓	✓	✓
County Controls	×	✓	×	✓	×	✓	×	✓

NOTES.—Each panel and cell refers to a separate 2SLS regression. The outcome variable in panel (a) is the logged number of undergraduate Science, Technology, Engineering, and Mathematics (STEM) degrees (bachelor's) conferred. In panel (b), the dependent variable is the logged number of non-STEM bachelor's degrees. A lead of "t+i" (ranging from 3 to 6) indicates the ith lead of the outcome variable. For example, the lead t+6 refers to the student cohort that completes bachelor's education in 6 years. CIP refers to the Classification of Instructional Programs, which is used to categorize degree programs by specific majors (see Table A.25 for the subset of STEM majors). The unit of analysis is at the institution, program (6-digit CIP), and year level. The models include year, program, and institution fixed effects, as well as time-varying county and state covariates. The state-level controls include poverty rate, logged population aged 18 to 24, per-capita personal income, minimum wage, number of Supplemental Nutrition Assistance Program (SNAP) recipients per 100,000 population, and number of Temporary Assistance for Needy Families (TANF) recipients per 100,000 population. The county-level controls include logged per-capita personal income, total population, unemployment rate, percentage of population by race (Hispanic, Black, and White), percentage of male population, and birth and death rates per 100,000 population. Standard errors are clustered at the institution level. *p < 0.10, **p < 0.05, ***p < 0.01

The first-stage F-statistics provide strong evidence for the relevance of the instrumental variable in both the STEM and non-STEM models. For STEM degrees, the F-statistics range from 99.742 to 205.53 across different lead specifications, indicating a robust correlation between the instrument and the endogenous regressor. The non-STEM models exhibit even larger F-statistics, ranging from 260.50 to 678.11. This difference in magnitude is likely attributable, at least in part, to the substantially larger sample size for non-STEM programs, which affords greater precision in the first-stage estimates. These robust F-statistics suggest that the instrument is highly correlated with the endogenous regressor, mitigating concerns about weak instrument bias.[14]

Building on my previous specification with year, institution, and program fixed effects, I explore the robustness of my findings by introducing more granular fixed effects in Table A.6. I estimate models with program-by-year and institution-by-program fixed effects to control for time-varying factors common across programs and institutions, as well as for characteristics unique to each program within an institution. For example, this approach accounts for changes in national accreditation standards for engineering programs and the reputation of a specific university's biology department. The results remain quantitatively and qualitatively consistent across these more demanding specifications. The instrumented state appropriations continue to show a statistically significant effect only on STEM degrees, with no significant impact on non-STEM programs.

Table A.7 shows similar 2SLS results for master's degrees, indicating that state appropriations have a positive effect exclusively on STEM-related master's degrees. The doctoral degrees, as shown in Table A.8, indicate a null effect of state appropriations on either major type. This will be discussed further in the heterogeneous effect section, which shows that the state appropriation effect is concentrated within non-selective universities, implying that selective institutions offering higher degrees like doctorates are not affected.

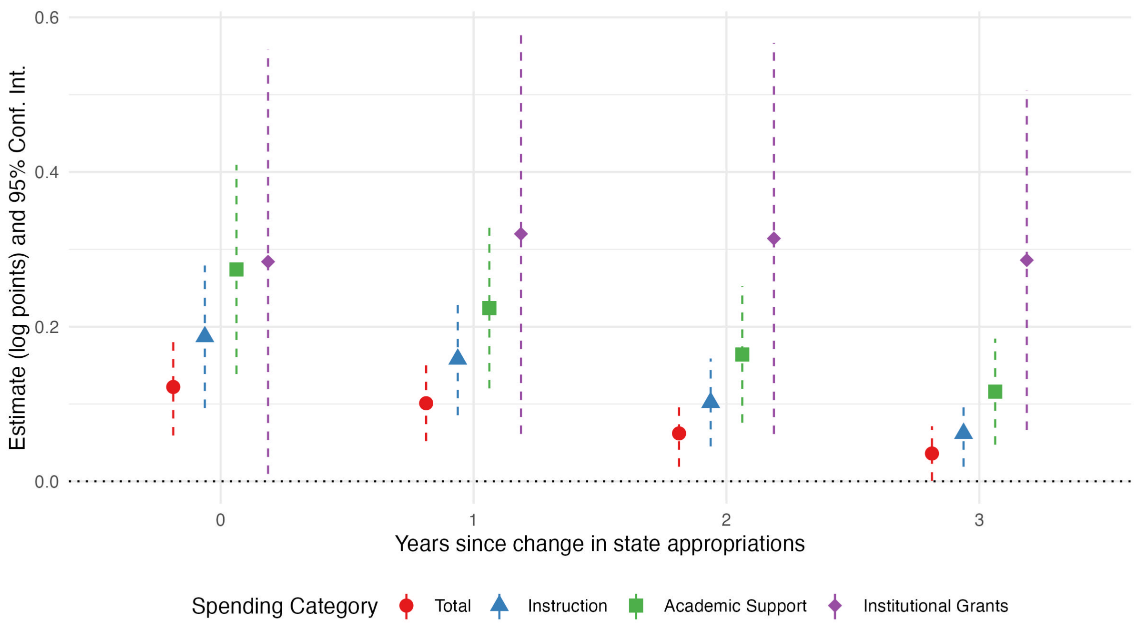
While the extant literature (e.g., Bound et al. (2019); Deming and Walters (2018)) has established that state appropriations influence college aggregate enrollment and completion, this study delves into the nuances of college completion by examining the differential impacts on STEM and non-STEM degrees. The findings reveal that STEM degrees exhibit a higher sensitivity to state funding.

The subsequent section explores the underlying mechanisms driving the disproportionate effect of state appropriations on STEM degrees. Specifically, I examine two potential mechanisms: college spending (with a focus on specific areas such as academic support) and the number of programs offered (the choice set of majors within each field).

[14]Table A.5 presents the OLS estimates. Adding controls increases STEM coefficients and decreases non-STEM coefficients, suggesting potential downward bias for STEM and upward bias for non-STEM in simpler models. This indicates that parsimonious specifications may underestimate STEM effects and overestimate non-STEM effects on the outcome variable.

1.5.2 The Effect of State Appropriations on Spending and Number of Programs offered

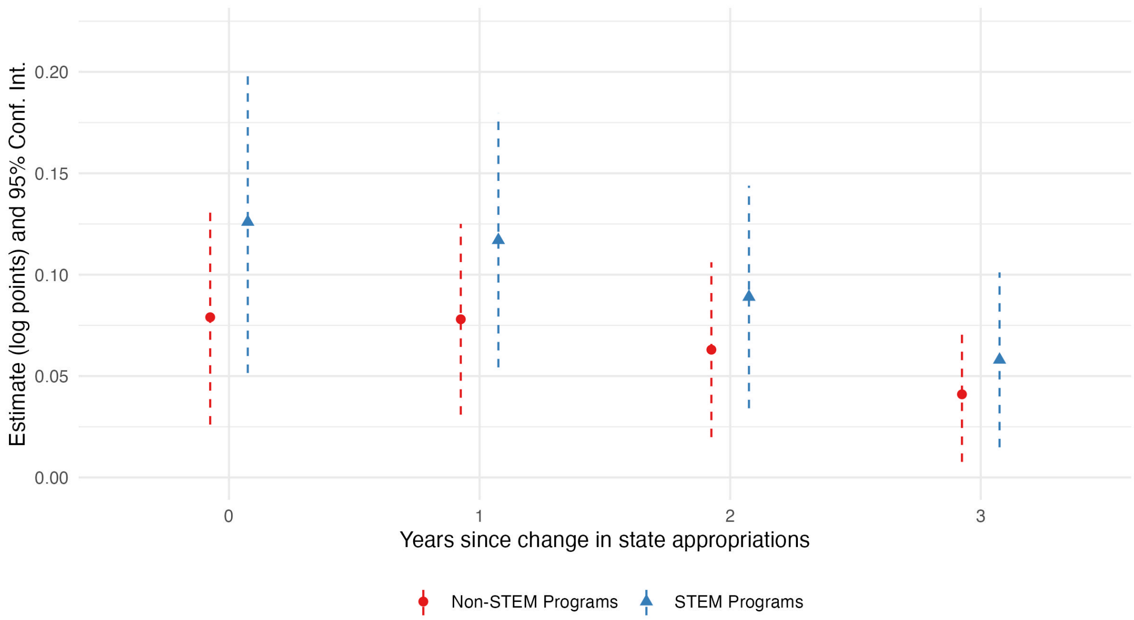
Figure 5: Effect of State Appropriations on Spending



NOTES: Each estimate refers to a separate 2SLS regression. Academic support spending includes expenditures on tutoring, advising, mentoring, and other activities and services that support the institution's primary missions. The models include year and institution fixed effects, as well as time-varying county and state covariates, as described in Table 1.2. A lead of "t+i" (years since change in state appropriations, from 0 to 3) indicates the i^{th} lead of the outcome variable. The unit of analysis is at the institution-year level. Standard errors are clustered at the institution level. Further estimation details are provided in Table A.9.

State support influences public universities' financial decisions, affecting the amount universities invest in their core operations, such as instruction, academic support, and institutional grants. Figure 5 (see Table A.9 for further estimation details) indicates that state support leads to higher aggregate and specific expenditures. Interestingly, the elasticities for academic support related expenditures (e.g., tutoring, advising, and mentoring) surpass those for instructional expenditures (e.g., faculty salaries and wages). Hence, state support increases spending on tutoring and other supportive academic activities. As STEM courses tend to be more time-intensive to study for, increased tutor spending at the institutional level may benefit STEM majors more than non-STEM majors. The last panel (d) shows that institutions also invest more in grants (fellowships and aid). Consequently, state support is directly linked to the investment levels in core operations at public institutions, with elasticities ranging from 0.036 to 0.31.[15]

Figure 6: Effect of State Appropriations on Number of Programs



NOTES: Each estimate refers to a separate 2SLS regression. "N Distinct STEM/Non-STEM Programs" refers to the institution-level number of distinct programs offered within Science, Technology, Engineering, and Mathematics (STEM) or non-STEM fields. The models include year and institution fixed effects, as well as time-varying county and state covariates, as described in Table 1.2. A lead of "t+i" (years since change in state appropriations, from 0 to 3) indicates the i^{th} lead of the outcome variable. The unit of analysis is at the institution-year level. Standard errors are clustered at the institution level. Further estimation details are provided in Table A.10.

State appropriations can also impact the ability of colleges to offer new programs to meet market demand. To study this effect, I investigate how state funding affects the number of distinct STEM and non-STEM degree programs offered by each college over the data period. Figure 6 (see Table A.10 for more estimation details) reports the results. State appropriations impact the number of distinct programs for both STEM and non-STEM majors. However, the elasticity of state appropriations for STEM degrees is approximately 1.6 times that of non-STEM degrees at $t + 0$. This larger effect on STEM programs is consistent with the rapid market developments in STEM fields, such as advancements in artificial intelligence, data science, and programming languages.[16]

The results in all tables are robust to the inclusion of more granular demographic and

[15]Spending measures are only available at the institution level from IPEDS surveys, not disaggregated by department or program. The analysis thus uses college-year level data instead of college-program-year level data

[16]Appendix Table A.12 shows higher state appropriations lead to lower college prices. A 10% funding increase reduces sticker prices by 1.04% and net prices by 1.8%. Effects persist but decrease over time. While state funding may influence degree numbers through pricing, small elasticities suggest limited impact.

economic covariates at the universities' location counties. The first-stage F-statistic exceeds 20 for the concurrent effect (t+0), indicating relatively strong instrument relevance.

Although state funding has varying impacts on different factors (spending and program offerings), it remains unclear whether these factors significantly influence bachelor's degree completion. Hence, it is important to establish whether these factors have differing influences on STEM and non-STEM bachelor's degree completion.

1.5.3 The Effect of Prices, Spending and Number of Programs offered on Completion

This mechanism section involves determining the effect of college financial variables (expenditures), number of distinct STEM or non-STEM majors offered, and prices (tuition and net price) on degree completion. The use of the previously employed Bartik instrumental variable is not feasible in this case because spending historical shares remain nearly constant over time. For example, flagship universities in each state consistently maintain the highest share of aggregated expenditures, as their relative size remains intact.

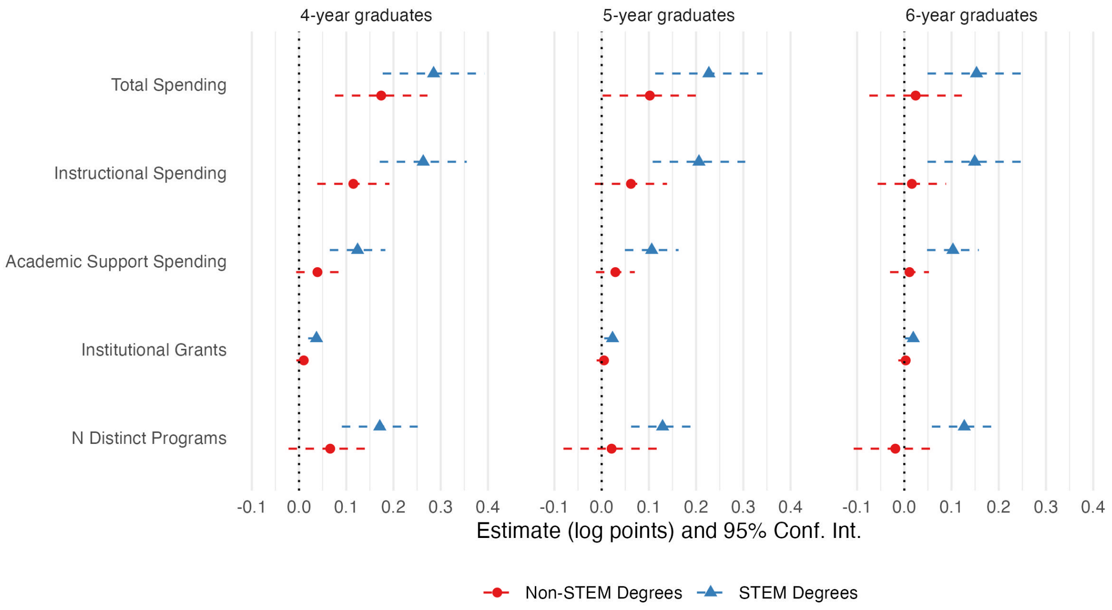
For this purpose, I employ OLS as in Equation (3), with $Appr$ replaced by the expenditure, distinct number of programs, and price variables. I supplement these OLS results with an assessment of omitted variable bias using the approaches of Diegert et al. (2022) and Oster (2019). Specifically, I present breakdown points, which are the thresholds at which the estimated results could be overturned due to omitted variable bias. A higher breakdown point indicates that the results are less likely to be significantly influenced by unobserved factors.[17]

Figure 7 (see Table A.11 for further estimation details) indicates that most of these factors are significant, and their impact is more pronounced for STEM degrees. Panel (a) shows that the number of distinct programs offered is associated with more bachelor's degrees only for STEM majors. Combined with the previous findings, state appropriations have a greater impact on STEM degree completion by increasing the number of distinct STEM majors, which in turn attracts new students. The insignificance of the coefficient for non-STEM implies that increases

[17]While these methods do not necessarily provide causal estimation, they are highly effective in quantifying the risk of omitted variable bias. The Diegert et al. (2022) approach is more robust as it relaxes the assumption that omitted variables must be uncorrelated with the included controls.

in non-STEM majors due to state appropriations lead only to student shuffling within these majors.[18]

Figure 7: Effect of Spending and Programs on Number of STEM and Non-STEM Degrees



NOTES: Each estimate refers to a separate OLS regression. The STEM degree outcome variable (triangle estimates) is the logged number of undergraduate Science, Technology, Engineering, and Mathematics (STEM) bachelor's degrees conferred. For the non-STEM degrees (point estimates), the dependent variable is the logged number of non-STEM bachelor's degrees. "N Distinct Programs" refers to the institution-level number of distinct programs offered within STEM for the first three columns and within non-STEM for the last three columns. "Academic Support Expenditures" includes spending on tutoring, advising, mentoring, and other activities and services that support the institution's primary missions. The models include year and institution fixed effects, as well as time-varying county and state covariates, as described in Table 1.2. A lead of "t+i" (ranging from 3 to 6) indicates the ith lead of the outcome variable. For example, the lead t+6 refers to the student cohort that completes bachelor's education in 6 years. The unit of analysis is at the institution-year level. Standard errors are clustered at the institution level. Further estimation details are provided in Table A.11.

The table (panel a) also reports the breakdown point ($\hat{r}_X^{bp}(\%)$) following Diegert et al. (2022). This breakdown statistic refers to the threshold of the allowed maximum share of selection on unobservables out of selection on observables such that the coefficient would no longer statistically be different from zero. The breakdown point of 80% indicates that selection on unobservables due to omitted variables must be at least 80% as large as selection on observables to reject the OLS estimates and conclude that the coefficient is zero. A breakdown point close to zero would imply that the OLS results are unlikely to be causal. Since the breakpoints are high, the findings are robust and omitted variable bias is less of a concern.

Figure 7 indicates that total spending is positively associated with the number of both STEM

[18]The number of distinct programs is collinear with the program fixed effects. Therefore, STEM and non-STEM degrees are summed at the institution and year level. Other variables vary only at the institution level.

and non-STEM degrees, but the elasticity for STEM degrees is approximately 64% higher than that for non-STEM degrees (4-year graduates). The figure shows that academic support spending has more than triple the effect on STEM degrees for graduation within 4 years (4-year graduates). Additionally, academic support spending has a statistically significant effect only on STEM degrees for graduation within 5 and 6 years. Similarly, the findings indicate that instructional spending has a greater effect on STEM degrees at 4-year graduatesâĂŤthe elasticity for STEM is more than double that of non-STEM degrees. For institutional grants (fellowships and aid), the effect is significant only for STEM majors, but the elasticities are smaller than those of spending. The breakpoints from Table A.11 support the robustness of the estimates in panels b, c, and e, but the breakpoint in panel d is relatively smaller, implying sensitivity to selection on unobservable factors.

Panels (f) and (g) in Table A.11 show that prices have minimal or no effect on non-STEM major completion. Sticker price has a negative effect on STEM degrees with elasticities of -0.13 and -0.11 for cohorts graduating in 4 and 5 years, respectively. This implies that some STEM-oriented students on the margin may have shifted their major choice or college choice, opting for the private sector. The outside option could also be no enrollment in college or other preferences such as seeking vocational education. Nonetheless, net price has no statistically significant effect on the number of degrees, which implies that the actual cost of college attendance net of aid and grants does not influence the number of degrees conferred. One potential explanation for this is that grant aid from various sources has been increasing since 2006, reaching approximately $8,000 on average at public 4-year universities in 2019, while the net cost of attendance has remained stagnant since 2015 at $20,000 (Ma, 2021).

Whereas findings in Figures 5 and 6 show that state appropriations have a significant positive effect on college attendance cost, spending, and number of offered majors, Figure 7 shows that most of these factors are highly important for STEM completion. Consequently, state appropriations have a disproportionately greater effect on STEM major completion due to the dependence of STEM fields on resources and the ability of institutions to respond to new emerging demand by offering new majors. Figure A.2 shows that the sensitivity analysis breakdown points remain at the same level or improve for different assumptions of correlation

(0 to 1) between observed and unobserved variables.[19]

1.6 Heterogeneous effects

1.6.1 Selectively

Selectivity in higher education refers to the degree of competitiveness and exclusivity in a university's admissions process. Highly selective institutions typically admit a small percentage of applicants, often those with exceptional academic credentials, standardized test scores, and extracurricular achievements. These institutions generally have stronger reputations and greater financial resources through alumni donations, research grants, and endowments. In contrast, less selective or non-selective institutions have more inclusive admissions policies, admitting a larger proportion of applicants with varying academic backgrounds.[20]

Selective universities tend to have a greater number of distinct revenue sources. Bound et al. (2019) show that such institutions are able to compensate for lost state appropriations by increasing private gifts and endowments. This suggests that my main findings in Table 1.2 are primarily driven by non-selective institutions that have limited sources of income to smooth any income shock. Figure 8 (see Table A.14 for further estimation details) confirms this hypothesis, showing that the effect is significant only for non-selective institutions. The effect remains significant only for STEM degrees, as in the main results.

These findings are based on the Carnegie classification of selectivity. Table A.19 shows similar results with an alternative definition of selectivity, instead using membership in the Association of American Universities or Barron's Profile of American Colleges. The coefficients for STEM degrees remain significant at the 5% level and are larger in magnitude.[21]

Land-grant institutions have a broad mission that includes a strong emphasis on STEM fields,

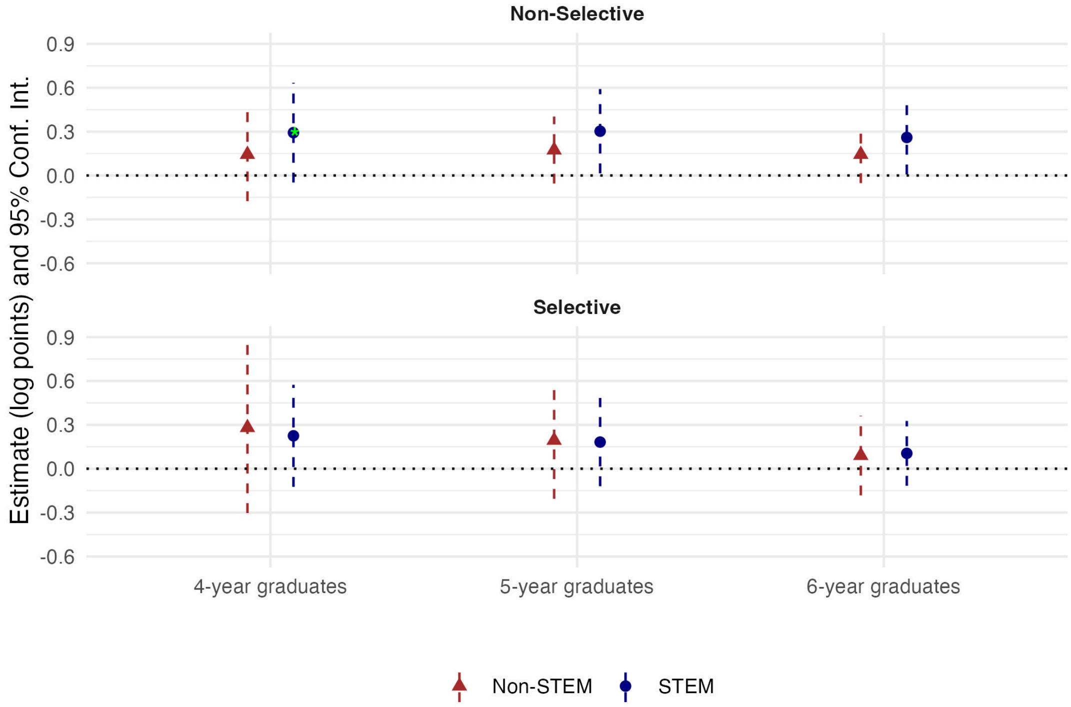
[19]As shown in Table A.11, Sticker price appears to have negative effects only on STEM completion within 4 or 5 years, but net price has no effect on either major's completion. The omitted variable sensitivity test is also robust to alternative estimation methods following Oster (2019), as shown in Table A.13.

[20]There is no consensus in the literature for the definition of selectivity. For example, Deming and Walters (2018) uses Barron's Profile of American Colleges, while Bound et al. (2019) uses membership in the American Association of Universities (AAU).

[21]Similarly, Appendix Table A.20 indicates that the effect is concentrated among non-doctoral institutions, most of which are also non-selective.

especially in areas like agriculture, engineering, and applied sciences, which align with their historical mandate (the Acts of 1862 and 1890). Appendix Table A.21 shows that the state appropriations elasticity of STEM degrees for land-grant institutions ranges between 1.19 and 1.76, suggesting increasing returns to scale at these institutions.

Figure 8: The Effect of State Appropriations on Bachelor's Degrees, by Institution Selectivity and Degree Type



NOTES: Each estimate refers to a separate 2SLS regression. The outcome variable is the logged number of undergraduate degrees (bachelor's) conferred for Science, Technology, Engineering, and Mathematics (STEM) majors or non-STEM majors. The star label refers to significance at 10% level. Selectivity is defined based on the Carnegie classification and refers to the institutions that are in the top 20th percentile of selectivity among all baccalaureate institutions with fewer than 20 percent of entering undergraduate transfers. Table A.19 shows similar results based on the selectivity definition of AAU membership or being listed as competitive in Barron's Profiles of American Colleges. Further estimation details are provided in Table A.14.

1.6.2 Majors

This paper makes a significant contribution by identifying specific disciplines within STEM and non-STEM majors that respond to changes in state funding. The 2SLS regression results presented in Table 1.3 indicate that the impact of state appropriations on the number of bachelor's degrees conferred in STEM fields is not uniform across all disciplines.[22]

[22]The table presents the top five most frequent majors at the 2-digit CIP code level. DHS considers the following majors at the 2-digit CIP code level as STEM: Engineering (CIP code 14), Biological and Biomedical Sciences

The analysis reveals a negligible effect on engineering majors, while a more pronounced influence is observed in the fields of biological and biomedical sciences, mathematics and statistics, as well as physical sciences. It is noteworthy that the 'other' category encompasses a few select majors from predominantly non-STEM disciplines that are recognized as STEM by the Department of Homeland Security (DHS). The insignificance of the coefficient for the engineering fields can be attributed to many factors, including peer effects and student cohort preparedness or quality. Hemelt et al. (2021) show that engineering programs are more expensive than English programs mainly due to higher wages and lower teaching loads, which may then factor into better student outcomes within these fields.

However, the next subsection reveals significance for engineering among some underrepresented gender and racial groups. Overall, the effect is concentrated within non-computer or information sciences and mathematics-related STEM fields.[23]

Table A.15 shows the heterogeneous effects for non-STEM majors; the marginal effect of state appropriations on non-STEM majors is concentrated within English majors and other fields (Table A.22 shows that English majors constitute most of the "other" category in panel f). Changes in teaching loads and class size within these fields, as demonstrated by Hemelt et al. (2021), may explain these results.

(CIP code 26), Mathematics and Statistics (CIP code 27), and Physical Sciences (CIP code 40). See Table A.25 for further details about DHS STEM designated programs.

[23]For example, only four programs in the social sciences are counted as STEM: "Econometrics and Quantitative Economics," "Cartography," "Archeology," and "Geographic Information Science and Cartography."

Table 1.3: The Effect of State Appropriations on STEM Bachelor's Degrees, by Program Type

	Dependent Variable: Log Number of STEM Bachelor Degrees							
	t+3		t+4		t+5		t+6	
Panel (a): Engineering								
Log State Appropriations	0.063	0.080	0.112	0.129	0.171	0.199	0.087	0.062
	(0.155)	(0.182)	(0.144)	(0.178)	(0.134)	(0.168)	(0.097)	(0.117)
N Institutions	326	326	326	326	326	326	326	326
N Programs	55	55	55	55	55	55	55	55
Observations	18,851	18,441	17,217	16,842	15,623	15,282	14,094	13,786
F-test (1st stage)	30.73	23.92	25.85	17.67	33.45	20.95	39.60	23.58
Panel (b): Physical Sciences								
Log State Appropriations	0.263	0.253	0.264*	0.303*	0.130	0.165	0.090	0.108
	(0.190)	(0.193)	(0.155)	(0.170)	(0.097)	(0.105)	(0.089)	(0.093)
N Institutions	485	485	485	485	485	485	485	485
N Programs	35	35	35	35	35	35	35	35
Observations	18,170	17,817	16,723	16,402	15,297	15,003	13,887	13,619
F-test (1st stage)	23.65	22.20	33.86	30.70	61.82	57.47	56.78	53.28
Panel (c): Biological and Biomedical Sciences								
Log State Appropriations	-0.018	0.005	0.452	0.507	0.521**	0.550**	0.426**	0.425**
	(0.382)	(0.392)	(0.337)	(0.353)	(0.258)	(0.264)	(0.211)	(0.207)
N Institutions	526	526	526	526	526	526	526	526
N Programs	66	66	66	66	66	66	66	66
Observations	15,406	15,250	13,988	13,846	12,623	12,494	11,320	11,203
F-test (1st stage)	14.22	12.78	18.35	16.34	30.24	28.40	25.41	24.82
Panel (d): Computer and Information Sciences and Support Services								
Log State Appropriations	-0.403	-0.554	-0.220	-0.287	0.071	0.060	0.303	0.331
	(0.372)	(0.417)	(0.286)	(0.311)	(0.203)	(0.204)	(0.250)	(0.256)
N Institutions	534	534	534	534	534	534	534	534
N Programs	26	26	26	26	26	26	26	26
Observations	9,754	9,590	8,774	8,625	7,841	7,707	6,960	6,839
F-test (1st stage)	10.63	9.227	13.95	12.77	22.16	21.67	18.78	18.11
Panel (e): Mathematics and Statistics								
Log State Appropriations	0.435*	0.407*	0.244	0.231	0.102	0.080	0.095	0.061
	(0.227)	(0.228)	(0.167)	(0.168)	(0.114)	(0.117)	(0.110)	(0.115)
N Institutions	503	503	503	503	503	503	503	503
N Programs	13	13	13	13	13	13	13	13
Observations	8,531	8,387	7,836	7,704	7,148	7,027	6,479	6,368
F-test (1st stage)	14.90	14.24	18.43	17.09	30.48	27.99	29.06	26.46
Panel (f): Others								
Log State Appropriations	0.534	0.499	0.867*	0.845	0.754*	0.660	0.691	0.568
	(0.531)	(0.536)	(0.497)	(0.520)	(0.444)	(0.419)	(0.446)	(0.410)
N Institutions	550	550	550	550	550	550	550	550
N Programs	144	144	144	144	144	144	144	144
Observations	23,537	23,319	21,036	20,843	18,674	18,506	16,488	16,342
F-test (1st stage)	19.28	18.95	22.94	22.61	33.00	33.02	26.99	27.55
Institution FE	✓	✓	✓	✓	✓	✓	✓	✓
Program FE (6-Digit CIP)	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓	✓	✓
County Controls	×	✓	×	✓	×	✓	×	✓

NOTES.—Each panel and cell refers to a separate 2SLS regression and includes a subset of all sub-majors within the given global major 2-digit CIP code. The outcome variable is the logged number of undergraduate degrees (bachelor's) conferred for Science, Technology, Engineering, and Mathematics (STEM) majors. A lead of "t+i" (ranging from 4 to 6) indicates the ith lead of the outcome variable. The models include year, program, and institution fixed effects, as well as time-varying county and state covariates, as described in Table 1.2. Standard errors are clustered at the institution level. *p < 0.10, **p < 0.05, ***p < 0.01

1.6.3 Gender and Race

There is a growing body of literature that attempts to explain the gender gap in STEM majors (Altonji, Arcidiacono, & Maurel, 2016; Kahn & Ginther, 2017). I contribute to this literature by examining the effect of state funding on STEM degrees by gender. Table 1.4 shows that the state appropriation effect for female students is null. This finding supports the growing set of literature that attributes the gap to non-resource-inputs related factors such as social expectations and stereotypes, peer group preferences, professional expectations, and non-STEM-oriented males being less likely to enter university (Brenøe & Zölitz, 2020; Card & Payne, 2021; Cheryan et al., 2015; Crosnoe et al., 2008; Shapiro & Williams, 2012; Speer, 2017). A study by Ahn et al. (2024) showed that curving grades towards a B for all courses can help diminish the gender gap in STEM fields, as female students tend to be more risk-averse towards low grades.

The primary findings, as presented in Table 1.4, indicate that the influence of state appropriations on STEM bachelor's degrees predominantly affects male students. Further analysis, detailed in Tables A.16, A.17, and A.18, reveals that the effect varies by race, gender, and STEM degree types. The findings indicate a significant effect of state appropriations on Black male graduates in engineering-related fields. This implies that production inputs are critical for some underrepresented groups within the engineering discipline. Additionally, the analysis for females shows that state funding has a significant effect only on biological and biomedical science degree completion.

Future research opportunities lie in examining the underlying factors that drive the effects on specific racial groups, notably the null effects of state funding on Hispanic students across all majors, in contrast to other racial groups. One potential explanation for the lack of significance for Hispanic students could be the effectiveness of numerous programs designed to promote underrepresented minorities in STEM fields, particularly in states with large Hispanic populations such as New Mexico, Texas, Florida, and California (Crisp, Nora, & Taggart, 2009; Estrada et al., 2016; Hernandez, Schultz, Estrada, Woodcock, & Chance, 2013). Further investigation into these programs and their interaction with state funding could provide valuable insights into the mechanisms underlying the observed disparities in STEM degree completion across racial groups.[24]

Table 1.4: The Effect of State Appropriations on STEM Bachelor's Degrees, by Gender

	Dependent Variable: Log Number of STEM Bachelor Degrees							
	t+3		t+4		t+5		t+6	
Panel (b): Male								
Log State Appropriations	0.098	0.098	0.287*	0.305*	0.281**	0.296**	0.198*	0.195*
	(0.168)	(0.176)	(0.152)	(0.161)	(0.128)	(0.131)	(0.102)	(0.101)
Panel (c): Female								
Log State Appropriations	0.005	-0.003	0.130	0.157	0.085	0.083	0.091	0.078
	(0.183)	(0.187)	(0.152)	(0.164)	(0.109)	(0.114)	(0.101)	(0.104)
N Institutions	617	617	617	617	617	617	617	617
N Programs	339	339	339	339	339	339	339	339
Observations	94,249	92,804	85,574	84,262	77,206	76,019	69,228	68,157
F-test (1st stage)	110.51	99.74	129.83	114.24	205.53	182.92	193.32	171.02
Institution FE	✓	✓	✓	✓	✓	✓	✓	✓
Program FE (6-Digit CIP)	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓	✓	✓
County Controls	×	✓	×	✓	×	✓	×	✓

NOTES.—Each panel and cell refers to a separate 2SLS regression. The outcome variable in panels (a) refers to the logged number of undergraduate Science, Technology, Engineering, and Mathematics (STEM) degrees (bachelors) conferred to male students, and in panel b to female students. A lead of "t+i" (ranging from 4 to 6) indicates the ith lead of the outcome variable. The models include year, program, and institution fixed effects, as well as time-varying county and state covariates, as described in Table 1.2. Standard errors are clustered at the institution level. *p < 0.10, **p < 0.05, ***p < 0.01

1.7 Robustness checks

To validate the reliability and consistency of the main findings, I conduct a series of robustness checks. This section presents three distinct approaches to assess the stability of the results. First, following Goldsmith-Pinkham et al. (2020), I examine the sensitivity of the instrumental variable (IV) estimates to the exclusion of influential universities. Second, I employ an alternative IV strategy using state-level appropriations, as in Bound et al. (2019, 2020); Webber (2017). Finally, I perform a placebo test using Monte Carlo simulations to further corroborate the robustness of the main results. These checks collectively provide a comprehensive evaluation of the validity and stability of the empirical findings, addressing potential concerns related to instrument construction, endogeneity, and the possibility of spurious correlations.

[24]Some of the prominent programs include the New Mexico Alliance for Minority Participation, the Society of Hispanic Professional Engineers (SHPE), and Hispanic-Serving Institution (HSI) STEM programs.

1.7.1 IV Robustness

Following Goldsmith-Pinkham et al. (2020), I compute the relevance of each university in generating the identifying variation in each instrument (see equation 4). I find that public 4-year institutions from Colorado have the largest power in explaining variation of the instrument. Universities such as Colorado State University, University of Colorado, and other colleges from Colorado have the highest weights in the construction of the instrument (see Table A.23 in the Appendix).

Table 1.5: Impact of State Appropriations on Bachelor's Degrees (Excluding Top 5 Rotemberg Weight Institutions)

	Dependent Variable: Log Number of Bachelor Degrees					
	t+4		t+5		t+6	
Panel (a): STEM						
Log State Appropriations	0.458*	0.483*	0.508**	0.510**	0.395**	0.378*
	(0.245)	(0.268)	(0.233)	(0.253)	(0.192)	(0.205)
N Institutions	617	617	617	617	617	617
N Programs	339	339	339	339	339	339
Observations	84,606	83,294	76,335	75,148	68,444	67,373
F-test (1st stage)	69.426	59.948	74.257	64.125	78.022	63.600
Panel (a): Non-STEM						
Log State Appropriations	0.272	0.248	0.310	0.262	0.267	0.202
	(0.259)	(0.280)	(0.211)	(0.213)	(0.184)	(0.180)
N Institutions	663	663	663	663	663	663
N Programs	783	783	783	783	783	783
Observations	236,321	233,278	212,596	209,823	189,866	187,359
F-test (1st stage)	220.67	194.32	258.75	241.38	241.29	224.88
Institution FE	✓	✓	✓	✓	✓	✓
Program FE (6-Digit CIP)	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓
County Controls	×	✓	×	✓	×	✓

NOTES.—Each panel and cell refers to a separate 2SLS regression. The outcome variable in panel (a) is the logged number of undergraduate Science, Technology, Engineering, and Mathematics (STEM) degrees (bachelor's) conferred. In panel (b), the dependent variable is the logged number of non-STEM bachelor's degrees. A lead of "t+i" (ranging from 3 to 6) indicates the ith lead of the outcome variable. For example, the lead t+6 refers to the student cohort that completes bachelor's education in 6 years. CIP refers to the Classification of Instructional Programs, which is used to categorize degree programs by specific majors (see Table A.25 for the subset of STEM majors). The Table replicates the main findings in Table 1.2 after omitting the top 5 institutions that have the highest Rotemberg weights shown in Table A.23. Table A.24 shows similar results after dropping top 10 weights instead. Standard errors are clustered at the institution level. *p < 0.10, **p < 0.05, ***p < 0.01

Table 1.5 demonstrates that the main estimates are robust to excluding the five universities with the highest Rotemberg weights. This exclusion increases the magnitude of the coefficients, with statistical significance remaining relevant for STEM fields only. Table A.24 shows similar results when dropping the top 10 universities driving the identifying variation in the instrument. These findings indicate that the main results are robust and represent a lower bound estimate.

1.7.2 Alternative IV

Since states often use higher education appropriations as the balancing wheel for state budgets a few years in advance, some literature uses the aggregate state appropriations as an instrumental variable for the specific university-level appropriations. Similar to the instrument implemented by Bound et al. (2019, 2020); Webber (2017), I use the state-level appropriations as an instrument and report the findings in Table 1.6. The coefficients are smaller than the main results, but they remain significant only for the STEM degrees, indicating that my overall story holds.

Table 1.6: The Effect of State Appropriations on STEM Bachelor's Degrees (Alternative IV)

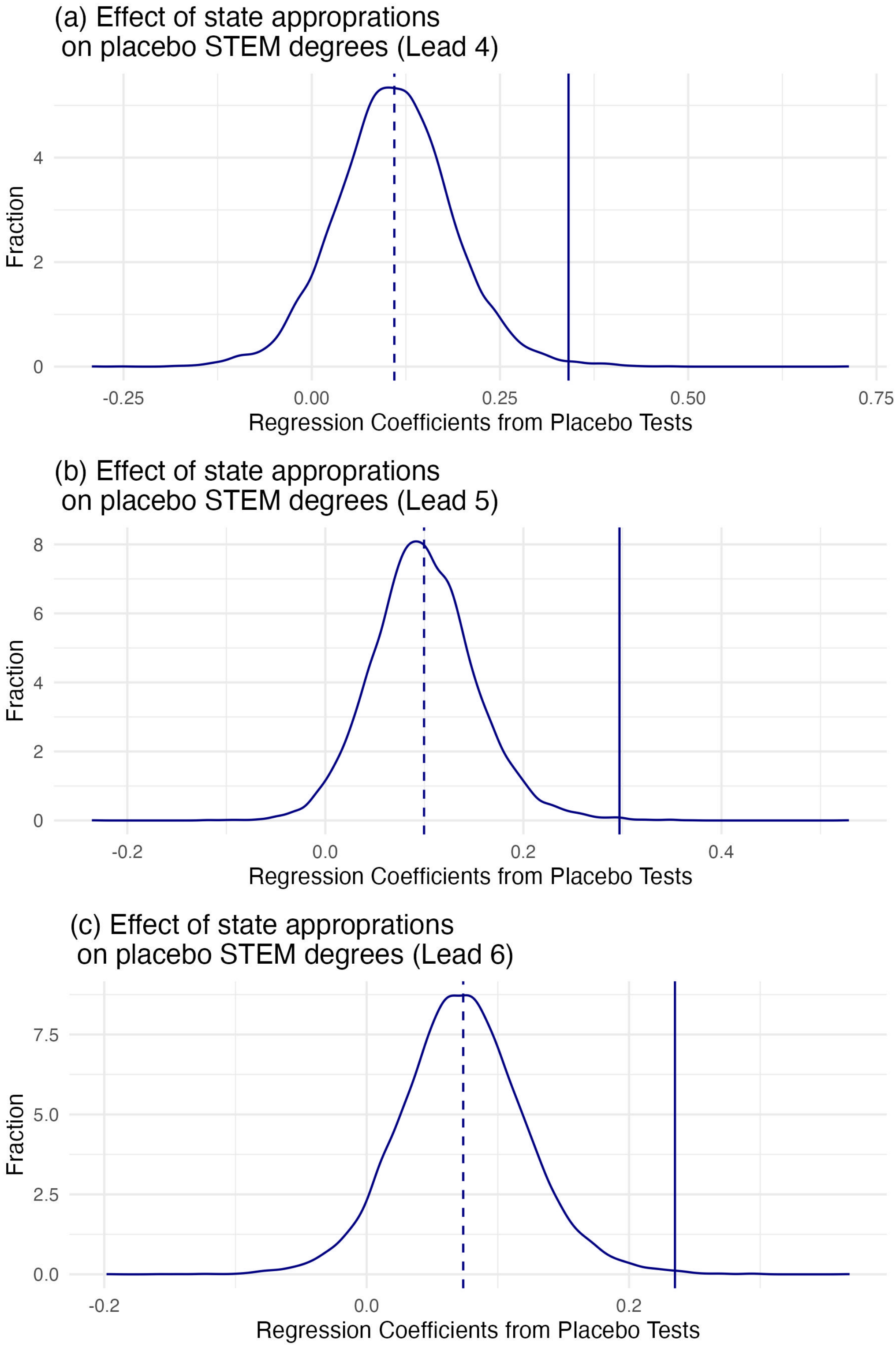
	Dependent Variable: Log Number of STEM Bachelor Degrees							
	t+3		t+4		t+5		t+6	
Panel (a): STEM								
State appropriations	0.088	0.081	0.168*	0.176*	0.150**	0.150**	0.126**	0.118*
	(0.103)	(0.107)	(0.094)	(0.101)	(0.073)	(0.074)	(0.062)	(0.061)
N Institutions	617	617	617	617	617	617	617	617
N Programs	339	339	339	339	339	339	339	339
Observations	94,249	92,804	85,574	84,262	77,206	76,019	69,228	68,157
F-test (1st stage)	80.041	69.530	99.237	84.214	168.12	145.82	160.65	138.64
Panel (b): Non-STEM								
State appropriations	0.100	0.094	0.118	0.109	0.113*	0.098	0.095*	0.073
	(0.123)	(0.140)	(0.094)	(0.102)	(0.065)	(0.063)	(0.055)	(0.051)
N Institutions	663	663	663	663	663	663	663	663
N Programs	783	783	783	783	783	783	783	783
Observations	263,276	259,969	238,383	235,340	214,454	211,681	191,533	189,026
F-test (1st stage)	202.20	166.57	302.13	261.18	550.99	504.18	500.29	472.32
Institution FE	✓	✓	✓	✓	✓	✓	✓	✓
Program FE (6-Digit CIP)	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓	✓	✓
County Controls	×	✓	×	✓	×	✓	×	✓

NOTES.—Each panel and cell refers to a separate 2SLS regression. The table replicates the main results in Table 1.2 using an alternative instrumental variable–the aggregate state-level appropriations following Bound et al. (2019). The outcome variable in panels (a) refers to the logged number of undergraduate Science, Technology, Engineering, and Mathematics (STEM) degrees (bachelors) conferred, and in panel b Non-stem. A lead of "t+i" (ranging from 4 to 6) indicates the ith lead of the outcome variable. The models include year, program, and institution fixed effects, as well as time-varying county and state covariates, as described in Table 1.2. Standard errors are clustered at the institution level. *p < 0.10, **p < 0.05, ***p < 0.01

1.7.3 Placebo test

To test the robustness of the main findings in Table 1.2, I conduct a falsification test involving Monte Carlo analysis. In this analysis, I first limit the data to non-STEM majors and randomly assign false STEM majors such that the share of STEM to non-STEM is preserved as in the original data, where approximately 30% of the majors are STEM. This random sample is drawn 10,000 times to perform the 2SLS estimation for each, as in the main results. Figure 9 shows the kernel density of the generated placebo coefficients on each variable lead. The dashed lines indicate the estimated placebo parameters, while the solid lines denote the actual estimated parameters in Table 1.2. The estimated true coefficients are further away from the center (randomly generated parameters). This indicates that the main findings are not random but robust.

Figure 9: Placebo Test for the Effect of State Appropriations on STEM Degrees



NOTES.—The figure shows the kernel density plot of the coefficient on each outcome variable lead obtained from 10,000 different sets of placebo 2SLS regressions as in Table 1.2. The placebo test involves a Monte Carlo analysis in which STEM majors are randomly assigned across a bootstrap sample (random sample with replacement) drawn from the dataset that contains non-STEM majors only. The number of placebo STEM majors drawn is consistent with the share of STEM majors in the original data, which is 30.2% of all the majors. The dashed line gives the average point estimate of the coefficient on each placebo distribution, and the solid vertical line gives the estimated slope coefficient of each outcome variable as shown in Table 1.2.

STEM refers to Science, Technology, Engineering, and Mathematics programs or majors as designated by the Department of Homeland Security (DHS). Table A.25 presents the complete list of STEM programs.

1.8 Conclusion

This study provides compelling evidence that state appropriations have a significant and disproportionate effect on STEM degree completion in public four-year institutions. The findings reveal that a 10% increase in state funding leads to a 3.4% increase in STEM degrees conferred, primarily manifesting four years after the funding change. This effect is particularly pronounced in biological and biomedical sciences, mathematics, and physical sciences, while being largely concentrated among male students and non-selective institutions.

The analysis of potential mechanisms suggests that increased state support influences STEM degree completion through multiple channels. Higher institutional spending (especially on academic support), and an expanded array of STEM programs all contribute to improved outcomes in STEM fields. These findings underscore the resource-intensive nature of STEM education and the critical role of state funding in supporting these programs (Altonji & Zimmerman, 2019; Hemelt et al., 2021).

The heterogeneous effects observed across institution types, gender, and race highlight important considerations for policymakers. The concentration of effects in non-selective institutions suggests that these schools may be more vulnerable to changes in state funding and may require particular attention in funding decisions. These results have several important implications for higher education policy. They suggest that cuts to state appropriations may have outsized negative effects on STEM education, potentially hampering efforts to increase the STEM workforce (Bottia, Stearns, Mickelson, & Moller, 2018; Griffith, 2010; Sjoquist & Winters, 2015). The findings indicate that increased state funding could be an effective tool for boosting STEM degree completion, particularly at non-selective institutions.

While state support for operating costs has diminished over time, public universities have continued to increase their real assets per student almost linearly (see Figure A.3). This trend suggests a shift in state funding toward long-term projects (e.g., STEM buildings or laboratory equipment), warranting investigation into the effect of college assets on student outcomes. In conclusion, this study underscores the critical role of state appropriations in supporting STEM education at public four-year institutions, providing important evidence to inform policymakers' deliberations on budget decisions and strategies to strengthen the STEM workforce.

Chapter 2

From High School to Higher Education: Is Recreational Marijuana a Consumption Amenity for US College Students?

Recreational marijuana legalization (RML) is a contentious issue with widespread attention from politicians, researchers, and citizens. Twelve states have legalized recreational marijuana as of 2020 (Carnevale Associates, 2022), and there is a growing trend of high marijuana use among college freshmen. Nearly 35% of high school seniors and 44% of college students reported using marijuana (National Institute on Drug Abuse, 2020; Schulenberg et al., 2021). Marijuana legalization's impact has been extensively studied across disciplines. These include health (Raman & Bradford, 2022; Sarvet et al., 2018), crime or substance abuse (Sabia, Dave, Alotaibi, & Rees, 2021; Wu, Wen, & Wilson, 2021), as well as income (Chakraborty, Doremus, & Stith, 2020; Dave, Liang, Muratori, & Sabia, 2023; Jiang & Miller, 2022). Others have shown that students value some college amenities (i.e., student activities and sports) resulting in positive or negative effects on achievements (Jacob, McCall, & Stange, 2018a; Lindo, Swensen, & Waddell, 2012; Webber & Ehrenberg, 2010). Therefore, the impact of RML on undergraduate enrollment remains an unanswered question.

The intuitive connection between RML and enrollment is based on two theoretical frameworks. The first is the human capital model by Becker (1964), which views education as an investment that depends on costs, earnings, and consumption amenities. The second approach involves the neoclassical model of spatial equilibrium, as proposed by Rosen (1974) and Roback (1982). This model implies that there are no utility gains at equilibrium, given that the rental and wage adjustments driven by amenities counterbalance the utility differentials. These theoretical frameworks imply that RML increases enrollment in states that have legalized recreational marijuana (RM states). However, measuring the value of amenities for location choice is challenging because it involves accounting for various confounding and even dynamic factors. Due to this difficulty, a large literature including Chen, Oliva, and Zhang (2022); Mayer and Trevien (2017); Zambiasi and Stillman (2020) relies on natural experiments that exploit an exogenous shock to amenities' value.[1]

Related literature focuses on limited geographical locations and uses student-level data or natural experiments to examine how marijuana use affects educational performance, including standardized grades (Marie & Zölitz, 2017; A. C. Wright & Krieg, 2020), quantity and quality of applicants (Blake, Hess, & Thomas, 2023), time to graduation, (Arria, Caldeira, Bugbee, Vincent, & O'Grady, 2015), school failure (Duarte, Escario, & Molina, 2006), and dropout rate or truancy (Roebuck, French, & Dennis, 2004; Suerken et al., 2016).

I contribute to the above literature by examing how RML affects postsecondary education for first-time enrollment in the US and whether the effect is due to the perception of marijuana as a college consumption amenity. To investigate this question, I utilize the state-level and temporal variation in RML and build a unique college-level dataset.[2] My main empirical strategies are as follows: the standard two-way-fixed effects difference-in-differences (TWFE DID) and the dynamic TWFE DID. These methods compare the mean enrollment differences between RM and non-RM states, controlling for both college-level time-invariant unobserved heterogeneity (college fixed effects) and common time shocks (time fixed effects). I also include time-varying county-level economic variables such as unemployment and per-capita income, as well as various

[1] According to the neoclassical model of spatial equilibrium, people relocate in response to amenities until their utility from amenities equals the disutility from wage and rent differentials. Similarly, the human capital model implies that students take into account the availability of amenities when choosing a college. College sports serve as one such amenity, as shown in Chapter 3 (El Fatmaoui & Ransom, 2025), while recreational marijuana can potentially be perceived as another college amenity.

[2] I construct this panel dataset from Integrated Postsecondary Education Data System (IPEDS) surveys, the Bureau of Economic Analysis, and the Bureau of Labor Statistics

time-varying college-level characteristics, in the regressions. To address potential concerns about the temporal and cohort variation in the treatment effects, I follow the procedures proposed by Sun and Abraham (2021) and Goodman-Bacon (2021) .[3]

Standard errors are clustered at the state level at which the treatment is assigned.[4] The validity of clustering techniques depends on the number of clusters, as overrejection can occur when the number of clusters is small (Cameron, Gelbach, & Miller, 2008). Therefore, to account for the potential bias in the cluster-robust standard errors due to the small number of clusters and few treated units, I also report the p-values or confidence intervals from the wild-cluster bootstrap procedure proposed by MacKinnon and Webb (2018) for the main coefficients of interest.[5]

The estimates based on TWFE DID show that RML leads to approximately a 4.6% to 9% increase in total first-time undergraduate enrollments.[6] Goodman-Bacon (2021) decompositions indicate that the estimates are driven by early policy adopters and mostly driven by the RM and non-RM states part of the decomposition. These findings are consistent with those of Cheng, Mayer, and Mayer (2018), Zambiasi and Stillman (2020) and Hodge and Hazel (2022) that respectively indicate RML leads to a 6% increase in Colorado's housing values, approximately a 10% to 20% increase in in-migration to Colorado, and up to a 6% increase in per-capita taxable food sales in Washington.

The potential impact of marijuana use on cognitive abilities also raises concerns about its influence on students' overall performance. Studies show divergent findings on the relationship between marijuana use and cognitive abilities, with some suggesting a negative effect (Bourque & Potvin, 2021; Scott et al., 2018) and others indicating no significant or positive impact on creativity (Hart et al., 2010; Schafer et al., 2012). Hence, these conflicting results imply

[3]To address spillover effects, I exclude control units near the treated state border and utilize alternative control groups, such as non-treated states that legalized medical marijuana.

[4]This clustering approach is conservative because it yields larger standard errors compared to those based on institution clustering (see Figure 2).

[5]The number of clusters when using states that legalized marijuana for medical use as a control group is 30, with 6 treated states and 24 control states. Similarly, the number of clusters is 47 when using all the states, omitting Maine, Michigan and Alaska due to noncontiguity or opening dispensary stores in or after 2019, the last year of my data. Following recent related literature (Agan & Makowsky, 2023), I report wild bootsrap p values or confidence intervals from 10,000 bootstrap iterations at the state-year subcluster following MacKinnon and Webb (2018).

[6]These estimates refer to enrollment growth in RM states relative to non-RM states (i.e., the mean differences in RM and non-RM states).

that the rise in enrollment due to RML may have uncertain effects on college completion and graduation rates. My findings reveal that RML has statistically significant positive effect on overall performance. Specifically, RML results in at least 10% and 13% more bachelor's and associate degrees being conferred. I also find significant effects on graduation rates at 100% and 150% of normal time. These results indicate that RML does not have a detrimental effect on degree completion or graduation rates.[7] Next, I attempt to determine the channels leading to this RML-induced enrollment growth.

In line with the neoclassical model of spatial equilibrium (Rosen, 1974; Roback, 1982) and the human capital model (Becker, 1964), it is important to also examine how RML affects first-time enrollments from in-state and out-of-state students in RM states. I find that RML increases out-of-state enrollments by at least 15% compared to neighboring non-RM states, but has no effect on in-state enrollments.[8] These findings indicate that RML improves the competitiveness of colleges in RM states compared to neighboring states by providing an additional college amenity, which is consistent with the gravity model of college enrollment (Leppel, 1993).

The second channel investigates two alternative explanations for enrollment growth in RM states: lower costs or higher quality of education.[9] However, I find no evidence that RML affects tuition and fee revenue per student or retention rate. Thus, these factors most likely did not contribute to the enrollment increase in RM states, supporting the hypothesis that RML serves as a consumption amenity. This is in line with the findings of Zambiasi and Stillman (2020) which show that RML-induced migration to Colorado is primarily because potential migrants regard marijuana as a positive amenity.

Next, I explore the heterogeneous effects of RML by institution type and find that RML only affects non-selective colleges with open admission policies, while selective colleges are unaffected. I also show that RML mainly affects public institutions and those offering only

[7]Previous studies (i.e., Marie and Zölitz (2017); A. C. Wright and Krieg (2020)) have found a negative effect of marijuana use on student grades. However, this does not imply that marijuana use affects college completion or graduation rates, which are the outcome of interest in my analysis.

[8]This result is distinguished from the main results in that it refers to how RML affects out-of-state enrollments relative to colleges that are within proximity to those in the treated states.

[9]Besides marijuana use, lower costs or higher quality of education may make RM states more attractive. For instance, RM states may have improved their postsecondary education quality or lowered their tuition fees by using the extra revenue from marijuana sales taxes (Marijuana Policy Project, 2022a).

associate and bachelor degrees, rather than doctoral research universities.[10] These effects are stronger for early adopters of RML, suggesting that RML attracts more students to non-selective colleges in early adopting states. However, as RML becomes more widespread and popular among Americans (i.e., 68% support RML (Gallup, 2021)), student sorting or relocation has gradually declined over time and that can explain why the enrollment growth is limited to early adopters. In addition, RM states vary in the laxity of their policies, allowing different amounts of plant cultivation per individual. I investigate how the RML effect depends on the permitted cultivation levels. I find that the policy effect is invariant to the amount of allowable cultivation.

I test the validity of the TWFE DID estimator using two methods. First, I conduct an event study analysis to determine whether the parallel trends assumption is most likely satisfied and find no significant pre-treatment enrollment differences between treated and control colleges, supporting the assumption. Second, I examine the potential spillover effects from nearby untreated locations. This could happen if students near the border of a treated state move or commute in order to have legal access of marijuana. Following literature that uses the same empirical approach (i.e., Chen et al. (2022); Mayer and Trevien (2017)), I drop control units or colleges within 80 miles of a treated border. The estimate decreases slightly when dropping control units up to 50 miles away, but remains stable for further distances up to 80 miles. This indicates that RML negatively affects control units within a 50-mile radius, supporting the hypothesis that the RML effect is due to student sorting rather than being a net gain.

The paper is organized as follows. I first provide background on policy definitions and trends in marijuana use and prices. I then describe the data and summary statistics. Next, I explain the empirical methods used. I then present the main results, the underlying mechanisms, and the heterogeneous effects. Finally, I conclude by summarizing the key points and discussing policy implications.

[10] These findings are in line with those of Marie and Zölitz (2017) and A. C. Wright and Krieg (2020) suggesting that RML has a greater effect on low-performing students.

2.1 Background

Marijuana legalization in the US is a complicated issue that involves varying degrees of legality. According to the federal law, cannabis is a Schedule I substance under the Controlled Substances Act of 1970, which means it is considered highly addictive and illegal to produce, distribute, or sell. However, some states have legalized marijuana for recreational use since 2012, creating a conflict with the federal law. The Justice Department responded by issuing guidelines that allowed states to regulate and enforce their own marijuana laws, as long as they met certain criteria, such as preventing the access of minors to marijuana (Cole, 2013). Nevertheless, there are ongoing efforts in Congress to decriminalize marijuana at the federal level, as evidenced by several bills that have been introduced (i.e., Marijuana Opportunity Reinvestment and Expungement Act (2022); States Reform Act (2021), and Cannabis Administration and Opportunity Act (2022)).

Recreational marijuana legalization is a policy trend that has swept across the US states in recent years. Before 2020, only 11 states had legalized cannabis for recreational use, either by ballot measure or by state legislature. The first two states to do so were Colorado and Washington, which passed ballot initiatives to legalize marijuana for adult use in 2012, followed by Alaska and Oregon in 2014, and California, Maine, Massachusetts, and Nevada in 2016. In 2018, Vermont became the first state to legalize cannabis by state legislature, while Michigan passed another ballot measure. The most recent state to join the list was Illinois, which legalized cannabis by state legislature in 2019. Figure 1 panel (a) illustrates the timing of RML for the states, which are included in the treatment group of this study. To gain a broader understanding of the policy effects, I consider the timing of law passage and the availability of cannabis for retail sale.[11]. However, 10 states joined the legalization trend in the past three years or between 2020 and 2022 (Marijuana Policy Project, 2022c).

One key assumption is that RML affects college choices because some students perceive the policy as a positive amenity. A large body of literature found a high prevalence of marijuana

[11] I omit DC, Alaska, Maine, Michigan, Hawaii, and all US territories from the analysis because they either do not share borders with other states or they lack sufficient post-policy data within the sample periods. The policy became effective in Oregon and Nevada at the end of the year, so I round up to the next year.

use among high school seniors and college students (National Institute on Drug Abuse, 2020; Schulenberg et al., 2021), as well as a positive association between marijuana legalization and marijuana use based on survey data (Cerdá, Wall, Keyes, Galea, & Hasin, 2012; Dave et al., 2023; Martins et al., 2016; Wen, Hockenberry, & Cummings, 2015). I present additional support for this assumption in Appendix B.3 by using Google Trends and PriceOfWeed.com data to show that RML is correlated with increased marijuana demand and decreased marijuana prices.

2.2 Data and descriptive statistics

The Integrated Postsecondary Education Data System (IPEDS) is a set of surveys conducted by the National Center for Education Statistics (NCES) that provides data on various aspects of postsecondary education, such as enrollment, degree completion, and institutional characteristics. In the 2019-20 academic year, there were 3982 title IV degree-granting institutions in the US. These institutions are eligible to receive federal student aid funds and are required to report their data to NCES through IPEDS surveys (National Center for Education Statistics, 2019, 2022a).

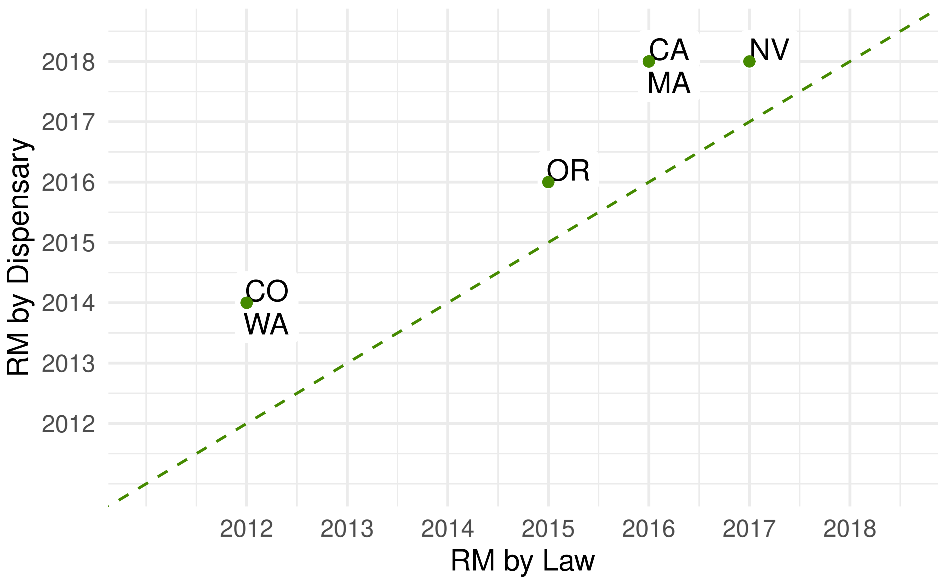
I use several IPEDS surveys, as well as county-level data, to construct a panel dataset of colleges and their characteristics from 2009 to 2019. I use the fall enrollment survey to measure first-time enrollments, the directory survey to obtain the geographical information of each institution, and other surveys (finance, residence and migration of first-time undergraduate students, and admission) to control for potential confounders and explore possible mechanisms. I also use the reviewed or complete versions of these surveys, which are released with a two-year lag and have higher data quality and accuracy. In addition, I control for some county-level variables that affect the enrollment decisions, such as the young population (age 18 to 24), the per-capita income, and the unemployment rate. I obtained these variables from the Bureau of Economic Analysis and the Bureau of Labor Statistics.

I report the summary statistics of all variables in Table 2.1. The table compares the mean values of the control and treatment groups. The treatment group consists of states that legalized recreational marijuana (RM), while the control group includes states that legalized medical marijuana (MM) only. The use of MM states as a control group may reduce the confounding effects of unobserved factors that affect both marijuana legalization and marijuana use.[12] Nonetheless, I use both MM states and all non-RM states for further robustness checks.

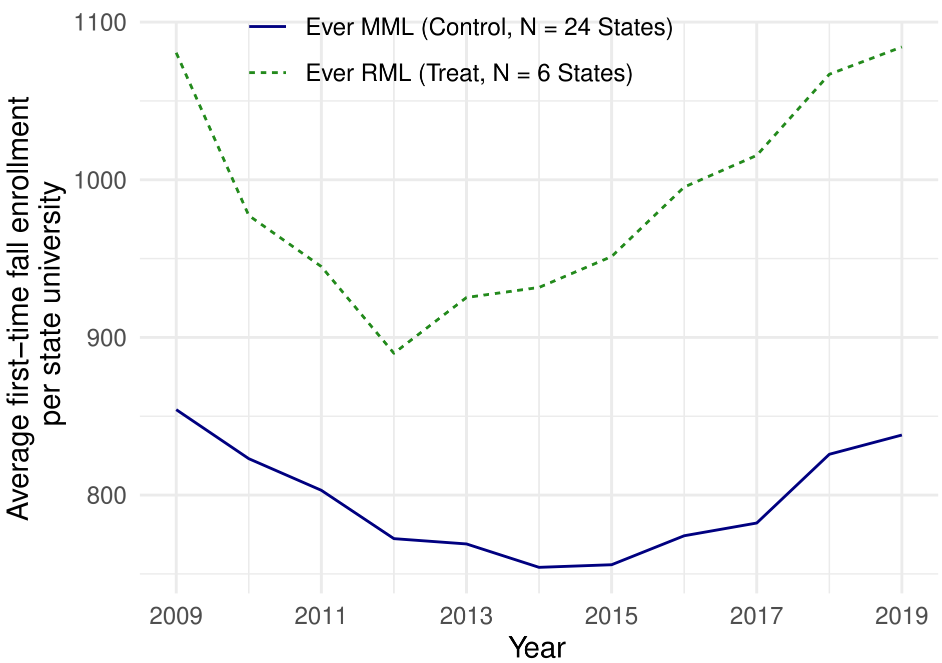
[12]This control group is preferable because it is more plausibly comparable to the treatment group in terms of unobserved characteristics, such as social attitudes towards marijuana or political environment. In Appendix Table B.1, I present the summary statistics for the alternative control group that includes all non-RM states.

Figure 1: Recreational Marijuana Legalization and Enrollment Trends

(a) Treated States Legalization Timeline



(b) Average Freshmen Enrollment, by Treatment Group



Note: Figure (a) displays the legalization dates of recreational marijuana (RM) for each state in the sample. The x-axis indicates the year when RM laws became effective (RM by Law), while the y-axis indicates the year when the first RM dispensary opened (RM by Dispensary). Appendix Figure B.1 provides more details on the RM legalization timeline. Figure (b) plots the average first-time enrollment for the treatment states (Ever RML) shown in (a) and the control states (ever MML), which legalized marijuana for medical use only by 2019. The treated and control states are mapped in Appendix Figure B.7. The analysis excludes vocational institutions and focuses on academic institutions only.

Table 2.1: Summary Statistics, by Recreational Marijuana Legalization Status (2009 to 2019)

	Non-RM States N Observations=16993 N institutions=1814		RM States N Observations=6534 N institutions=733		
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means
Log total first-time enrollment	5.676	1.660	5.765	1.844	0.089***
Log male first-time enrollment	4.717	1.901	4.775	2.110	0.058*
Log female first-time enrollment	5.057	1.697	5.179	1.814	0.122***
Log number of degrees	6.426	1.615	6.518	1.733	0.092
Aggregate enrollment over 20k dummy	0.059	0.236	0.099	0.299	0.040***
Offering medical degree dummy	0.038	0.192	0.029	0.168	-0.009***
Offering ROTC program dummy	0.309	0.462	0.226	0.418	-0.083***
Offering distance programs dummy	0.789	0.408	0.762	0.426	-0.028***
Log age 18 to 24 population	10.281	1.484	11.495	1.464	1.214***
Age 18 to 24 population female share	0.491	0.024	0.487	0.022	-0.005***
Unemployment rate	6.258	2.343	7.267	3.397	1.009***
Log per capita income	10.696	0.271	10.793	0.258	0.097***
Log Net migration	11.287	0.142	11.099	1.284	-0.187***
Log total out-of-state enrollment	6.448	1.685	6.507	1.852	0.058
Log RHG out-of-state enrollment	5.764	2.279	5.652	2.465	-0.112*
Tuition and fees revenue	16.573	1.699	16.586	1.611	0.012
Tuition and fees revenue per student	8.896	1.190	8.497	1.683	-0.399***
Log number of admissionss	7.139	1.612	7.131	1.960	-0.008
Log number of ACT and SAT submissions	6.284	1.215	6.681	1.345	0.397***
Bachelor's graduation rate (100% normal time)	37.490	22.796	41.507	25.451	4.017***
Bachelor's graduation rate (150% normal time)	51.994	20.853	57.255	22.455	5.260***
Bachelor's graduation rate (200% normal time)	53.914	20.338	59.888	21.538	5.974***
Associate's graduation rate (100% normal time)	19.179	19.326	16.618	18.001	-2.561***
Associate's graduation rate (150% normal time)	30.137	19.957	32.238	19.697	2.100***
Associate's graduation rate (200% normal time)	34.500	19.348	38.473	18.752	3.974***
Retention rate	67.119	17.120	70.376	16.816	3.257***
ACT 75th Math score	25.314	3.356	26.528	3.863	1.213***
SAT 75th Math score	592.911	71.508	611.573	77.858	18.662***
ACT 75th English score	25.883	3.833	27.042	4.437	1.159***
ACT 75th Crititcal Reading score	587.533	69.224	605.316	72.090	17.783***

Note: Enrollment refers to the number of first-time degree-seeking undergraduate students. RM states refer to Colorado, Washington, Oregon, California, Nevada, and Massachusetts which respectively opened the first dispensary in 2014, 2014, 2016, 2018, 2018, and 2018. Territories, Hawaii, Alaska, Maine, and Michigan are excluded either because they are noncontiguous to other states or because they open their first dispensary in 2019 or after (see Figure 1). The unit of observation is university-year. RHG stands for Recent High School Graduates and ROTC for Reserve Officers' Training Corps. Bachelor's and associate's degrees make up the total number of degrees conferred. The control group in this table consists of all states that have legalized marijuana for medical use, as shown in Figure B.1. Table B.1 presents similar summary statistics based on the control group of all states. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The table shows that institutions in RM states have significantly higher first-time enrollments than institutions in MM states, with a mean difference of about 9%. This suggests a potential positive effect of RML on college enrollment. Similarly, Figure 1 panel (b) indicates that college enrollment rates are higher in states that have legalized recreational marijuana (Ever RML) after 2012. While enrollment rates were higher in states that ever legalized RML at a later point, there is even greater growth in enrollment after 2012. However, this simple comparison does not account for other factors that may affect enrollment (i.e., demographics or economic conditions); therefore, I conduct a more rigorous analysis using a difference-in-differences approach, which I discuss in the next section.

2.3 Empirical strategy

I use two-way fixed effects (TWFE) difference-in-differences to estimate the overall impact of RML and I use the event study (i.e., dynamic TWFE difference-in-differences) to determine the effect over time. To ensure the robustness of my findings, I adopt alternative specifications as suggested by Sun and Abraham (2021) and Goodman-Bacon (2021).

2.3.1 Two-way fixed effects difference-in-differences

In order to estimate the effect of marijuana pervasiveness on student enrollment, I use the following TWFE difference-in-differences specification.[13] β_1 is the coefficient of interest measuring the RML effect on enrollments.

$$\log(Y_{ikjt}) = \beta_0 + \beta_1 \mathrm{RM}_{jt} + \delta_1 X_{kt} + \delta_2 Z_{it} + \delta_3 MM_{jt} + \psi_i + \theta_t + \epsilon_{ikjt} \qquad (1)$$

The variable of interest, denoted as Y_{ijkt}, represents the number of first-time fall enrollments in institution i located in county k and state j during the specific time period or year t. To investigate the impact of recreational marijuana legalization, I introduce the binary variable RM_{jt} as an indicator for the adoption of recreational marijuana laws. It is computed as the

[13] The canonical difference-in-differences does not allow for institution and time fixed effects because of their collinearity respectively with the treatment dummy (Treat) and policy time dummy (Post).

interaction between two dummy variables, $Post_t$ and $Treat_j$, following the standard difference-in-differences methodology. In this context, RM_{jt} is assigned a value of one if state j legalized marijuana in or after year t, as depicted below.

$$\text{RM}_{jt} = \begin{cases} 1 & \text{if state } j \text{ legalized RM in or after time } t \\ 0 & \text{Otherwise.} \end{cases}$$

X represents the set of baseline county covariates, while Z denotes an extensive set of college-level covariates which are reported in Table 1. The notation ψ_i is the college fixed effects that address any time-invariant heterogeneity within an institution. As shown in equation (1), I use the most robust model specification by incorporating the college fixed effects. However, this choice restricts the ability to simultaneously control for county-fixed effects because most colleges are situated in separate counties. The county covariates are added to control for potential high school graduates within proximity to the college and other important county-level factors that may influence enrollments such as unemployment rate, per-capita income, and net migration. I also control for medical marijuana legalization to account for any potential medical marijuana policy effect on enrollments. MM_{jt} is a control dummy for states that legalized marijuana for medical use. MM_{jt} is a binary variable for the adoption of medical marijuana law. MM and RM are staggered dummies where RM represents a continuation of MM for states that changed the legalization status from medical to recreational.

$$\text{MM}_{jt} = \begin{cases} 1 & \text{if state } j \text{ legalized medical marijuana but not RM in or after time } t \\ 0 & \text{Otherwise.} \end{cases}$$

As shown in Figures 1 and B.1, all the RM states have first legalized medical marijuana before proceeding to recreational legalization. This indicates states that legalized medical marijuana are more likely to have similar unobserved characteristics as RM states (i.e., social movements leading to legalization). For the RML definition, I consider both the timing of policy passage and the opening of the first dispensary.[14]

[14]The main analysis shows the results obtained using both MM and all non-RM states as control groups. For the sake of parsimony and to streamline the presentation of results, the medical states control group and the dispensary

2.3.2 Event study

I run the following event study or dynamic difference-in-differences model in which I regress the logged first-time enrollments on G leads and M lags and normalized them to the first pre-shock year (reference group). This specification allows for the policy effect to change over time. It is also crucial to examine policy anticipation or trends in RM and non-RM states before the shock event. As a robustness check, I incorporate the event study specification by Sun and Abraham (2021) which assimilates pre-treatment and post-treatment dynamics to resolve the problem of multiple units being treated in different time periods.

$$\log(Y_{ikjt}) = \psi_i + \theta_t + \sum_{m=-G}^{M} \beta_m RM_{i,t+m} + \delta_1 X_{kt} + \delta_2 Z_{it} + \delta_3 MM_{jt} + \epsilon_{ikjt} \tag{2}$$

2.3.3 Other specifications

I relax the assumption of constant average treatment effect in the TWFE DID by using two alternative methods that account for the heterogeneity in treatment effects due to the staggered policy adoption (Callaway & Sant'Anna, 2021; De Chaisemartin & d'Haultfoeuille, 2020; Goodman-Bacon, 2021). First, I use Sun and Abraham (2021), which extends the TWFE DID to a dynamic setting and allows for varying treatment effects across states and time periods. Second, I use Goodman-Bacon (2021), which decomposes the TWFE DID into a weighted average of all possible pairwise DID estimators. I show that both methods yield similar results to the TWFE DID, suggesting that the estimates are unbiased. I also consider different definitions of RML and control groups, as discussed previously, to enhance the credibility of the estimates.

RML timing will be used in all subsequent analyses. For the mechanism analysis, I show results obtained using the RML law passage timing and the different control groups in the appendix.

2.4 Results

2.4.1 Main results

The aim of this paper is to examine the effect of RML on academic enrollment.[15] I estimate the RML policy coefficient on first-time enrollment using OLS based on Equation 1.[16] Figure 2 displays the results, which vary by the definition of RML (either the timing of dispensary openings or the timing of law approval), the type of control group (either MM states or all non-RM states), and the type of standard error clustering. Each sub-figure has two rows and two columns, showing a different combination of these factors. These main results indicate that RML increases enrollment by 4.6% to 9%.[17] The upper bound of approximately 9% remains statistically significant at the 5% level even when using the most-conservative wild cluster bootstrap standard errors.[18] Table B.2 shows that the exclusion of control covariates yields statistical significance at the 10% level. To account for potential bias due to few clusters at the state level, I also report wild bootstrap confidence intervals following MacKinnon and Webb (2018). Throughout the rest of the paper I report error estimates that account for the fact that treatment is assigned at the state level rather than the unit level (i.e. university level).[19]

The results also show that RML has a significant positive effect on both men and women at the 10% level, particularly based on the preferred control group and RML by law (i.e., the passage of law without dispensaries opening). However, the difference between men and women's estimates is not statistically significant at the 5% level.[20]

[15] As shown in Figure B.3, RML does not significantly affect the enrollment of vocational institutions. This result is consistent with previous studies (e.g., Marie and Zölitz (2017); Roebuck et al. (2004); A. C. Wright and Krieg (2020)) that mainly explore the relationship between marijuana use or legalization and academic outcomes.

[16] Tables B.3 and B.2 provide more details on the estimation. In Table B.3, the control group consists of all non-RM states from 2009 to 2019 (see Figure B.7 for the list of MM and non-MM states).

[17] When I account for the timing of dispensary openings and use all unaffected states as a control group, I obtain a slightly lower effect of 4.6%

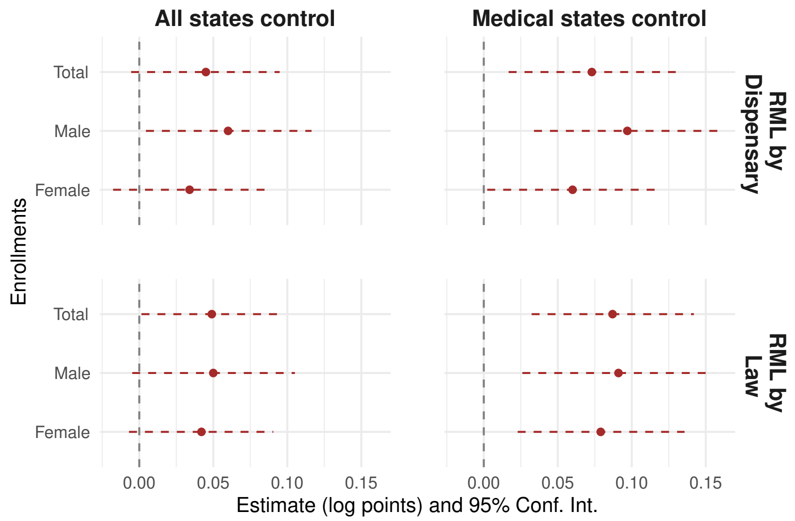
[18] Blake et al. (2023) found a 5.2% increase in college applications due to RML. However, my analysis considers actual enrollment figures, noting that applications to non-open admission universities may not always result in enrollments, particularly at highly selective institutions with limited capacity.

[19] Clustering at the unit level (i.e. university level) may be natural in other instances since the fixed effects are taken at that level in order to account for the heterogeneity in college policies (e.g., substance use regulations) and other unobserved college-specific factors that may affect enrollment. However, I prefer the state level cluster because the treatment is assigned at the state level, and this clustering approach provides conservative estimates.

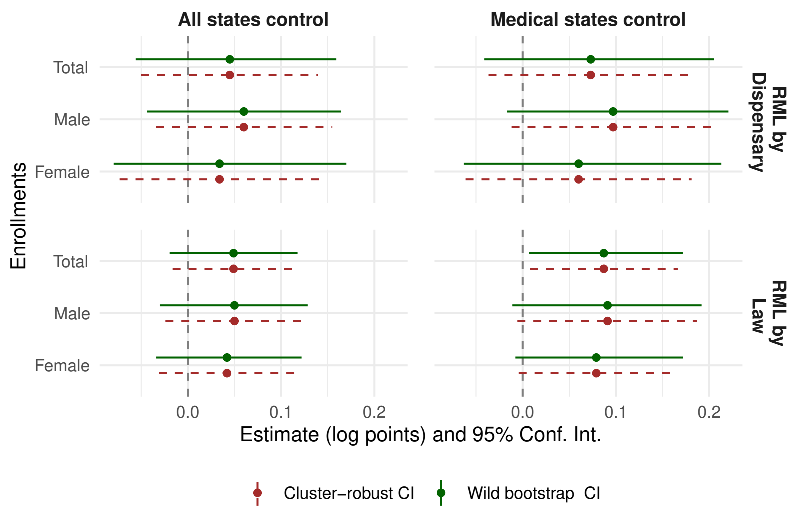
[20] The preferred control group comprises states that have legalized marijuana solely for medical purposes. This selection is considered superior for its enhanced comparability with the treatment group in terms of unobserved characteristics, such as societal attitudes toward marijuana and the political environment.

Figure 2: RML Effect on First-Time Enrollment, by Legalization, Control Group, and Standard Errors Clustering Types

(a) College Level Clustering

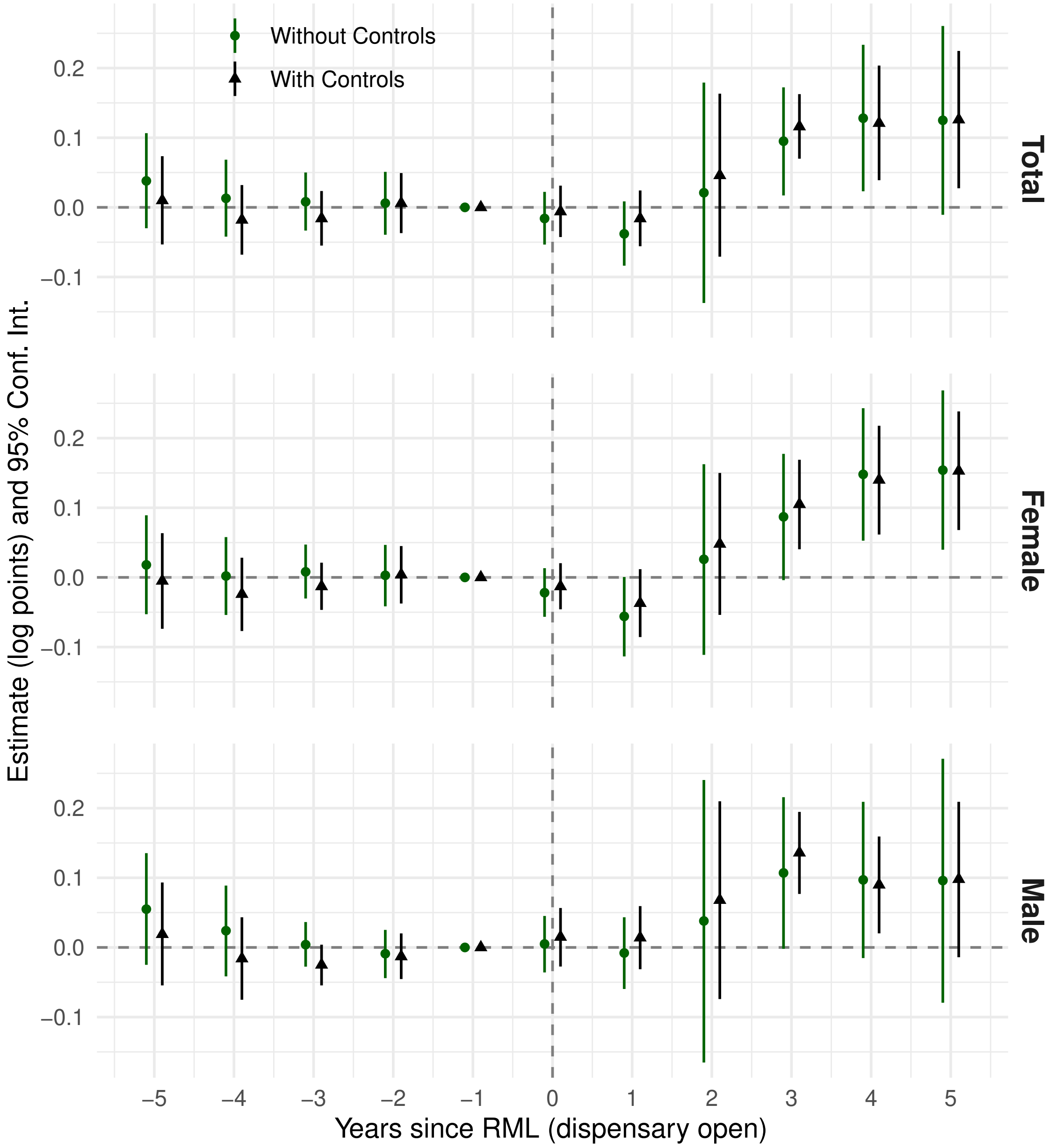


(b) State Level and Wild-Bootstrap Clustering



Note: The plots depict the RML effect estimates on log first-time enrollments in academic institutions based on the TWFE difference-in-differences. Each row varies the estimate by how legalization is defined, and each column varies the estimate by how the control group is defined. The legalization status refers to the type of recreational marijuana legalization used for the policy definition as shown in Figure 1. Medical states are those states that legalized marijuana for medical use as of 2019 as illustrated in Figure B.7. The models include year and institution fixed effects as well as baseline covariates. The institution-level controls are as follows: medical degree dummy, college size dummy (having over 20k total enrollments), student-to-faculty ratio, distance program dummy, and Reserve Officers' Training Corps (ROTC) program dummy. The county-level controls include the age 18 to 24 population, age 18 to 24 female population share, per-capita income, unemployment rate, and net migration. Standard errors are clustered at institution level in (a) and at the state level in (b). Due to the small number of treated clusters for the RML estimates in (b), I follow MacKinnon and Webb (2018) to report the RML wild bootstrap confidence intervals (Wild bootstrap CI). Further estimation details are shown in Tables B.2 and B.3.

Figure 3: Event Study Graph of RML Effect on Enrollment



Note: The plots depict the RML effect estimates on log first-time enrollments over time based on Sun and Abraham (2021) model. The zero value on the x-axis refers to the first year of legalization. RML policy is defined by the first dispensary opening time as shown in figure 1, and the states which legalized marijuana for medical use are used as a control group. Control variables and fixed effects are as described in Figure 2. Standard errors are clustered at the state level.

The above DID analysis results show that on average RML has a positive effect on college enrollment in RM states. To better understand how the effect changes over time, I present the event study analysis results (equation 2) following Sun and Abraham (2021). Figure 3 shows that total first-time enrollment increased significantly by 10% to 12% after the third year of policy implementation.[21] Estimates with and without controls remain significant at the 5% or

[21] Figure B.4 displays similar results when RML is measured by legal status alone without dispensary availability.

10% level in the third and fourth year of legalization for all outcomes.

The results indicate that both women's and men's enrollments rose significantly after the 4th year of the first dispensary opening. However, the p-value for men is smaller than 10% but larger than 5% in the last year. The law abiding behavior could be one of the possible explanation for the high statistical significance of women's estimates over time. Portillo and DeHart-Davis (2009) empirically found that women abide by the law more than men even when having organizational power. [22] There are also some possible reasons for the delayed response to the policy.

First, there is a slow and gradual development of a marijuana consumption culture or network among college students who use more marijuana (Rinker, Krieger, & Neighbors, 2016). Marijuana use among young adults is positively associated with higher perceived rates of marijuana use by peers (Koval, Kerr, & Bae, 2019), and the internalization of a marijuana use culture within college settings has led to increased consumption of the substance among students (Pearson et al., 2018). Moreover, the college choice and application process usually starts at least a year before the policy adoption year (College Board, 2022). Another explanation could be the slow growth of the marijuana industry after the legalization since most states take two years to start licensing retail stores.

These event study graphs also support the DID's parallel trend assumption. Figure 3 shows that there is no significant difference in enrollment between RM and non-RM states before the policy intervention. There is no major difference in significance level or magnitude between the estimates based on TWFE DID and Sun and Abraham's method.

The above findings indicate that some college students perceive marijuana as a positive amenity through increasing college enrollments. However, I acknowledge that other factors may confound the relationship between RML and enrollments. RM states may use the additional tax revenue from marijuana sales to subsidize their higher education sector. The human capital model of Becker (1964) and related literature (e.g., Behrman, Constantine, Kletzer, McPherson,

These results are also robust to the use of all non-RM states as a control group, as shown in Figure B.5. Zero in the x-axis refers to the first year in which the first dispensary opened.

[22]RML may have a greater impact on women's enrollment for several other reasons. Supporting this notion, Marie and Zölitz (2017) and H. G. Pope, Jacobs, Mialet, Yurgelun-Todd, and Gruber (1997) demonstrated that marijuana prohibition or consumption had a more pronounced impact on female test performance.

and Schapiro (1996); Cook and Zallocco (1979); Hemelt and Marcotte (2011)) suggest that college quality and price are the main determinants of college choice. I present my analysis of these mechanisms in the next section. First, I discuss whether RML also leads to an increase in degree completions.

2.4.2 RML effect on completion and graduation rates

I test whether RML affects students' overall performance by examining its impact on the number of degrees conferred. I restrict the treatment group to Washington and Colorado, which are the only states that have legalized recreational marijuana for a sufficiently long period.[23] This allows me to use the outcome leads to measure the effect of RML on student cohorts who have different exposure durations (i.e., time needed to complete the degree) to the policy. Table 2.2 shows that RML has an approximately 10% to 13% significant effect on undergraduate degree completions (bachelor and associate degrees). In addition, RML exhibits a notable positive impact on graduation rates, contributing to increases of up to 2.7 percentage points for bachelor's degrees and 5.6 percentage points for associate degrees. The graduation rate for associate degrees shows a significant effect from the fourth lead, indicating a delayed response to the policy as discussed in the main result section. On the other hand, the graduation rate for bachelor degrees shows a significant effect from the second lead, which may reflect increasing student transfers. Taken together, RML does not seem to be detrimental to the overall performance of students (i.e., degree completion on time and graduation rates). [24]

Despite Marie and Zölitz's (2017) findings that RML leads to diminished grades, particularly in courses requiring numerical skills, my analysis reveals that RML does not undermine overall student success. My results points to an uptick in college graduations associated with the increase in enrollments due to RML. However, this does not negate Marie and Zölitz's (2017) findings. Due to the lack of Grade Point Average (GPA) data in the IPEDS dataset, I cannot assess the impact of RML on overall student scores or GPA.

[23]For example, a student cohort that started in year t finishes bachelor education in t+4, t+5, or t+6. This means the cohort should be exposed to the policy for 4 to 6 years.

[24]When I use all states that did not legalize marijuana for recreational use as a control group instead of those that only legalized medical marijuana (see Table B.4 in appendix), some of these results become insignificant.

Table 2.2: Effects of RML on Undergraduate College Completion and Graduation Rate

	Lead 1	Lead 2	Lead 3	Lead 4	Lead 5	Lead 6
Panel (a): Log Number of Undergraduate Degrees						
	Bachelor Degrees					
RM	−0.028	0.007	0.035	0.120***	0.125***	0.105**
	(0.068)	(0.055)	(0.042)	(0.027)	(0.025)	(0.029)
Wild bootstrap p	0.736	0.932	0.518	0.017	0.007	0.023
N Obs.	12 137	10 644	9198	7794	6473	5214
	Associate Degrees					
RM	0.043	0.063	0.099	0.142**	0.138*	0.097
	(0.072)	(0.058)	(0.051)	(0.049)	(0.052)	(0.048)
Wild bootstrap p	0.604	0.401	0.160	0.044	0.075	0.153
N Obs.	13 938	12 125	10 393	8748	7206	5746
Panel (b): Bachelor Degrees Graduation Rate (Percentage Points)						
	100% of Normal Time (4-years)					
RM	−0.146	0.747	2.480*	0.614	0.590	1.222
	(0.815)	(0.990)	(0.974)	(0.547)	(1.276)	(1.568)
Wild bootstrap p	0.886	0.567	0.102	0.462	0.730	0.583
	150% of Normal Time (6-years)					
RM	0.593	1.857*	2.744*	1.007	0.832	−0.243
	(1.114)	(0.829)	(1.073)	(0.695)	(1.954)	(1.476)
Wild bootstrap p	0.665	0.127	0.093	0.341	0.768	0.902
N Obs.	9289	9345	8517	7676	6840	5991
Panel (c): Associate Degrees Graduation Rate (Percentage Points)						
	100% of Normal Time (2-years)					
RM	1.965	0.950	2.380	3.802*	5.087**	5.599***
	(2.580)	(2.901)	(1.944)	(1.495)	(1.572)	(0.936)
Wild bootstrap p	0.522	0.812	0.442	0.108	0.071	0.010
	150% of Normal Time (3-years)					
RM	3.064	2.615	0.911	0.645	0.859	0.889
	(1.900)	(1.685)	(1.527)	(1.038)	(0.953)	(1.510)
Wild bootstrap p	0.181	0.212	0.631	0.577	0.475	0.712
N Obs.	4958	4888	4391	3904	3424	2954
Controls	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
College FE	Y	Y	Y	Y	Y	Y

Note: The outcome variable in panel (a) refers to the logged number of all undergraduate degrees (associates and bachelors) conferred. In panels (b) and (c), the dependent variable is the graduation rate within the standard degree completion time (i.e., 4 years for bachelor) or or within 150% of the normal time (i.e., 6 years for bachelor). The graduation rate refers the percentage of students in a cohort who completed their degrees within the specified times (100% or 150% of normal degree time). The policy is defined by law approval timing (see Figure 1) to allow the use of 6 leads. This is in contrast to prior exhibits which are based on the dispensary opening. A lead of "i" (ranging from 1 to 6) indicates that the ith lead of the outcome variable was used in the analysis. Colorado and Washington are the only treated states. Nevada, California, Oregon, and Massachusetts are dropped from this analysis due to the nonexistence of observations for leads of the outcome variable. The control group includes states that legalized marijuana for medical use (see figure B.7 for more details). I report wild bootstrap p values following MacKinnon and Webb (2018) using 10,000 bootstrap iterations at state-year subcluster. Control variables and fixed effects are as described in Figure 2. Standard errors are clustered at the state level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

2.5 Mechanisms

I investigate the mechanisms and causes of the positive effect of RML on first-time enrollment. First, I show that RML increases the number of out-of-state enrollments compared to neighboring states, suggesting that RML provides a competitive advantage for colleges in RM states. Second, I examine whether college prices and quality could have changed due to potential investments from marijuana sales tax revenue in higher education.

2.5.1 RML effect on out-of-state enrollment, by proximity status

In this section, I examine how RML affects the enrollment decisions of and out-of-state students, and whether the effect varies by the distance to the treated states. Following the neoclassical model of spatial equilibrium by Rosen (1974) and Roback (1982), an exogenous change in amenities (RML) will lead to student movement. In addition, students are more likely to enroll in a college if the distance to the college is small (Leppel, 1993).

Figure B.8 illustrates the states included in the treatment and control groups for contiguous and noncontiguous states. The treatment group includes only the three early policy adopters that have enough observations after the policy implementation.[25] Figure 4 shows that RML improves the total out-of-state enrollments by more than 30% relative to contiguous states. The effect is not significant relative to noncontiguous states.[26] In addition, Table B.7 shows that the policy has no effect on in-state enrollment. These findings suggest that RML increases enrollment by providing a competitive advantage relative to neighboring states. [27]

Given that RML positively affects out-of-state enrollment, it is also vital to explore its impact on out-of-state recent high school graduate (RHG) enrollments. Figure 4 shows that RML has no effect on RHG enrollments at 5% level. The insignificance of the effect on RHG could be due to the legalization policy that permits only individuals aged 21 or above to buy cannabis legally, so perhaps non-RHG have a greater incentive to attend colleges where recreational marijuana is

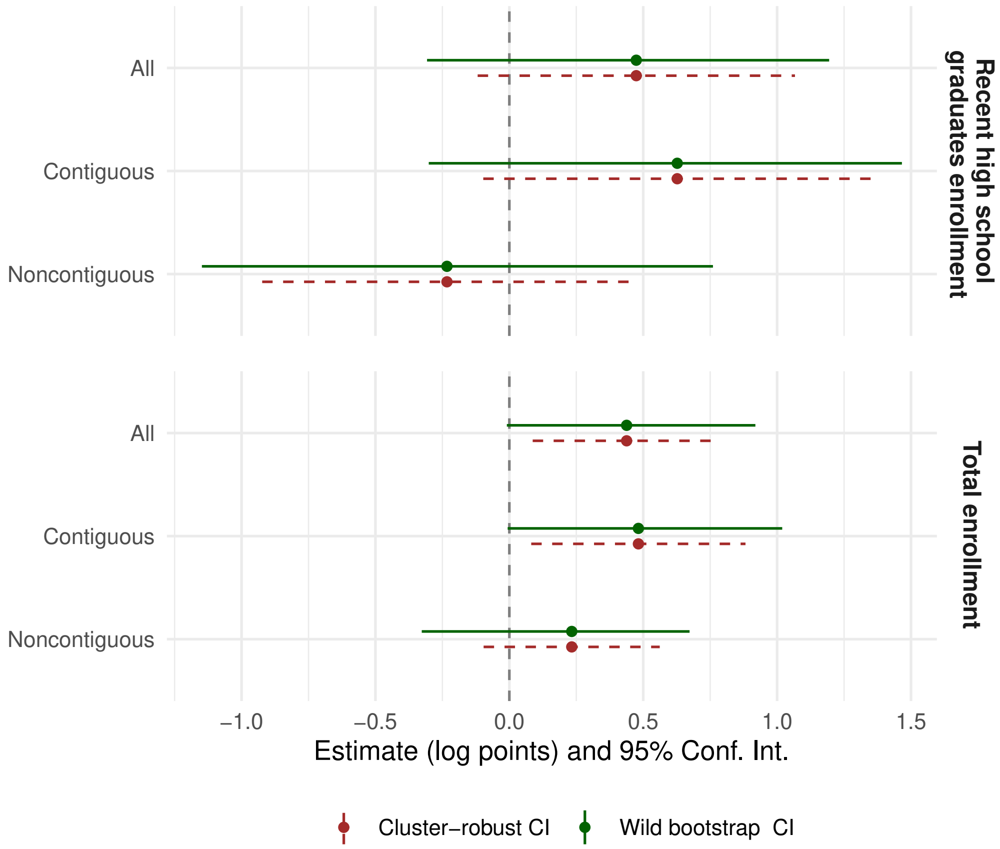
[25]The IPEDS survey on residence and migration of first-time undergraduate students is only available for even-numbered years from 2008 to 2018.

[26]Figure B.10 shows these results are robust to the use of a different RML definition (mere legality). Contiguous and noncontiguous groups are shown in figures B.8, and medical states are shown in figure B.7.

[27]Since the control group contains contiguous units, these estimates may overestimate the true RML effect as the spillover analysis and results from Figure 6 indicate.

legal. This follows the literature side documenting negative or no association between marijuana laws and the odds of marijuana use among youths (D. M. Anderson, Hansen, Rees, & Sabia, 2019; Dilley et al., 2019).[28]

Figure 4: RML Effect on Out-of-State Enrollment, by Contiguity Status



Note: The plots depict the RML effect estimates on log total and recent high school graduates' out-of-state first-time enrollments in academic institutions based on TWFE difference-in-differences. Contiguous refers to states that are contiguous or within close proximity to RM-states as shown in Figure B.8. The treatment group includes early movers (Colorado, Washington, and Oregon) because out-of-state data is available only biennially from 2008 to 2018. RML policy is defined by the first dispensary opening time as shown in Figure 1. Control variables and fixed effects are as described in Figure 2. Standard errors are clustered at the state level. Dashed confidence intervals are based on cluster-robust standard errors, and the solid ones on 10,000 subcluster (state-year) wild-bootstrap iterations as suggested by MacKinnon and Webb (2018).

Consistent with the gravity model (Leppel, 1993), these results suggest that distance from the affected states is important and support the hypothesis that improvement in first-time enrollment is driven by the gain of competitive advantage relative to neighboring states. These findings also indicate that the RML effect on enrollment is not a net gain, but rather a redistribution of students across states.

[28]Cerdá et al. (2017) showed that RML increases marijuana use among younger adolescents by decreasing their perception of harm from marijuana use.

2.5.2 RML effect on college price and quality

Table 2.3 shows that RML has no effect on tuition per student and retention rate.[29]This suggests that the growth in enrollments is not due to any price or quality changes but most likely because some students perceive marijuana as a college consumption amenity. Column (2) in Table 2.3 shows that the total tuition revenue grew significantly by approximately 8%.[30] To gain a broader understanding of the policy's effect, the next section examines its heterogeneous effects.

Table 2.3: Effects of RML on Tuition Revenue and Retention Rate

	Tuition Revenue		Tuition Revenue Per Student		Retention Rate	
	(1)	(2)	(3)	(4)	(5)	(6)
RM	0.088	0.079*	0.075	0.027	−0.009	−0.028
	(0.051)	(0.038)	(0.044)	(0.019)	(0.021)	(0.022)
Wild bootstrap p	0.117	0.059	0.117	0.317	0.675	0.217
N Obs.	22 558	22 558	22 556	22 556	21 998	21 998
N college	2547	2547	2547	2547	2547	2547
Controls	Y	Y	Y	Y	Y	Y
College FE	Y	Y	Y	Y	Y	Y
Year FE	N	Y	N	Y	N	Y

Note: The tuition and fees variables are derived by binding all the IPEDS finance surveys. Since these surveys depend on the institution types (public or private) and the accounting standards (GASP and FASB), I controlled for these types in the tuition models. Tuition refers to all the tuition and fees (logged) collected during a year period. Retention rate is defined on IPEDS's survey as "the percent of the (fall full-time cohort from the prior year minus exclusions from the fall full-time cohort) that re-enrolled at the institution as either full- or part-time in the current year". I report wild bootstrap p values following MacKinnon and Webb (2018) using 10,000 bootstrap iterations at state-year subcluster. Control variables and fixed effects are as described in Figure 2. Standard errors are clustered at the state level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

2.6 Heterogeneous effects

In this section, I analyze the impact of the policy on various types of colleges and the states that implemented it. First, I demonstrate that colleges with admission requirements remain unaffected, indicating that the policy's effect is primarily relevant to colleges without stringent admission criteria. Second, I find that the policy does not influence private colleges and that the effect on enrollment is likely driven by states that adopted the policy earlier.

[29]Table B.5 shows that RML has no impact on tuition per student and the retention rate under different policy definitions and control groups.

[30]This result is also significant when using a different policy definition (RML by law) as shown in B.5.

Table 2.4: Effects of RML on Admissions and Test Scores for Selective Colleges

	N Admission	N Tests Submissions	ACT Math	SAT Math	ACT English	SAT Reading
	(1)	(2)	(3)	(4)	(5)	(6)
RM	0.027	−0.011	0.072	0.076	0.142*	0.004
	(0.041)	(0.043)	(0.048)	(0.043)	(0.056)	(0.042)
Wild bootstrap p	0.585	0.830	0.201	0.153	0.043	0.925
N Obs.	11 258	8395	6862	7938	6863	7881
N college	1329	933	841	909	841	909
Controls	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Colleges FE	Y	Y	Y	Y	Y	Y

Note: The dependent variables in columns (1) and (2) refer respectively to the logged number of admitted students and the logged number of SAT and ACT test scores received for admission purposes to an undergraduate program. In columns (3) to (6), the outcome variables are the standardized 75th percentile scores for each test type. The sample control units include the states which legalized marijuana for medical use. The included states are depicted in Figure B.7. The policy dummy variable is defined at the dispensary opening time of legalization. I report wild bootstrap p values following MacKinnon and Webb (2018) using 10,000 bootstrap iterations at state-year subcluster. Control variables and fixed effects are as described in Figure 2. Standard errors are clustered at the state level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.4 displays the impact of RML on admissions, test score submissions, and standardized ACT and SAT scores for selective colleges (those with admission requirements).[31] While RML does not significantly affect the number of admissions, it does show a negative but insignificant impact on the number of test submissions and a positive impact on ACT English percentile score.[32] One possible explanation for this result is that RML may have influenced students' preferences, leading some to prioritize open policy institutions. These findings align with Jacob et al.'s (2018a) indicating that high-achieving students prioritize academic quality over consumption amenities.

Next, I examine how RML affects first-time enrollments in different types of academic institutions. Figure 5 (panel a) shows that RML has a positive and significant impact at 10% level based on wild bootstrap only on public colleges that offer undergraduate degrees (bachelor's). Panel (b) reveals that the effect of RML varies by the timing of policy adoption. Colorado and Washington (the pioneers in RML) experience a large increase in enrollments in institutions that offer bachelor's degrees as their highest level of education by about 21%. One possible reason

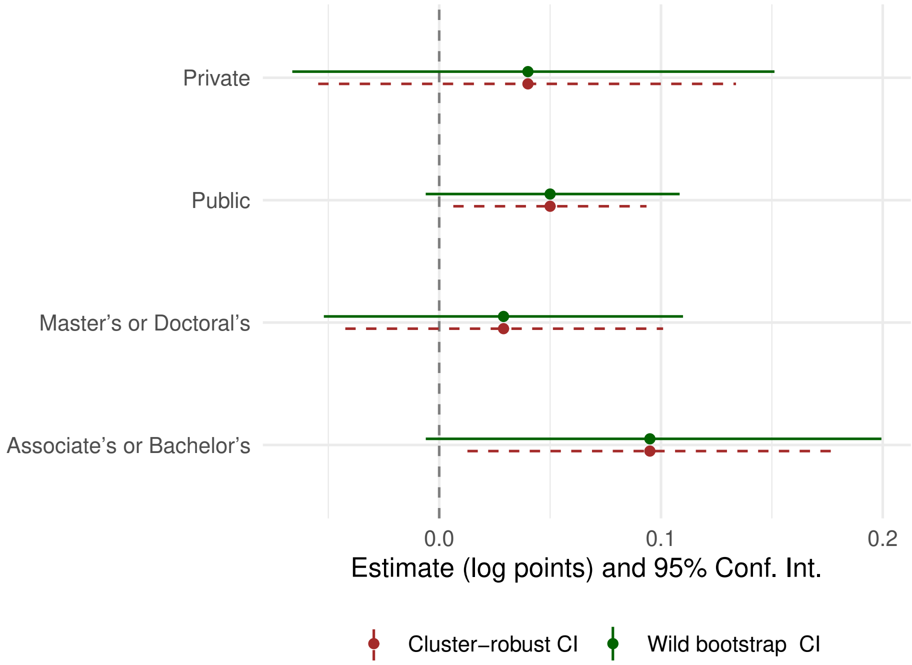
[31] Selective colleges here refer to all institutions with any type of admission policies such as high school rank or GPA, completion of a college preparatory program, or teachers' recommendation.

[32] Table B.6 shows that these results are robust to different control groups and RML definitions.

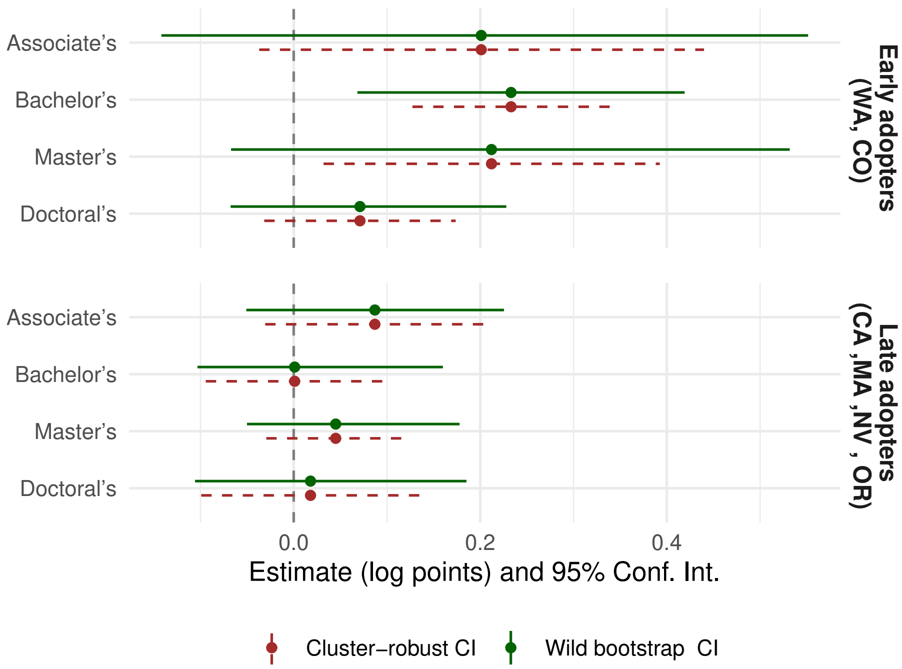
for this impact is that they face no competition from neighboring states for at least three years after legalization (see Figure 1).

Figure 5: Heterogeneous Effects of Recreational Marijuana Legalization

(a) RML Effect by All Institution Types



(b) RML Effect by Public Institutions Types and The Timing of Adoption



Note: RML policy is defined by the dispensary opening as shown in Figure 1. Degree names refer to the institutions' highest degree conferred. For example, "Associate's" indicates institutions (community colleges) that offer an associate degree as their highest degree, and "Doctoral's" refers to research institutions that offer Ph.D. programs. Panel (a) includes both public and private institutions, and panel (b) is restricted to only public institutions. Control variables and fixed effects are as described in Figure 2. Standard errors are clustered at the state level. Dashed confidence intervals are based on cluster-robust standard errors, and the solid ones on 10,000 subcluster (state-year) wild-bootstrap iterations as suggested by MacKinnon and Webb (2018).

Last, I examine whether the effect of recreational marijuana legalization depends on the policy strictness. For example, some states with RML allow or prohibit individuals from cultivating cannabis. I use the data from Borbely, Lenhart, Norris, and Romiti (2022) to categorize RM states by the amounts of cultivation allowed. To test this heterogeneity, I adjust equation 1 by interacting the policy dummy (RM) with an indicator for different cultivation categories (4, 6, or no plants).[33] Results in Table B.8 show no effect for states (CA, NV, OR, CO, and MA) that allow 4 to 6 plants cultivation. The effect is positive and significant for the remaining state (WA) which prohibits cultivation. Nonetheless, this positive effect is most likely due to the early adoption of the policy. Therefore, the policy effect does not seem to vary by the strictness of RML.

2.7 Robustness checks

2.7.1 Sensitivity to proximity spillover

While the common trend assumption in difference-in-differences (DID) analysis is likely satisfied, as discussed earlier and supported by Figure 3, the stability of the treatment unit assumption requires further investigation. This assumption assumes that the treatment and control groups remain stable over time, with no systematic changes or spillover effects between them.

A major factor that could introduce bias in the main results in Figure 2 is the proximity of students to state borders where recreational marijuana is legal.[34] These students may be attracted to the policy and opt to commute or migrate to the nearest college in the neighboring state where marijuana is legally allowed. For instance, if students from neighboring states start enrolling in Colorado colleges, it could lead to an overestimation of the true RML effect.

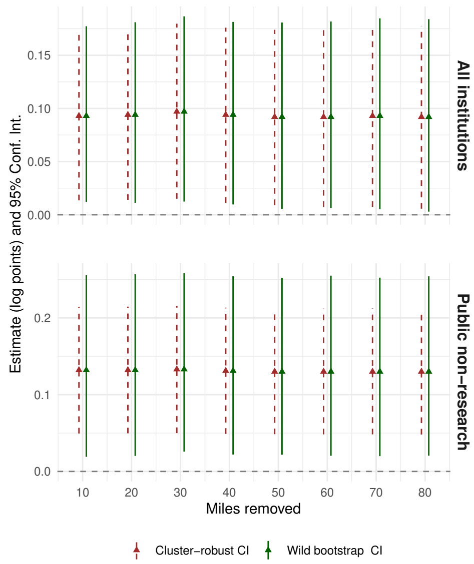
To investigate this potential bias from extensive commuting or migration within borders' proximity, I exclude colleges located near the border as illustrated in Figure B.9. Figure 6 shows how the estimates vary as I remove control units within close distance to RM states.[35]

[33]The data shows the cultivation amounts by state with recreational marijuana legalization in Table A.2 (Borbely et al., 2022).

[34]A significant spillover effect across borders can potentially violate the stable unit treatment assumption, which is a key assumption of the DID specification.

[35]For standard error clustering at the university level, Table B.9 presents the results, which demonstrates that the

Figure 6: Effects of RML on Enrollment Sensitivity to Proximity Spillover



Note: The outcome variable is log first-time enrollments. All institutions refer to main results sample, which contains all academic institutions. Public non-research refers to the interaction between RML dummy and a binary for public institutions that do not offer graduate degrees (i.e., offering only bachelor or associate degrees). To assess potential bias in the estimates caused by spillover effects from students commuting or migrating to policy-affected colleges, each column shows the results after excluding control colleges located within a certain distance of the policy-affected border. For instance, 80 miles removal displays the results obtained after excluding all untreated colleges within 80 miles of the border of the treated states. Figure B.9 illustrates the geographical distribution of control and treatment institutions. Control variables and fixed effects are as described in Figure 2. Standard errors are clustered at the state level. Dashed confidence intervals are based on cluster-robust standard errors, and the solid ones on 10,000 subcluster (state-year) wild-bootstrap iterations as suggested by MacKinnon and Webb (2018).

The top panel in Figure 6 replicates the main results after removing these within-proximity control units.[36] The effect stays at approximately 8% when removing units over 40 miles. This small spillover indicates that RML attracted some students within proximity. Interacting the policy variable (RM in equation 1) with a binary indicator for public non-research institutions, the bottom panel reveals a more pronounced effect for these institutions, maintaining just below 14% even when excluding distances over 50 miles. This follows the neoclassical model of spatial equilibrium which suggest students' relocation to maximize their utility from college amenities. The results also show that the RML effect is concentrated among non-research public institutions.

2.7.2 Alternative estimation methods

The timing effect plays a significant role in the policy analysis context, particularly as different states legalized recreational marijuana at different times. Therefore, it is crucial to comprehend

effect remains significant even after removing colleges within 80 miles away from a treated state border.

[36]These results are based on the upper bound estimate of 9% in Figure 2.

the underlying factors influencing the overall point estimate. In line with Goodman-Bacon (2021), the DID method allows for the decomposition of the point estimate into subcomponents, namely treated versus treated and treated versus untreated. Figure B.6 shows the results of this decomposition. Larger weights are assigned to the treated versus untreated subcomponent. This indicates that differences in enrollments between states with and without recreational marijuana legalization are the primary drivers of my estimate. Specifically, the first two early movers, Colorado and Washington, heavily influence the estimate by contributing to a positive 20% improvement in enrollments, with relatively large weights.[37] As expected, the weighted sum of the decomposition is approximately 9%, the estimate from the main results.

2.8 Conclusion

This paper examines the impact of RML on fall freshmen enrollments. The findings indicate that RML leads to a 4.6% to 9% increase in freshmen enrollments in RM states compared to non-RM states. The policy positively affects out-of-state enrollment and has no effect on enrollment, college quality or price and the effect is concentrated among early adopters and public non-research colleges. The policy has no effect on selective colleges, possibly because of their limited capacity or their students' preference for college quality and expected future earnings over college amenities. I further find that RML does not negatively impact college completion, suggesting that some students perceive RML as a positive amenity that influences their college choice.

These findings follow at least two conceptual frameworks. They are consistent with the gravity model (Leppel, 1993) because the proximity to the treated colleges drives the RML effect. Second, the neoclassical model by Rosen (1974) and Roback (1982) implies that students' relocation due to the rise of new amenities will result in no utility gain at equilibrium.[38] These theoretical frameworks suggest that students relocate to maximize their consumption of college amenities. The findings are in line with the increasing support for marijuana legalization in the

[37]The negative coefficient can perhaps be due to already-established student connections in early treated states or ease of marijuana access in these early mover states.

[38]The utility gain from consumption amenities will converge to zero as wages and rents on either side of the policy offset each other. For instance, higher enrollment on one side may increase the rent while decreasing it on the other side.

US, as indicated by a 68% approval rate according to Gallup (2021). Additionally, the results align with the recent literature which finds that some migrants view marijuana legalization as a favorable amenity (Zambiasi & Stillman, 2020).

Although my findings indicate that RML has no significant effect on overall academic performance (degree completion), it also raises intriguing questions about its impact on other aspects of student behaviors. Additional research is needed to investigate how this policy affects students' choice of majors, including differentiating between STEM and non-STEM disciplines. Understanding this aspect could shed light on potential shifts in educational and career preferences and trajectories. Moreover, it is important to investigate in future studies how RML may contribute to the segregation of low-income or low-performing students (Chetty, Friedman, Saez, Turner, & Yagan, 2020a).[39] This may have further implications for peer effects within educational institutions.

In addition, the current results highlight the dominant impact of early RML adopters and predict null effects for future adopters. As more states legalize marijuana for recreational use and more post-policy data becomes available, future research can explore the long-term consequences of the policy on college enrollment. Another avenue for research may examine all potential externalities and spillovers (Cheng et al., 2018; Choi, Dave, & Sabia, 2019; Dave et al., 2023; Dragone, Prarolo, Vanin, & Zanella, 2019; Sabia et al., 2021) to conduct a comprehensive cost-benefit analysis of RML from a social welfare perspective. Understanding the broader implications of RML can help policymakers make informed decisions and assess the overall societal impact of marijuana legalization.

[39] To more effectively address these policy considerations, obtaining student-level data is essential. This necessity could guide future research endeavors in exploring this aspect more comprehensively.

Chapter 3

What is the Role of Sports at Elite Universities in the US?

College athletics is a defining feature of American higher education, distinguishing it from university systems around the world. Intercollegiate sports have expanded substantially since their origins in the nineteenth century, evolving from student-organized activities into professionally managed enterprises overseen by the National Collegiate Athletic Association (NCAA). While much attention has focused on revenue sports like football and basketball at large public universities, less is known about the role of athletics at elite private institutions. These universities, despite their small student bodies, often field more athletic teams and enroll proportionally more student-athletes than their larger public counterparts.

This paper examines athletics at elite universities in the United States, with a focus on how sports offerings shape the socioeconomic composition of these institutions. Using comprehensive data on NCAA programs and a novel dataset of athlete backgrounds, we document several striking patterns. First, elite private universities allocate a disproportionate share of their scarce enrollment capacity to varsity athletes. While large public universities typically have 1-2% of undergraduates participating in varsity athletics, this figure reaches 20-25% at some elite private institutions. Second, athletes at these universities–particularly in sports like squash, sailing, and fencing–come from substantially more advantaged backgrounds than the general student population. Finally, using a structural model of university decision-making, we show that these patterns reflect underlying institutional preferences, with Division 1 schools exhibiting stronger preferences for athletes from higher-income backgrounds.

Understanding the role of athletics at elite universities is particularly important given the growing competition for admission to these institutions. Elite universities offer substantial returns to their graduates, both in terms of earnings and access to leadership positions (Chetty, Deming, & Friedman, 2023). As seats become increasingly scarce, how universities allocate their limited enrollment capacity has significant implications for economic mobility and the efficient allocation of educational opportunity. If a substantial portion of seats is reserved for athletes from advantaged backgrounds, this may reduce opportunities for lower-income students with strong academic credentials.

Our research builds on several strands of literature. First, work on college admissions preferences has documented substantial advantages for recruited athletes (Bowen & Levin, 2003; Espenshade & Chung, 2005). Arcidiacono, Kinsler, and Ransom (2022b) find that recruited athletes at Harvard receive a substantial admissions boost, comparable to that of legacy students and children of donors. Second, research on the socioeconomic backgrounds of youth athletes has shown that participation in organized sports–particularly sports like tennis, swimming, and lacrosse–is strongly correlated with family income (Pennington, 2019; Ransom & Ransom, 2018). Third, studies on the economics of higher education have examined how universities compete for students and resources (Epple, Romano, Sarpça, & Sieg, 2017; Hoxby, 1997), though this work has generally not focused on the role of athletics in this competition.

We make several contributions to this literature. First, we provide the most comprehensive documentation to date of athlete representation at American universities, showing how this varies across institutions and athletic conferences. Second, we assemble novel data on the socioeconomic backgrounds of athletes using publicly available information about their high schools, allowing us to examine how athlete backgrounds vary across sports and institutions. Third, we develop and estimate a structural model of university decision-making that reveals the preferences driving observed patterns in sports offerings.

Our findings have important implications for policy debates about educational access and opportunity. As universities, policymakers, and courts reconsider admissions practices in light of recent Supreme Court decisions on affirmative action, our research suggests that athletic recruitment warrants similar scrutiny as legacy preferences. Both practices may reduce opportunities for disadvantaged students, potentially limiting the role of elite universities in promoting socioeconomic mobility. Using our structural model, we can simulate the effects of policies that would restrict the proportion of athletes at universities, providing insights into how such reforms might affect student body composition and institutional outcomes.

The remainder of the paper proceeds as follows. Section 2 describes our data on universities, sports programs, and individual athletes. Section 3 presents descriptive evidence on athlete representation and socioeconomic backgrounds. Section 4 develops our structural model of university decision-making. Section 5 presents estimation results from the model. Section 6 concludes.

3.1 Related Literature

Our work intersects with several strands of research on higher education, sports participation, and economic mobility. We highlight three areas most relevant to understanding the role of athletics at elite universities.

3.1.1 Socioeconomic Mobility and Elite Higher Education

A growing body of research documents the role of elite universities in shaping socioeconomic mobility. Chetty, Friedman, Saez, Turner, and Yagan (2020b) show that children from high-income families are 77 times more likely to attend an Ivy League college than children from low-income families. This disparity matters for long-term outcomes: Chetty et al. (2023) find that elite universities have large effects on students' likelihood of reaching the top 1% of the income distribution and working at prestigious firms. Similarly, Brint, German, Anderson-Natale, Shuker, and Wang (2020) demonstrate that elite universities have an outsized effect on placement into leadership positions across various sectors of American society.

While some earlier research suggested minimal economic returns to attending elite institutions (Dale & Krueger, 2014), more recent work using larger samples and administrative data has documented substantial heterogeneity in returns. Specifically, elite universities appear to have particularly large effects for students from disadvantaged backgrounds (Chetty et al., 2023) and for outcomes in the tails of the distribution. These findings highlight the importance of understanding how scarce seats at elite institutions are allocated.

3.1.2 Admissions Advantages and Athletic Recruitment

Several studies have documented substantial admissions advantages for recruited athletes at selective universities. Bowen and Levin (2003), in their influential book *Reclaiming the Game*, find that recruited athletes at selective institutions have substantially lower academic credentials than their peers yet receive significant preferential treatment in admissions. Espenshade and Chung (2005); Espenshade, Chung, and Walling (2004) quantify this advantage, finding that being a recruited athlete is equivalent to adding 200 points to an applicant's SAT score at elite private universities.

More recently, Arcidiacono et al. (2022b) analyze internal admissions data from Harvard, finding that recruited athletes receive an admissions boost comparable to legacy students and children of donors. Their analysis shows that these preferences effectively function as quotas, reserving a substantial portion of seats for these preferred groups. Arcidiacono, Kinsler, and Ransom (2022a) extend this work to examine racial disparities within preferred admissions categories, finding that white applicants disproportionately benefit from athletic and legacy preferences.

3.1.3 Socioeconomic Selection in Youth Sports

The socioeconomic composition of college athletes is shaped by patterns of participation in youth sports. Ransom and Ransom (2018) document substantial income gradients in youth sports participation, with children from higher-income families more likely to participate in organized sports and to specialize in particular sports. This gradient is especially pronounced for sports that have high equipment costs, require specialized facilities, or lack accessibility in public schools and community programs.

Jayakumar and Page (2021) examine racial and socioeconomic segregation across different sports, finding that sports like lacrosse, swimming, and tennis are disproportionately played by white and affluent youth. Pennington (2019) highlight how these patterns persist into college athletics, with certain sports at elite universities drawing primarily from private schools and affluent communities. These selection effects are important for understanding how athletics shape the socioeconomic composition of elite universities.

3.1.4 Higher Education Market Structure

Our structural approach to university decision-making builds on an established literature on higher education markets. Hoxby (1997) pioneered the analysis of competition between universities, showing how increased competition affected pricing and expenditure decisions. Epple, Romano, and Sieg (2006) developed equilibrium models of the higher education market, incorporating student sorting based on ability and income along with university optimization decisions.

More recent work by Jacob, McCall, and Stange (2018b) examines how universities compete through amenities, including sports facilities and programs. They find that students value consumption amenities, including athletics, and that universities strategically invest in these amenities to attract students. Garthwaite, Keener, Notowidigdo, and Ozminkowski (2020) analyze how athletic success, particularly in revenue sports, affects university outcomes beyond athletics. This literature provides the theoretical foundation for our model of university decision-making regarding sports offerings.

3.1.5 Economic Returns to Athletics for Universities

Universities invest in athletics for reasons beyond student preferences. Several studies examine how athletics affect alumni engagement and donations. Meer and Rosen (2009) find that former varsity athletes are more likely to donate to their alma mater, particularly when their teams were successful during their playing days. M. L. Anderson (2017) shows that athletic success leads to larger donations in the long run. These patterns help explain why universities might maintain large athletic programs despite their costs.

Additionally, successful athletics programs may increase applications and improve institutional visibility. D. G. Pope and Pope (2009, 2014) document that athletic success, particularly in men's basketball and football, increases applications to universities. This effect, known as the "Flutie Effect," allows universities to either increase enrollment or become more selective, potentially improving student quality. These benefits must be weighed against the opportunity costs of allocating scarce seats to recruited athletes with potentially lower academic credentials.

3.2 Data

Our analysis combines data from multiple sources to create a comprehensive dataset at both the university and athlete levels. This approach allows us to examine the market structure of college athletics while also investigating the socioeconomic backgrounds of student-athletes across institutions and sports.[1]

3.2.1 University-level Data

Our university-level dataset includes information on all 1,100 NCAA Division I, II, and III institutions for the 2019–2020 academic year. We collect general institutional characteristics from the Integrated Postsecondary Educational Data System (IPEDS), accessed through the R package `rscorecard` (Skinner, 2023). These data include undergraduate enrollment, student demographics, graduation rates, and financial information.

Information on the number of sports teams and student-athletes at each institution comes from CollegeFactual.com. We supplement these with data on athletic championships from the NCAA website and university rankings from Forbes.com. For measures of socioeconomic mobility, we utilize data from the Opportunity Insights project (Chetty et al., 2020b), which provides institution-level statistics on parental income distributions and intergenerational mobility outcomes for students.

To understand the financial dimensions of college sports, we incorporate data from the Equity in Athletics Data Analysis (EADA) database, which contains detailed information on

[1]Appendix Table C.3 provides further details about the data sources.

athletic revenues and expenses by institution and sport. These financial data allow us to assess the resource allocation within athletics departments and across different types of institutions.

3.2.2 Athlete-level Data

The core of our athlete-level dataset comes from athletic rosters published on university websites for the 2019–2020 academic year. We collect this information for all NCAA Divisions' institutions. From these rosters, we extract each athlete's high school or previous institution, which serves as our primary indicator of socioeconomic background.

We link each high school to demographic and socioeconomic characteristics using data from the Common Core of Data (CCD) for public schools and the Private School Survey (PSS) for private institutions, both administered by the National Center for Education Statistics (NCES).[2] These datasets provide information on school enrollment, racial composition, student-to-faculty ratios, and whether the school is public or private.

To develop more detailed measures of socioeconomic status, we match each high school's ZIP code to tax data from the Internal Revenue Service (IRS) and demographic information from the American Community Survey (ACS).[3] The ZIP code-level variables include average income, education levels, racial composition, and income distribution metrics.

In total, our athlete-level dataset contains information on 400,100 student-athletes enrolled at 1,100 universities during the 2019–2020 academic year. For each athlete, we observe the university attended, their sport, and the characteristics of their high school and its surrounding community. We choose this time period specifically because it predates both the COVID-19 pandemic and changes to name, image, and likeness (NIL) rules that have significantly altered the landscape of college athletics (Supreme Court of the United States, 2021).

[2]We access the CCD data via the `educationdata` R package (Tyagi, Ueyama, & The Urban Institute, 2022). The PSS data come directly from the NCES website.

[3]We access the ACS data using the `tidycensus` R package (Walker, Herman, & Eberwein, 2023). The IRS data come directly from the IRS website.

3.3 Descriptives

Our descriptive analysis reveals striking patterns in the distribution of athletes across universities and the socioeconomic background of these athletes. These patterns provide important context for understanding the market structure of college athletics and its implications for educational access.

3.3.1 Market Structure of College Athletics

Our descriptive analysis begins by examining the distribution of athletes across NCAA institutions. Figure 1 presents this distribution at both the university and conference levels, revealing a clear pattern of disproportionate athlete representation at elite private institutions.

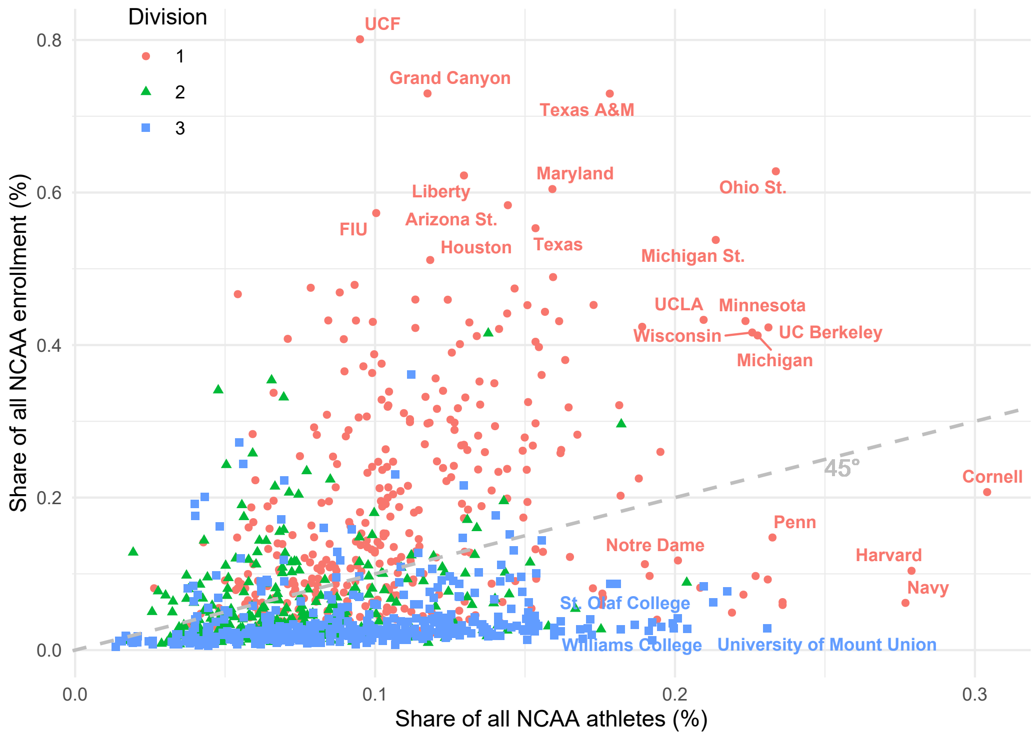
Panel (a) displays the relationship between each university's share of total NCAA enrollment and their share of NCAA athletes. Many elite private universities appear above the 45-degree line, indicating that they enroll a disproportionately high percentage of athletes relative to their overall student populations. For example, Cornell University accounts for over 0.2% of total enrollment in NCAA-affiliated universities, yet accounts for over 0.3% of all athletes at such schools.

This pattern becomes even more pronounced when we aggregate universities by athletic conference, as shown in Panel (b). Elite private university conferences (Ivy League, NESCAC, Patriot League) consistently appear in the upper-left portion of the graph, indicating a much higher athlete-to-student ratio than conferences dominated by large public universities.

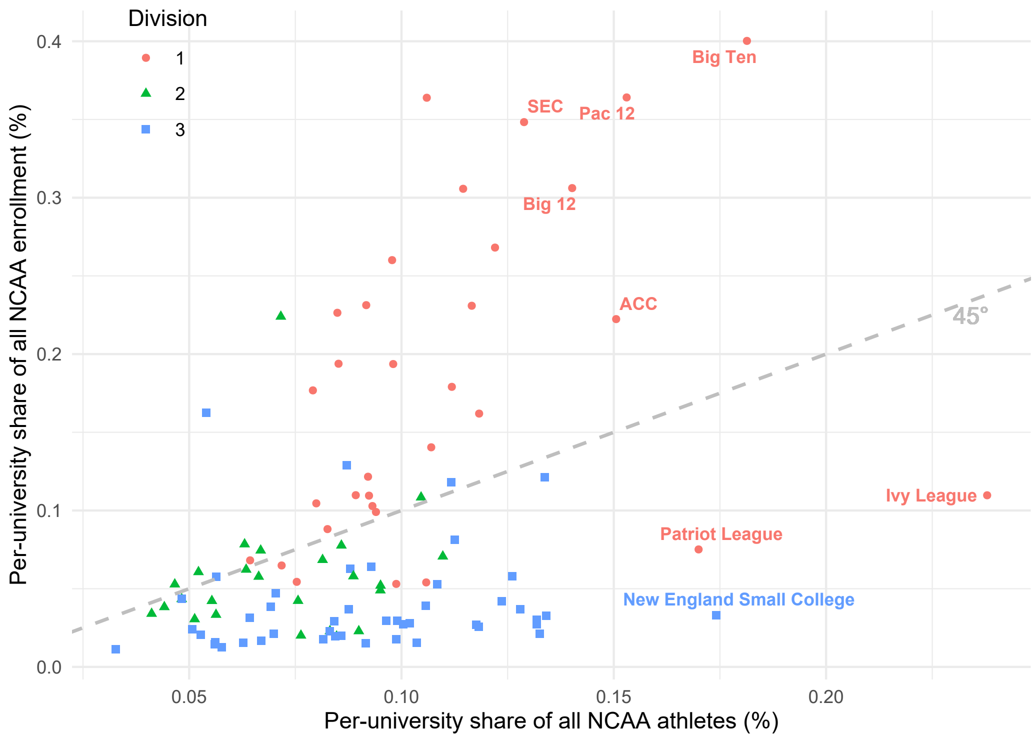
Table 3.1 provides summary statistics for our university sample. The average NCAA institution fields 15.8 sports teams and has an athlete population that comprises 15% of the undergraduate student body. However, this average masks substantial heterogeneity. At elite private institutions, athletes often constitute 20-25% of the student body, while at large public universities, this percentage typically ranges from 1-2%. For instance, at Dartmouth College, 22.6% of undergraduates are varsity athletes, compared to just 1.1% at Texas A&M University.

Figure 1: Share of NCAA Athletes vs. Share of NCAA Enrollment (2019–2020)

(a) By Individual University



(b) By Athletic Conference



NOTES: Panel (a) plots each NCAA university's share of total enrollment against its share of all NCAA athletes. Panel (b) plots each NCAA conference's share of total enrollment against its share of all NCAA athletes, with both shares divided by the number of universities in the conference to produce per-university estimates. In both panels, points above the 45-degree line represent institutions or conferences with disproportionately high athlete representation. Elite private universities and their conferences (Ivy League, NESCAC, Patriot League) consistently appear above this line, while large public universities typically fall below it.

Table 3.1: Summary Statistics for Universities and Colleges

	N	Mean	SD	Min	Max
Number of sports	1067	15.81	4.34	4.00	35.00
Share of athletes	1067	0.15	0.13	0.00	0.93
Number of athletes	1067	418.10	197.85	60.00	2297.00
Number of female athletes	1067	181.56	105.14	24.00	2018.00
Number of male athletes	1067	236.54	113.51	0.00	726.00
Median parental income of students (in $1k)	1024	93.80	29.27	28.80	226.70
Pecent of parents in the top 1% of the national income distribution	1024	2.53	3.64	0.00	21.91
Pecent of parents in the bottom 20% of the national income distribution	1024	8.06	5.94	1.12	45.46
Percent of student whose earnings in top 20% and whose parents' in bottom 20%	1024	1.90	1.36	0.03	12.94
Percent of student whose earnings in top 1% and whose parents' in bottom 20%	1024	0.08	0.12	0.00	1.25
Undergraduate enrollments	1067	6966.45	8716.73	294.00	80 170.00
Six year graduation rate	1067	0.60	0.17	0.12	0.98
Percent of pell grant recipient	1064	0.34	0.14	0.09	0.89
Admission rate	1034	0.66	0.20	0.04	1.00
Average SAT score	907	1160.39	127.18	855.00	1566.00
Federal loan rate	1064	0.55	0.17	0.02	0.93
Pecent full time faculty	1067	0.70	0.22	0.05	1.00
Share of women students	1067	0.56	0.10	0.11	1.00
Share of men students	1067	0.44	0.10	0.00	0.89
Share of white students	1067	0.58	0.21	0.00	0.93
Share of African Amr. students	1067	0.13	0.17	0.00	0.98
Share of Latin Amr. students	1067	0.12	0.12	0.00	1.00
Share of Asian Amr. students	1067	0.05	0.06	0.00	0.40
Share of International students	1067	0.04	0.04	0.00	0.33

This pattern of disproportionate athlete representation at elite private universities is particularly noteworthy given the socioeconomic profiles of these institutions. As shown in Table 3.1, the average NCAA institution has a median parental income of $93,800, with 2.53% of students coming from the top 1% of the income distribution. However, at elite private universities, these figures are substantially higher.

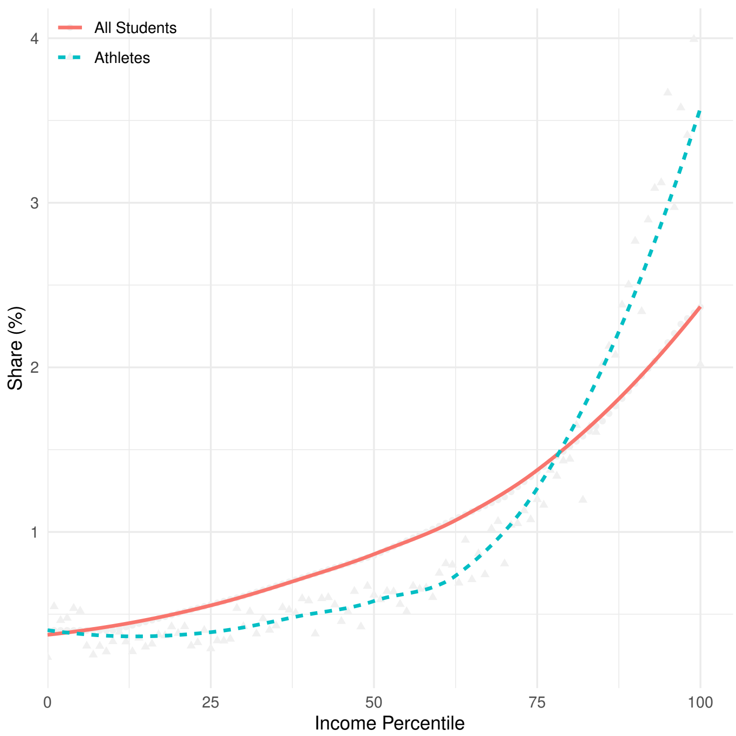
3.3.2 Athlete Socioeconomic Backgrounds

Figure 2 compares the family income distributions of athletes to the general student population across all NCAA institutions. The graph reveals that athletes come from higher-income backgrounds than the typical student, with the difference being particularly pronounced at the upper end of the income distribution. While both distributions show substantial right-skewness, the athlete curve lies consistently above the general student curve across most of the income range.

Table 3.1 provides insights into the characteristics of student-athletes in our sample. On average, 37% of athletes in our sample attended a private high school, compared to approximately 10% of all high school students nationally. This suggests that athletes are more than three times as likely to have attended private high schools as the typical American student. Additionally, 17% of athletes in our sample are international students, indicating the global reach of collegiate athletic recruitment.

The socioeconomic advantage of athletes is further evidenced by the characteristics of their high schools' ZIP codes. The average ZIP code salary for athletes' high schools is $83,130, substantially above the national average. Additionally, the average athlete's high school ZIP code has higher educational attainment than the national average, with 24.5% of residents holding bachelor's degrees and 17.7% holding advanced degrees.

Figure 2: Income Background of Athletes vs. All Students



NOTES: The figure shows the predicted share of family income rank for all students (red) and athletes (blue dashed) at 1051 NCAA-affiliated universities as of the 2019-2020 academic year. Curves are fit using local linear regression (LOESS).

3.3.3 Heterogeneity by Sport and Institution Type

Our data reveal substantial heterogeneity in the socioeconomic backgrounds of athletes across different sports and institution types. To examine these patterns in detail, Figure 3 presents a multi-panel analysis of income distributions.

Panel (a) of Figure 3 shows income distributions by sport, revealing substantial stratification. "Upper-crust" sports like squash, sailing, and fencing draw disproportionately from the highest income deciles, with average ZIP code incomes often falling in the top 20% of the national distribution. These sports tend to have high barriers to entry in terms of facilities, equipment costs, and limited geographic accessibility. By contrast, sports like football, basketball, and wrestling show more balanced income distributions, though still skewed upward relative to the general population.

Panel (b) displays income distributions by athletic conference, highlighting institutional patterns. Elite conferences like the Ivy League and NESCAC show markedly higher income distributions than conferences composed primarily of public institutions such as the Big Ten or SEC. This pattern likely reflects both the higher overall socioeconomic profile of elite private universities and their disproportionate investment in sports that draw from higher-income backgrounds.

Panels (c) and (d) provide additional dimensions of heterogeneity. Panel (c) reveals that Division III athletes come from higher-income backgrounds than those in Divisions I or II, despite the absence of athletic scholarships in Division III. This counterintuitive finding suggests that the socioeconomic composition of athletes is driven more by institutional selectivity and sport offerings than by financial aid considerations. Panel (d) confirms the substantial gap between private and public institutions, with private school athletes coming from ZIP codes with significantly higher average incomes.

Panel (a) of Appendix Figure C.1 reveals similar striking differences in educational backgrounds across sports. "Upper-crust" sports like squash draw disproportionately from private high schools, with over 60% of athletes in these sports coming from private institutions. By contrast, wrestling, football, and basketball show much lower private school representation. International recruitment also varies dramatically, with sports like tennis drawing heavily from international talent pools, while football and wrestling remain predominantly domestic.

Panel (b) illustrates how these patterns vary across athletic conferences. Elite conferences like the Ivy League and NESCAC recruit substantially higher percentages of private school graduates (over 40%) compared to power conferences like the SEC and Big Ten (approximately 20-25%). The gap is particularly striking given that elite private universities have significantly higher tuition costs and typically offer less generous athletic scholarships than large public universities.

Panels (c) and (d) explore how these patterns vary even within specific sports categories. Panel (c) focuses on football and basketball players, revealing that even in these more accessible sports, elite conferences recruit a substantially higher percentage of private school graduates. Panel (d) examines cross-country runners and soccer players, showing even more pronounced differences in private school representation across conference types and particularly high rates of international recruitment in certain conferences.

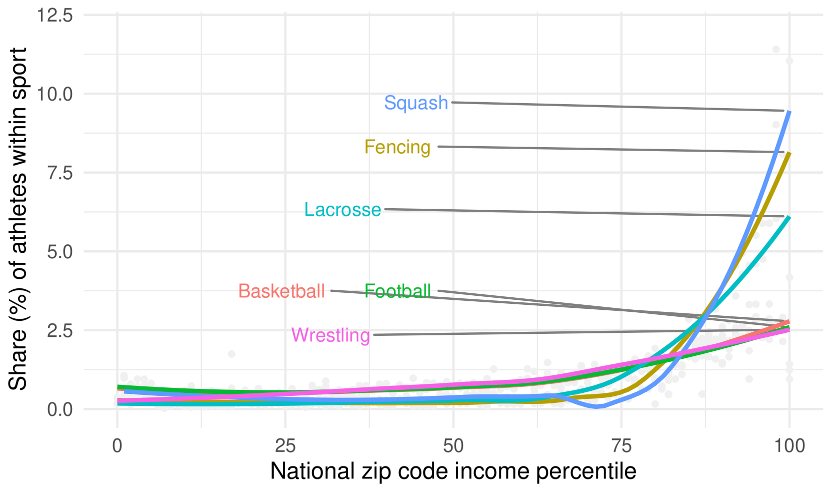
These patterns of educational background reinforce the income-based findings and suggest that athletics at elite institutions may function as a channel for maintaining socioeconomic advantage rather than promoting diversity. The substantial overrepresentation of private school graduates–institutions that themselves are disproportionately attended by higher-income students–indicates multiple layers of socioeconomic selection in the pipeline to elite college athletics.

Table C.1 provides additional context by comparing the characteristics of high schools that send athletes to NCAA institutions with those that do not. These patterns further underscore the socioeconomic gradients in athletic recruitment and participation across institution types and sports.

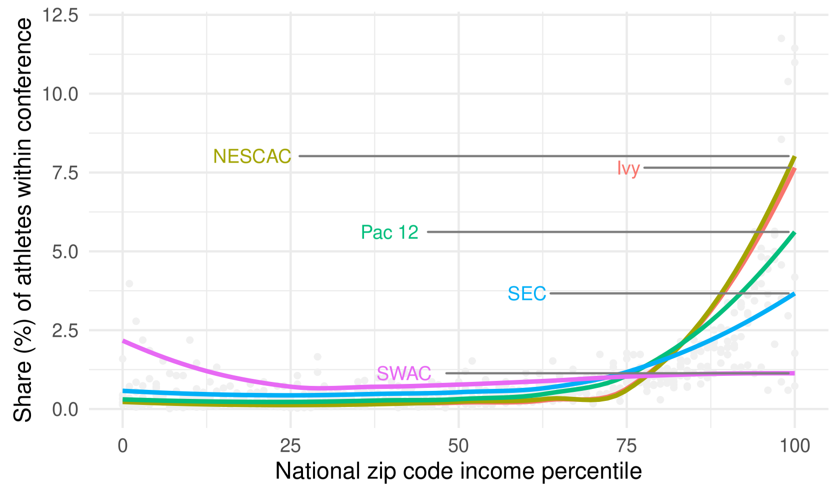
Similar patterns hold for public high schools, with those sending athletes to NCAA institutions being substantially larger (1,022 students) than those that do not (276 students). These schools have higher proportions of white students (58.7% vs. 50.4%) and lower proportions of Hispanic American students (18.4% vs. 24.6%). They also come from ZIP codes with higher average salaries.

This differential patterns suggest selection effects in athletic recruitment that favor larger high schools, likely due to more robust athletic programs, better facilities, and more competitive sports environments. The data also reveal socioeconomic and racial gradients in which high schools send athletes to NCAA institutions, with schools in higher-income areas and with lower minority enrollments more likely to place students in college athletics.

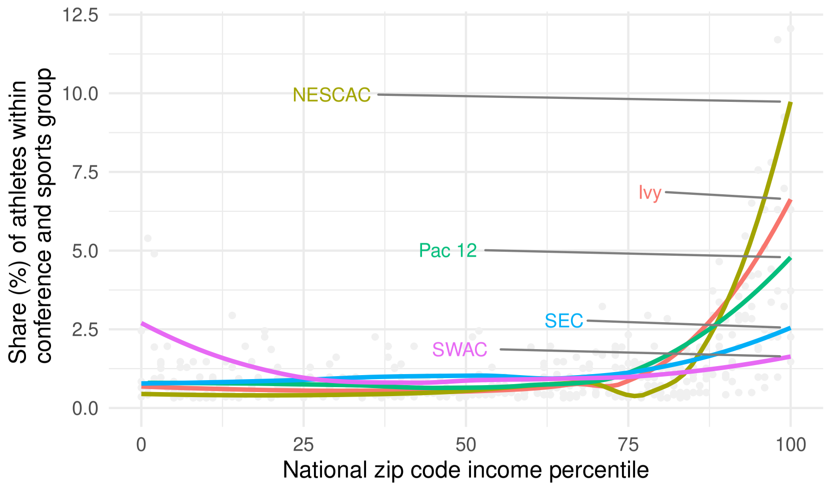
Figure 3: Income Distributions by Sport and Institution Type



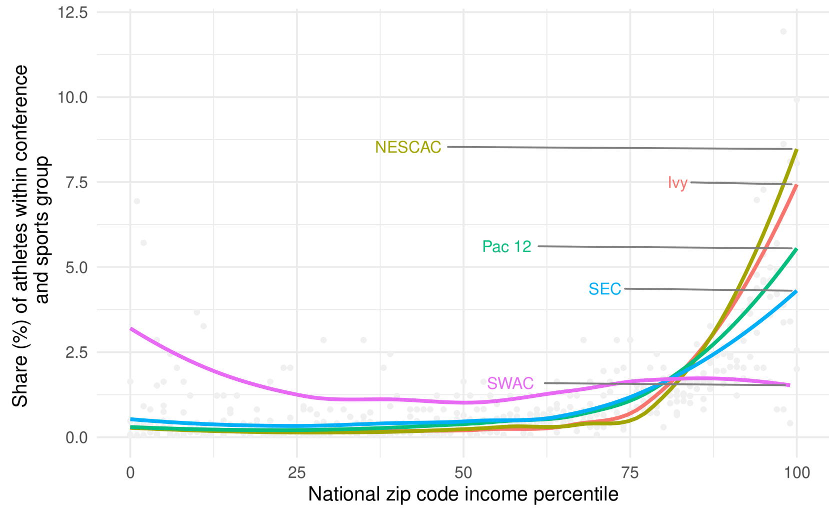
(a) By sport, all athletes



(b) By athletic conference, all athletes



(c) By athletic conference, football and basketball players only



(d) By athletic conference, cross-country and soccer players only

NOTES.—The figures plot the probability mass of athletes at each income percentile at different levels of aggregation, along with local linear regression lines (LOESS). The rank is calculated using the zip code of the athlete's secondary school. Zip codes are ranked according to the average income of the tax returns filed from that zip code in 2017.

3.3.4 Implications for Educational Access

These descriptive patterns have important implications for educational access and opportunity. Elite private universities–which offer some of the highest returns to education and strongest pathways to leadership positions in American society (Chetty et al., 2023)–allocate a disproportionate share of their limited enrollment capacity to athletes who come from more advantaged backgrounds than the general student population.

This pattern suggests that athletics may function as a channel through which seats at elite institutions are reserved for students from higher socioeconomic backgrounds, potentially at the expense of lower-income students with stronger academic credentials. The concentration of this effect in "upper-crust" sports like squash, sailing, and fencing–which are disproportionately offered at elite private universities–further heightens concerns about equity and efficiency in the allocation of educational opportunity.

In the sections that follow, we develop and estimate a structural model of university decision-making about sports offerings. This model allows us to quantify the underlying preferences that drive the patterns observed in our descriptive analysis and to simulate the effects of counterfactual policies that would restrict the proportion of athletes at universities.

3.4 Model of College Sports Choices

To understand how universities determine which sports to offer, we develop a structural model of university decision-making. This approach allows us to identify the underlying preferences that drive observed sports portfolios and to conduct counterfactual policy simulations.

3.4.1 University Choice Problem

We model each university i as choosing a bundle of sports j from a finite set of alternatives $\mathcal{J}$ to maximize utility. The utility that university i derives from choosing bundle j is specified as:

$$U_{ij} = \beta_1 \log(R_{ij}) - \beta_1 \log(E_{ij}) + \beta_2 Z_{ij} + \gamma X_j + \varepsilon_{ij} \quad (3.1)$$

where:

- R_{ij} represents the expected revenues that university i would receive if it were to select bundle j
- E_{ij} represents the expected expenses that university i would incur if it were to select bundle j
- Z_{ij} is a vector containing the expected socioeconomic characteristics of athletes if university i were to select bundle j
- X_j is a vector of bundle-specific characteristics
- ε_{ij} is an error term with a Type I extreme value distribution

The first two terms capture the financial implications of offering each sports bundle. By constraining the coefficients on $\log(R_{ij})$ and $\log(E_{ij})$ to be equal and opposite, we impose the restriction that universities care about the net financial return of their athletics programs. The vector Z_{ij} contains measures of the socioeconomic status of athletes, including average ZIP code income of high schools attended, the percentage attending private high schools, and the percentage of international students. The vector X_j includes indicators for whether the bundle contains football and categorical variables for the intensity of men's and women's sports across different categories.

3.4.2 Defining the Choice Set

A key challenge in estimating this model is defining the choice set $\mathcal{J}$. With 77 possible NCAA sports, the raw number of potential combinations (over 2^{77}) is computationally intractable. We address this challenge by aggregating sports into meaningful categories and using observed combinations to define a feasible choice set.

First, we categorize sports along three dimensions:

- **Standard sports:** Widely offered across universities (e.g., basketball, baseball/softball)
- **Regional sports:** Concentrated in specific geographic regions (e.g., lacrosse, ice hockey)

- **Niche sports:** Less commonly offered, often requiring specialized facilities (e.g., squash, fencing)

We treat football separately due to its unique financial position and cultural significance. For each category, we define intensity levels (low, medium, high) based on the number of sports offered within the category. Using hierarchical agglomerative clustering on the observed combinations of these intensities, we identify 74 representative bundles that form our final choice set.

3.4.3 Prediction of Bundle Characteristics

For each university-bundle pair, we need to predict the revenues, expenses, and athlete characteristics that would result if the university selected that bundle. We estimate these values using the following approach:

$$Y_{ij} = \delta_0 + \delta_1 W_i + \delta_2 B_j + \delta_3 (W_i \times B_j) + \xi_{ij} \tag{3.2}$$

where Y_{ij} represents the outcome of interest (revenues, expenses, or SES measures), W_i is a vector of university characteristics, B_j is a vector of bundle characteristics, and ξ_{ij} is an error term. We estimate this equation using observed university-bundle pairs and use the resulting coefficients to predict values for all counterfactual pairs.

3.4.4 Estimation Strategy

Given the Type I extreme value distribution for ε_{ij}, the probability that university i chooses bundle j takes the familiar logit form:

$$\Pr(i \text{ chooses } j) = \frac{\exp(V_{ij})}{\sum_{k \in \mathcal{J}} \exp(V_{ik})} \tag{3.3}$$

where V_{ij} represents the deterministic component of utility. We estimate the model using maximum likelihood, with the log-likelihood function given by:

$$\mathcal{L}(\beta, \gamma) = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij} \log \Pr(i \text{ chooses } j) \tag{3.4}$$

where d_{ij} is an indicator equal to 1 if university i chooses bundle j and 0 otherwise.

To account for preference heterogeneity across universities, we also estimate a mixed logit specification that allows random coefficients for key variables:

$$U_{ij} = \beta_{1i} \log(R_{ij}) - \beta_{1i} \log(E_{ij}) + \beta_{2i} Z_{ij} + \gamma X_j + \varepsilon_{ij} \tag{3.5}$$

where β_{1i} and β_{2i} are university-specific coefficients that follow normal distributions with means and standard deviations to be estimated.

Our identification strategy relies on variation in university characteristics and the sports bundles they choose. The key identifying assumption is that, conditional on the observed covariates, the error term ε_{ij} is uncorrelated with the explanatory variables.

The constraint that the coefficients on $\log(R_{ij})$ and $\log(E_{ij})$ are equal and opposite is motivated by the economic logic that universities should care about net financial returns. This constraint also helps identification by reducing the number of parameters to estimate.

Variation across NCAA divisions provides additional identification power. We exploit this by estimating the model separately for Division 1, Division 2, and Division 3 institutions, allowing us to examine how preferences vary with competitive level and institutional mission.

The results from estimating this model provide insights into how universities value different aspects of their athletics programs, including financial returns, socioeconomic composition, and sport-specific preferences.

3.5 Results

Our empirical analysis provides evidence that universities' sports portfolio choices are driven by a combination of financial motivations and socioeconomic considerations. Table 3.2 presents the main estimation results from our conditional logit model across all universities and separately by NCAA division.

Table 3.2: Conditional Logit Estimates of University Athletics Program Choices

	Dependent Variable: Chosen Sports Bundle			
	All	Division 1	Division 2	Division 3
Variables	(1)	(2)	(3)	(4)
Financial Variables				
log(Predicted Revenues)	15.750***	47.727***	28.140***	-9.373*
	(3.798)	(7.730)	(8.882)	(5.614)
log(Predicted Expenses)	-15.750***	-47.727***	-28.140***	9.373*
	(3.798)	(7.730)	(8.882)	(5.614)
Socioeconomic Status (SES) Variables				
Predicted ZIP Code Salary	0.114	1.146***	0.229	-0.439**
	(0.128)	(0.230)	(0.325)	(0.206)
Predicted % International	0.079***	0.092*	0.230***	0.127
	(0.029)	(0.050)	(0.055)	(0.081)
Predicted % Private HS	-0.559***	-0.894***	-0.236*	-0.624***
	(0.079)	(0.153)	(0.129)	(0.138)
Bundle Characteristics				
Has Football	-0.251	0.486	-0.822**	-0.074
	(0.175)	(0.377)	(0.358)	(0.266)
Men's Standards Category 2	-0.205**	-1.293***	-0.743***	0.803***
	(0.104)	(0.196)	(0.248)	(0.161)
Men's Standards Category 3	-0.164	-1.264***	-0.222	0.682***
	(0.111)	(0.203)	(0.258)	(0.175)
Women's Standards Category 2	0.174*	0.743***	-0.215	0.109
	(0.090)	(0.164)	(0.182)	(0.141)
Women's Standards Category 3	0.231*	1.475***	0.333	-0.481**
	(0.138)	(0.270)	(0.337)	(0.202)
Men's Regional Category 2	-0.301***	-1.128***	-0.332*	0.655***
	(0.087)	(0.154)	(0.172)	(0.159)
Men's Regional Category 3	-0.252**	-1.317***	-0.779***	0.924***
	(0.100)	(0.182)	(0.220)	(0.178)
Women's Regional Category 2	-0.303***	-0.681***	-0.148	-0.224
	(0.097)	(0.173)	(0.194)	(0.164)
Women's Regional Category 3	-0.474***	-0.318*	-1.056***	-0.395**
	(0.097)	(0.168)	(0.229)	(0.157)
Men's Niche Category 2	0.074	0.037	0.168	-0.046
	(0.077)	(0.137)	(0.167)	(0.120)
Women's Niche Category 2	-0.221***	-0.173	-0.480***	-0.022
	(0.075)	(0.136)	(0.154)	(0.119)
Women's Niche Category 3	-0.392***	-0.621***	-0.210	-0.365**
	(0.099)	(0.184)	(0.212)	(0.151)
Observations	74,740	25,308	20,498	28,934
Number of Universities	1,010	342	277	391
Log Likelihood	-4,281.15	-1,367.57	-1,091.82	-1,615.52

Notes: Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. All models are conditional logit specifications with the constraint that coefficients on log(Revenues) and log(Expenses) are equal and opposite. The choice set consists of 74 alternative sports bundles. Division refers to NCAA competitive division. Categories for standards, regional, and niche sports are relative to Category 1 (baseline).

3.5.1 Financial Considerations

The results in Table 3.2 show that financial considerations play a significant role in universities' sports choices, but with substantial heterogeneity across NCAA divisions. The coefficient on log predicted revenues is positive and highly significant in the pooled sample (15.750, $p<0.01$), indicating that universities generally prefer sports bundles with higher revenue potential, all else equal. However, this preference varies dramatically by division.

Division 1 institutions display the strongest financial orientation, with a coefficient (47.727, $p<0.01$) that is more than three times the overall average and nearly 70% larger than the Division 2 coefficient (28.140, $p<0.01$). This finding aligns with the higher commercialization of Division 1 athletics and the revenue pressures faced by these institutions. Surprisingly, Division 3 schools exhibit a negative relationship between predicted revenues and bundle selection (-9.373, $p<0.1$), suggesting that these institutions may prioritize other factors in their athletics decisions.

3.5.2 Socioeconomic Background Preferences

One of our key research questions concerns the role of athlete socioeconomic background in university decision-making. Our estimates reveal complex patterns in this dimension. Looking at the ZIP code salary coefficient, we find that Division 1 institutions show a strong positive preference for athletes from higher-income areas (1.146, $p<0.01$), while Division 3 schools exhibit a negative preference (-0.439, $p<0.05$). This striking difference suggests that Division 1 programs, particularly at selective universities, may use athletics as a channel to attract students from higher socioeconomic backgrounds.

The percentage of international athletes has a small but positive coefficient in the overall sample (0.079, $p<0.01$), with the strongest preference in Division 2 (0.230, $p<0.01$). This could reflect strategic recruitment of international talent or efforts to enhance global visibility and prestige.

Contrary to what might be expected, we find negative coefficients on the percentage of private high school graduates across all divisions, with the strongest effect in Division 1 (-0.894, $p<0.01$). This result is surprising given the descriptive evidence that many athletes at elite institutions come from private schools. One possible explanation is that, conditional on ZIP code income, universities prefer diversity in educational backgrounds or that private school attendance may correlate with other unobserved factors that universities consider in their decision-making.

3.5.3 Sport-Specific Preferences

The model also reveals distinct patterns in universities' preferences for different types of sports. Football, despite its revenue prominence, does not show a significant coefficient in the overall sample or for Division 1 and 3 schools. Division 2, however, shows a negative preference for football (-0.822, $p<0.05$), potentially reflecting the financial challenges of maintaining football programs at this competitive level.

For men's standard sports (basketball, baseball, etc.), Division 1 shows strong negative preferences for higher categories (-1.293 for Category 2, $p<0.01$), while Division 3 shows strong positive preferences (0.803 for Category 2, $p<0.01$). This inverse pattern suggests fundamentally different objectives across divisions: Division 1 may favor focused excellence in fewer sports, while Division 3 values broader participation.

Women's standard sports show different patterns, with Division 1 exhibiting strong positive preferences for higher categories (1.475 for Category 3, $p<0.01$), potentially reflecting Title IX compliance strategies or different competitive considerations for women's athletics.

Regional sports show similar divergent patterns between Division 1 (negative coefficients) and Division 3 (positive coefficients for men's sports), while niche sports generally have negative coefficients across divisions, particularly for women's sports in higher categories.

3.5.4 Preference Heterogeneity

To explore potential heterogeneity in preferences across universities, we estimate a mixed logit model that allows for random parameters. Table 3.3 presents these results.

Table 3.3: Mixed Logit Estimates of University Athletics Program Choices

Variables	Coefficient
Mean Parameters	
Financial Returns	
log(Predicted Revenues)	17.101***
	(4.027)
log(Predicted Expenses)	-17.101***
	(4.027)
Socioeconomic Status (SES) Variables	
Predicted ZIP Code Salary	0.130
	(0.132)
Predicted % International	0.034
	(0.045)
Predicted % Private HS	-0.606***
	(0.088)
Bundle Characteristics	
Has Football	-0.270
	(0.184)
Men's Standards Category 2	-0.164
	(0.108)
Men's Standards Category 3	-0.162
	(0.115)
Women's Standards Category 2	0.212**
	(0.096)
Women's Standards Category 3	0.274*
	(0.143)
Men's Regional Category 2	-0.304***
	(0.088)
Men's Regional Category 3	-0.264***
	(0.102)
Women's Regional Category 2	-0.286***
	(0.098)
Women's Regional Category 3	-0.455***
	(0.098)
Men's Niche Category 2	0.077
	(0.079)
Women's Niche Category 2	-0.221***
	(0.077)
Women's Niche Category 3	-0.375***
	(0.101)
Standard Deviations of Random Parameters	
SD(log(Predicted Revenues))	0.013
	(0.731)
SD(log(Predicted Expenses))	1.236***
	(0.378)
SD(Predicted ZIP Code Salary)	0.013
	(0.502)
SD(Predicted % International)	0.212**
	(0.086)
SD(Predicted % Private HS)	0.209
	(0.172)
Observations	74,740
Number of Universities	1,010
Log Simulated-Likelihood	-4,278.34
Integration Points	715

Notes: Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The model is a mixed logit specification with the constraint that coefficients on log(Revenues) and log(Expenses) are equal and opposite. The random parameters follow normal distributions. The choice set consists of 74 alternative sports bundles.

The mixed logit results confirm the main findings from the conditional logit model while revealing important sources of heterogeneity. The mean parameter for log predicted revenues (17.101, $p<0.01$) is similar to the conditional logit estimate, but we find significant variation in universities' sensitivity to expenses (SD = 1.236, $p<0.01$). This suggests that while revenue considerations are consistently important across institutions, the weight placed on expenses varies substantially.

We also find significant heterogeneity in preferences for international students (SD = 0.212, $p<0.05$), indicating that some universities actively seek international talent while others do not prioritize this dimension. The mean preferences for specific sport bundles are generally consistent with the conditional logit results, with women's standards receiving positive coefficients and most regional sports receiving negative coefficients.

To provide clearer interpretation of these variables, Table C.2 presents descriptions of the key measures in our analysis.

3.5.5 Implications for Educational Access

These results have important implications for educational access and opportunity. The strong positive preference for higher ZIP code salaries among Division 1 schools suggests that athletics may function as a channel through which these institutions, particularly elite private universities, maintain a student body with higher socioeconomic status. Coupled with the descriptive evidence that athletes at these institutions come disproportionately from higher-income backgrounds, our findings suggest that athletic recruitment may reduce socioeconomic diversity at precisely the institutions with the greatest resources and potential for promoting mobility.

The estimated preference parameters also reveal how different types of institutions navigate the complex trade-offs between financial returns, prestige, and compositional objectives through their sports offerings. Division 1 institutions appear to optimize financial returns and socioeconomic composition, while Division 3 schools show patterns more consistent with broader participation goals and less emphasis on financial or socioeconomic considerations.

3.6 Conclusion

This paper has examined the role of sports at elite universities in the United States, with particular attention to how athletic programs shape the socioeconomic composition of these institutions. Our analysis, combining comprehensive university-level data with novel athlete-level information, reveals several important patterns about college sports and educational opportunity.

First, we document that elite private universities allocate a disproportionately high share of their limited enrollment capacity to varsity athletes. At some elite institutions, athletes constitute 20-25% of the undergraduate student body, compared to just 1-2% at large public universities. This pattern holds despite the fact that these elite institutions have much smaller overall enrollments, indicating a fundamentally different approach to the role of athletics in the educational mission.

Second, we find that student-athletes, particularly at elite private institutions, come from significantly more advantaged backgrounds than the general student population. Athletes are more likely to have attended private high schools, to come from high-income ZIP codes, and to have attended larger schools with lower proportions of minority students. This socioeconomic advantage is particularly pronounced in sports like squash, sailing, and fencing–sports that are disproportionately offered at elite private institutions.

Third, our structural model of university sports choices reveals that institutions have heterogeneous preferences regarding athletic programs. Division 1 schools, particularly elite private universities, show stronger preferences for athletes from higher-income backgrounds, while Division 3 institutions exhibit preferences more consistent with broader participation goals. These results suggest that athletics may function as a channel through which elite institutions maintain a student body with higher socioeconomic status.

These findings raise important questions about the allocation of educational opportunity in the United States. As universities face increasing scrutiny over admissions policies, our results suggest that athletic recruitment deserves the same level of critical examination as legacy preferences and other forms of non-academic consideration. The significant resources devoted to sports that primarily benefit higher-income students may run counter to educational goals of promoting socioeconomic mobility and allocating talent efficiently.

Our work has several limitations that suggest directions for future research. While our data provide a rich picture of athlete backgrounds, we lack direct measures of family income and instead rely on high school characteristics as proxies. Additionally, our analysis is primarily cross-sectional and cannot identify causal effects of athletic programs on institutional outcomes. Finally, our model, while informative about university preferences, relies on simplifying assumptions about the choice set and decision-making process.

An important extension of this research is to use our structural model to simulate the effects of counterfactual policies that would restrict the proportion of athletes at universities. In future work, we plan to examine policies that would cap the percentage of varsity athletes at each institution, with varying thresholds; eliminate or reduce specific categories of sports with the strongest socioeconomic gradients; and impose minimum socioeconomic diversity requirements for athletic programs. For each policy scenario, we will estimate the effects on the socioeconomic composition of the student body, institutional finances, gender equity in athletic participation, and the distribution of athletic opportunities across different types of institutions.

Preliminary analysis suggests that capping the percentage of athletes at elite universities would increase socioeconomic diversity, primarily through reductions in niche sports that draw from the highest income brackets. However, such policies would have complex distributional consequences across sports types and genders, particularly given Title IX requirements. These counterfactual exercises will provide valuable insights for policymakers, university administrators, and researchers interested in the intersection of higher education, athletics, and social mobility.

In the broader context of research on higher education, our findings contribute to understanding the mechanisms through which socioeconomic advantage is maintained across generations. The prominent role of athletics at elite institutions–institutions that provide substantial returns to their graduates in terms of earnings and leadership opportunities–suggests that reform of athletic recruitment policies may be an important lever for increasing socioeconomic mobility and diversity in American society. As the competition for seats at elite universities continues to intensify, understanding and addressing the structural factors that shape access to these institutions becomes increasingly important for ensuring that educational opportunities are allocated both efficiently and equitably.

Bibliography

Agan, A. Y., & Makowsky, M. D. (2023). The minimum wage, eitc, and criminal recidivism. Journal of Human Resources, 58(5), 1712–1751.

Ahn, T., Arcidiacono, P., Hopson, A., & Thomas, J. (2024, May). Equilibrium grading policies with implications for female interest in stem courses. Econometrica, 92(3), 849–880. Retrieved from https://doi.org/10.3982/ECTA17876 doi: 10.3982/ECTA17876

Altonji, J. G., Arcidiacono, P., & Maurel, A. (2016). The analysis of field choice in college and graduate school: Determinants and wage effects. In Handbook of the economics of education (Vol. 5, pp. 305–396). Elsevier.

Altonji, J. G., Blom, E., & Meghir, C. (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers. Annu. Rev. Econ., 4(1), 185–223.

Altonji, J. G., & Zimmerman, S. D. (2019). The costs of and net returns to college major. In C. M. Hoxby & K. Stange (Eds.), Productivity in higher education (pp. 133–176). Chicago: University of Chicago Press.

Anderson, D. M., Hansen, B., Rees, D. I., & Sabia, J. J. (2019). Association of marijuana laws with teen marijuana use: new estimates from the youth risk behavior surveys. JAMA pediatrics, 173(9), 879–881.

Anderson, M. L. (2017). The benefits of college athletic success: An application of the propensity score design. Review of Economics and Statistics, 99(1), 119–134. doi: 10.1162/REST_a_00589

Arcidiacono, P. (2004). Ability sorting and the returns to college major. Journal of econometrics, 121(1-2), 343–375.

Arcidiacono, P., Hotz, V. J., & Kang, S. (2012). Modeling college major choices using elicited measures of expectations and counterfactuals. Journal of Econometrics, 166(1), 3–16.

Arcidiacono, P., Kinsler, J., & Ransom, T. (2022a). Divergent: The time path of legacy and athlete admissions at harvard. Journal of Human Resources, Forthcoming. doi: 10.3368/jhr.0421-11641R1

Arcidiacono, P., Kinsler, J., & Ransom, T. (2022b). Legacy and athlete preferences at harvard. Journal of Labor Economics, 40(1), 133–156. doi: 10.1086/713744

Arria, A. M., Caldeira, K. M., Bugbee, B. A., Vincent, K. B., & O'Grady, K. E. (2015). The academic consequences of marijuana use during college. Psychology of Addictive Behaviors, 29(3), 564.

Autor, D. H., & Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the us labor market. American economic review, 103(5), 1553–1597.

Autor, D. H., Dorn, D., & Hanson, G. H. (2013). The china syndrome: Local labor market effects of import competition in the united states. American economic review, 103(6), 2121–2168.

Bacovic, M., Andrijasevic, Z., & Pejovic, B. (2022). Stem education and growth in europe. Journal of the Knowledge Economy, 13(3), 2348–2371.

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

Becker, G. S. (1964). Human capital. New York: Columbia University Press for the National Bureau of Economic Research.

Behrman, J. R., Constantine, J., Kletzer, L., McPherson, M., & Schapiro, M. (1996). The impact of college quality on wages: Are there differences among demographic groups? Williams Project on the Economics of Higher Education, Discussion Paper, 38.

Bell, J. D. (2008). The nuts and bolts of the higher education legislative appropriations process (Policy Brief). Boulder, Colo.: Western Interstate Commission for Higher Education.

Berger, M. C. (1988). Predicted future earnings and choice of college major. ILR Review, 41(3), 418–429.

Bettinger, E. P., & Baker, R. B. (2014). The effects of student coaching: An evaluation of a randomized experiment in student advising. Educational Evaluation and Policy Analysis, 36(1), 3–19.

Betts, J. R., & McFarland, L. L. (1995). Safe port in a storm: The impact of labor market conditions on community college enrollments. Journal of Human resources, 741–765.

Blake, C. D., Hess, J., & Thomas, D. K. (2023). Higher education: The impact of recreational marijuana on college applications. Contemporary economic policy., 1.

Blanchard, O. J., Katz, L. F., Hall, R. E., & Eichengreen, B. (1992). Regional evolutions. Brookings papers on economic activity, 1992(1), 1–75.

Borbely, D., Lenhart, O., Norris, J., & Romiti, A. (2022). Marijuana legalization and mental health (Tech. Rep.). IZA Discussion Papers.

Borusyak, K., Hull, P., & Jaravel, X. (2022). Quasi-experimental shift-share research designs. The Review of Economic Studies, 89(1), 181–213.

Bottia, M. C., Stearns, E., Mickelson, R. A., & Moller, S. (2018). Boosting the numbers of stem majors? the role of high schools with a stem program. Science Education, 102(1), 85–107.

Bound, J., Braga, B., Khanna, G., & Turner, S. (2019). Public universities: The supply side of building a skilled workforce. RSF: The Russell Sage Foundation Journal of the Social Sciences, 5(5), 43–66.

Bound, J., Braga, B., Khanna, G., & Turner, S. (2020). A passage to america: University funding and international students. American Economic Journal: Economic Policy, 12(1), 97–126.

Bourque, J., & Potvin, S. (2021). Cannabis and cognitive functioning: from acute to residual effects, from randomized controlled trials to prospective designs. Frontiers in psychiatry, 12, 596601.

Bowen, W. G., & Levin, S. A. (2003). Reclaiming the game: College sports and educational values. Princeton, New Jersey: Princeton University Press.

Brenøe, A. A., & Zölitz, U. (2020). Exposure to more female peers widens the gender gap in stem participation. Journal of Labor Economics, 38(4), 1009–1054.

Brint, S., German, K. T., Anderson-Natale, K., Shuker, Z. F., & Wang, S. (2020). Where ivy matters: The educational backgrounds of u.s. cultural elites. Sociology of Education, 93(2), 153–172. doi: 10.1177/0038040719898505

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. Journal of Econometrics, 225(2), 200–230.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. The review of economics and statistics, 90(3), 414–427.

Cannabis administration and opportunity act. (2022). https://www.congress.gov/bill/117th-congress/senate-bill/4591/text. (Accessed: 2022-4-9)

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. National Bureau of Economic Research Cambridge, Mass., USA.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. Journal of Labor Economics, 19(1), 22–64.

Card, D., & Payne, A. A. (2021). High school choices and the gender gap in stem. Economic Inquiry, 59(1), 9–28.

Carnevale, A. P., Smith, N., & Melton, M. (2011). Stem: Science technology engineering mathematics. Georgetown University Center on Education and the Workforce.

Carnevale Associates. (2022). Status of state Marijuana Legalization - Carnevale Info Brief. https://www.carnevaleassociates.com/our-work/status-of-state-marijuana-legalization.html. (Online; accessed March 20, 2022)

Cerdá, M., Wall, M., Feng, T., Keyes, K. M., Sarvet, A., Schulenberg, J., … Hasin, D. S. (2017). Association of state recreational marijuana laws with adolescent marijuana use. JAMA pediatrics, 171(2), 142–149.

Cerdá, M., Wall, M., Keyes, K. M., Galea, S., & Hasin, D. (2012). Medical marijuana laws in 50 states: investigating the relationship between state legalization of medical marijuana and marijuana use, abuse and dependence. Drug and alcohol dependence, 120(1-3), 22–27.

Chakrabarti, R., Gorton, N., & Lovenheim, M. F. (2020). State investment in higher education: Effects on human capital formation, student debt, and long-term financial outcomes of students (Tech. Rep.). National Bureau of Economic Research.

Chakraborty, A., Doremus, J., & Stith, S. (2020). The effects of recreational cannabis access on labor markets: evidence from colorado. IZA Journal of Labor Economics, 10(1).

Chan, R. Y. (2016). Understanding the purpose of higher education: An analysis of the economic and social benefits for completing a college degree. Journal of Education Policy, Planning and Administration, 6(5), 1–40.

Chen, S., Oliva, P., & Zhang, P. (2022). The effect of air pollution on migration: Evidence from china. Journal of Development Economics, 156, 102833.

Cheng, C., Mayer, W. J., & Mayer, Y. (2018). The effect of legalizing retail marijuana on housing values: Evidence from colorado. Economic Inquiry, 56(3), 1585–1601.

Cheryan, S., Master, A., & Meltzoff, A. N. (2015). Cultural stereotypes as gatekeepers: Increasing girls' interest in computer science and engineering by diversifying stereotypes. Frontiers in psychology, 6, 49.

Chetty, R., Deming, D. J., & Friedman, J. N. (2023, July). Diversifying society's leaders? the causal effects of admission to highly selective private colleges (Working Paper No. 31492). National Bureau of Economic Research. doi: 10.3386/w31492

Chetty, R., Friedman, J. N., Saez, E., Turner, N., & Yagan, D. (2020a). Income segregation and intergenerational mobility across colleges in the united states. The Quarterly Journal of Economics, 135(3), 1567–1633.

Chetty, R., Friedman, J. N., Saez, E., Turner, N., & Yagan, D. (2020b, 02). Income segregation and intergenerational mobility across colleges in the united states. Quarterly Journal of Economics, 135(3), 1567–1633. doi: 10.1093/qje/qjaa005

Choi, A., Dave, D., & Sabia, J. J. (2019). Smoke gets in your eyes: medical marijuana laws and tobacco cigarette use. American Journal of Health Economics, 5(3), 303–333.

Cole, J. M. (2013, August 29). Memorandum to all united states attorneys. https://www.justice.gov/iso/opa/resources/3052013829132756857467.pdf.

College Board. (2022). 12th grade college application timeline – bigfuture. https://bigfuture.collegeboard.org/plan-for-college/applying-to-college/how-to-apply/college-application-timeline-12th-grade. (Accessed: 2023-06-18)

Cook, R. W., & Zallocco, R. (1979). A multi-attribute attitude model for the prediction of university preference. In Proceedings of the 1979 academy of marketing science (ams) annual conference (pp. 259–261). Cham: Springer International Publishing.

Crisp, G., Nora, A., & Taggart, A. (2009). Student characteristics, pre-college, college, and environmental factors as predictors of majoring in and earning a stem degree: An analysis of students attending a hispanic serving institution.

Crosnoe, R., Riegle-Crumb, C., Field, S., Frank, K., & Muller, C. (2008). Peer group contexts of girls' and boys' academic experiences. Child development, 79(1), 139–155.

Cummings, K., Laderman, S., Lee, J., Tandberg, D., & Weeden, D. (2021). Investigating the impacts of state higher education appropriations and financial aid. State Higher Education Executive Officers.

Dale, S. B., & Krueger, A. B. (2014). Estimating the effects of college characteristics over the career using administrative earnings data. Journal of Human Resources, 42(2), 323–358. doi: 10.3368/jhr.49.2.323

Dave, D. M., Liang, Y., Muratori, C., & Sabia, J. J. (2023). The effects of recreational marijuana legalization on employment and earnings (Tech. Rep.). National Bureau of Economic Research.

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. American Economic Review, 110(9), 2964–96.

Deming, D. J., & Walters, C. R. (2018). The impact of state budget cuts on us postsecondary attainment. Draft, Harvard University, 1567–1633.

Diegert, P., Masten, M. A., & Poirier, A. (2022). Assessing omitted variable bias when the controls are endogenous. arXiv preprint arXiv:2206.02303.

Dilley, J. A., Richardson, S. M., Kilmer, B., Pacula, R. L., Segawa, M. B., & Cerdá, M. (2019). Prevalence of cannabis use in youths after legalization in washington state. JAMA pediatrics, 173(2), 192–193.

Dragone, D., Prarolo, G., Vanin, P., & Zanella, G. (2019). Crime and the legalization of recreational marijuana. Journal of economic behavior & organization, 159, 488–501.

Duarte, R., Escario, J. J., & Molina, J. A. (2006). Marijuana consumption and school failure among spanish students. Economics of Education Review, 25(5), 472–481.

Dynarski, S. M. (2003). Does aid matter? measuring the effect of student aid on college attendance and completion. American Economic Review, 93(1), 279–288.

El Fatmaoui, A., & Ransom, T. (2025). What is the role of sports at elite universities in the us?

Epple, D., Romano, R., Sarpça, S., & Sieg, H. (2017). A general equilibrium analysis of state and private colleges and access to higher education in the U.S. Journal of Public Economics, 155, 164–178. doi: 10.1016/j.jpubeco.2016.09.003

Epple, D., Romano, R., & Sieg, H. (2006). Admission, tuition, and financial aid policies in the market for higher education. Econometrica, 74(4), 885–928. doi: 10.1111/j.1468-0262.2006.00690.x

Espenshade, T. J., & Chung, C. Y. (2005). The opportunity cost of admission preferences at elite universities. Social Science Quarterly, 86(2), 293–305. doi: 10.1111/j.0038-4941.2005.00303.x

Espenshade, T. J., Chung, C. Y., & Walling, J. L. (2004). Admission preferences for minority students, athletes, and legacies at elite universities. Social Science Quarterly, 85(5), 1422–1446. doi: 10.1111/j.0038-4941.2004.00284.x

Estrada, M., Burnett, M., Campbell, A. G., Campbell, P. B., Denetclaw, W. F., Gutiérrez, C. G., … others (2016). Improving underrepresented minority student persistence in stem. CBEâĂŤLife Sciences Education, 15(3), es5.

Gallup. (2021, November). Support for legal marijuana holds at record high of 68%. https://norml.org/blog/2021/11/04/gallup-poll-americans-support-for-marijuana-legalization-holds-at-record-high/.

Garthwaite, C., Keener, J., Notowidigdo, M. J., & Ozminkowski, N. F. (2020, August). Who profits from amateurism? rent-sharing in modern college sports (Working Paper No. 27734). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w27734 doi: 10.3386/w27734

Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2020). Bartik instruments: What, when, why, and how. American Economic Review, 110(8), 2586–2624.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. Journal of Econometrics, 225(2), 254–277.

Griffith, A. L. (2010). Persistence of women and minorities in stem field majors: Is it the school that matters? Economics of Education Review, 29(6), 911–922.

Hall, W., & Lynskey, M. (2016). Evaluating the public health impacts of legalizing recreational cannabis use in the united states. Addiction, 111(10), 1764–1773.

Handel, D. V., & Hanushek, E. A. (2022). Us school finance: Resources and outcomes (Tech. Rep.). National Bureau of Economic Research.

Hart, C. L., Ilan, A. B., Gevins, A., Gunderson, E. W., Role, K., Colley, J., & Foltin, R. W. (2010). Neurophysiological and cognitive effects of smoked marijuana in frequent users. Pharmacology Biochemistry and Behavior, 96(3), 333–341.

Hemelt, S. W., & Marcotte, D. E. (2011). The impact of tuition increases on enrollment at public colleges and universities. Educational Evaluation and Policy Analysis, 33(4), 435–457.

Hemelt, S. W., Stange, K. M., Furquim, F., Simon, A., & Sawyer, J. E. (2021). Why is math cheaper than english? understanding cost differences in higher education. Journal of Labor Economics, 39(2), 397–435.

Hernandez, P. R., Schultz, P., Estrada, M., Woodcock, A., & Chance, R. C. (2013). Sustaining optimal motivation: A longitudinal analysis of interventions to broaden participation of underrepresented students in stem. Journal of educational psychology, 105(1), 89.

Hinrichs, P. (2022). State appropriations and employment at higher education institutions. FRB of Cleveland Working Paper.

Hodge, T. R., & Hazel, C. (2022). The munchies: Marijuana legalization and food sales in washington. Southern Economic Journal, 89(1), 112–137.

Hoxby, C. M. (1997, December). How the changing market structure of u.s. higher education explains college tuition (Working Paper No. 6323). National Bureau of Economic Research. doi: 10.3386/w6323

Immigration, U., & Enforcement, C. (2022). Archived: 2016 dhs stem designated degree program list. https://www.ice.gov/doclib/sevis/pdf/stemList2016.pdf.

Jacob, B., McCall, B., & Stange, K. (2018a). College as country club: Do colleges cater to students' preferences for consumption? Journal of Labor Economics, 36(2), 309–348.

Jacob, B., McCall, B., & Stange, K. (2018b). College as country club: Do colleges cater to students' preferences for consumption? Journal of Labor Economics, 36(2), 309–348. doi: 10.1086/694654

Jayakumar, U. M., & Page, S. E. (2021). Cultural capital and opportunities for exceptionalism: Bias in university admissions. Journal of Higher Education, 92(7), 1109–1139. doi: 10.1080/00221546.2021.1912554

Jiang, S., & Miller, K. (2022). Watching the grass grow: does recreational cannabis legalization affect retail and agricultural wages? Journal of Cannabis Research, 4(1), 1–51.

Kahn, S., & Ginther, D. (2017). Women and stem (Tech. Rep.). National Bureau of Economic Research.

Kara, E., Tonin, M., & Vlassopoulos, M. (2021). Class size effects in higher education: Differences across stem and non-stem fields. Economics of Education Review, 82, 102104.

Ketenci, T., Leroux, A., & Renken, M. (2020). Beyond student factors: A study of the impact on stem career attainment. Journal for STEM Education Research, 3(3), 368–386.

Kim, D., O'Connor, S., & Norwood, B. (2020). Retail marijuana deregulation and housing prices. Available at SSRN 3537860.

Koval, A. L., Kerr, D. C., & Bae, H. (2019). Perceived prevalence of peer marijuana use: changes among college students before and after oregon recreational marijuana legalization. The American Journal of Drug and Alcohol Abuse, 45(4), 392–399.

Leppel, K. (1993). Logit estimation of a gravity model of the college enrollment decision. Research in Higher Education, 34, 387–398.

Lindo, J. M., Swensen, I. D., & Waddell, G. R. (2012). Are big-time sports a threat to student achievement? American Economic Journal: Applied Economics, 4(4), 254–274.

Ma, J. (2021). Trends in college pricing and student aid 2021. College Board.

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

Manski, C. F., & Wise, D. A. (1983). College choice in america. Harvard University Press.

Marie, O., & Zölitz, U. (2017). "high" achievers? cannabis access and academic performance. The Review of Economic Studies, 84(3), 1210–1237.

Marijuana Policy Project. (2019). Michigan. `https://web.archive.org/web/`

20191017091045/https://www.mpp.org/states/michigan/. (Accessed: Feb. 21, 2022)

Marijuana Policy Project. (2022a). Cannabis tax revenue in states that regulate cannabis for adult use. https://www.mpp.org/issues/legalization/cannabis-tax-revenue-states-regulate-cannabis-adult-use/. (Online; accessed March 20, 2022)

Marijuana Policy Project. (2022b). Maine. https://www.mpp.org/states/maine/. (Accessed: Feb. 21, 2022)

Marijuana Policy Project. (2022c). State policy. https://www.mpp.org/states/. (Online; accessed March 20, 2022)

Marijuana opportunity reinvestment and expungement act. (2022). https://www.congress.gov/bill/117th-congress/house-bill/3617. (Accessed: 2022-4-9)

Martins, S. S., Mauro, C. M., Santaella-Tenorio, J., Kim, J. H., Cerda, M., Keyes, K. M., ... Wall, M. (2016). State-level medical marijuana laws, marijuana use and perceived availability of marijuana among the general us population. Drug and alcohol dependence, 169, 26–32.

Mayer, T., & Trevien, C. (2017). The impact of urban public transportation evidence from the paris region. Journal of Urban Economics, 102, 1–21.

Meer, J., & Rosen, H. S. (2009, February). Altruism and the child cycle of alumni donations. American Economic Journal: Economic Policy, 1(1), 258–286. doi: 10.1257/pol.1.1.258

Miller, L., & Park, M. (2022). Making college affordable? the impacts of tuition freezes and caps. Economics of Education Review, 89, 102265.

Montmarquette, C., Cannings, K., & Mahseredjian, S. (2002). How do young people choose college majors? Economics of Education Review, 21(6), 543–556.

National Center for Education Statistics. (2019). Integrated postsecondary education data system (ipeds). https://nces.ed.gov/statprog/handbook/pdf/ipeds.pdf. (Accessed: 2022-4-9)

National Center for Education Statistics. (2022a). Fast Facts: Title IV Institutions. https://

nces.ed.gov/fastfacts/display.asp?id=1122.

National Center for Education Statistics. (2022b). Historically black colleges and universities. https://nces.ed.gov/fastfacts/display.asp?id=667. (Accessed: 2022-4-9)

National Center for Education Statistics. (2024). Table 333.10. total revenue of public degree-granting postsecondary institutions, by source of revenue and level of institution: Selected fiscal years, 2007-08 through 2021-22. https://nces.ed.gov/programs/digest/d23/tables/dt23_333.10.asp. (U.S. Department of Education, Digest of Education Statistics 2023, Table 333.10. Retrieved January 2024)

National Institute on Drug Abuse. (2020). Monitoring the future survey: High school and youth trends. https://monitoringthefuture.org/wp-content/uploads/2022/08/mtf-overview2020.pdf. National Institutes of Health. (Accessed: March 1, 2023)

National Student Clearinghouse Research Center. (2024). Term Enrollment Estimates Fall 2019. https://nscresearchcenter.org/wp-content/uploads/CTEE_Report_Fall_2019.pdf. (Accessed: 2024-07-19)

Norml. (2020). Federal laws and penalties. (https://norml.org/laws/federal-penalties-2/)

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. Journal of Business & Economic Statistics, 37(2), 187–204.

Parmley, K., Bell, A., L'Orange, H., & Lingenfelter, P. (2009). State budgeting for higher education in the united states: As reported for fiscal year 2007. State Higher Education Executive Officers Association.

PCAST. (2012). Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics (Tech. Rep.). Executive Office of the President. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/fact_sheet_final.pdf.

Pearson, M. R., Hustad, J. T., Neighbors, C., Conner, B. T., Bravo, A. J., Team, M. O. S., et al. (2018). Personality, marijuana norms, and marijuana outcomes among college students. Addictive Behaviors, 76, 291–297.

Pennington, B. (2019, November 7, 2019). The real cost of diversifying college rosters. The New York Times. Retrieved from https://www.nytimes.com/2019/11/07/sports/college-sports-diversity-amherst.html (Accessed November 7, 2019)

Pew Research Center. (2019). A Rising Share of Undergraduates Are From Poor Families. https://www.pewresearch.org/social-trends/2019/05/22/a-rising-share-of-undergraduates-are-from-poor-families-especially-at-less-selective-colleges/. (Accessed: 2024-07-19)

Pope, D. G., & Pope, J. C. (2009). The impact of college sports success on the quantity and quality of student applications. Southern Economic Journal, 75(3), 750–780. doi: 10.1002/j.2325-8012.2009.tb00930.x

Pope, D. G., & Pope, J. C. (2014). Understanding college application decisions: Why college sports success matters. Journal of Sports Economics, 15(2), 107–131. doi: 10.1177/1527002512445569

Pope, H. G., Jacobs, A., Mialet, J.-P., Yurgelun-Todd, D., & Gruber, S. (1997). Evidence for a sex-specific residual effect of cannabis on visuospatial memory. Psychotherapy and psychosomatics, 66(4), 179–184.

Portillo, S., & DeHart-Davis, L. (2009). Gender and organizational rule abidance. Public Administration Review, 69(2), 339–347.

Raman, S., & Bradford, A. C. (2022). Recreational cannabis legalizations associated with reductions in prescription drug utilization among medicaid enrollees. Health Economics, 31(7), 1513–1521.

Ransom, M. R., & Ransom, T. (2018). Do high school sports build or reveal character? bounding causal estimates of sports participation. Economics of Education Review, 64, 75–89. doi: 10.1016/j.econedurev.2018.04.002

Rinker, D. V., Krieger, H., & Neighbors, C. (2016). Social network factors and addictive behaviors among college students. Current addiction reports, 3, 356–367.

Roback, J. (1982). Wages, rents, and the quality of life. The Journal of Political Economy, 90(6), 1257–1278.

Roebuck, M. C., French, M. T., & Dennis, M. L. (2004). Adolescent marijuana use and school attendance. Economics of Education Review, 23(2), 133–141.

Rogers, Simon. (2022). What is google trends data âĂŤ and what does it mean? https://medium.com/google-news-lab/what-is-google-trends-data-and-what-does-it-mean-b48f07342ee8. (Accessed: 2022-4-9)

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. The Journal of Political Economy, 82(1), 34–55.

Rueben, K., & Randall, M. (2017, November). Balanced budget requirements: How states limit deficit spending (Fact Sheet). Washington, DC: Urban Institute. https://www.urban.org/.

Sabia, J. J., Dave, D. M., Alotaibi, F., & Rees, D. I. (2021). Is recreational marijuana a gateway to harder drug use and crime? (Tech. Rep.). National Bureau of Economic Research.

Sarvet, A. L., Wall, M. M., Fink, D. S., Greene, E., Le, A., Boustead, A. E., ... others (2018). Medical marijuana laws and adolescent marijuana use in the united states: A systematic review and meta-analysis. Addiction, 113(6), 1003–1016.

Schafer, G., Feilding, A., Morgan, C. J., Agathangelou, M., Freeman, T. P., & Curran, H. V. (2012). Investigating the interaction between schizotypy, divergent thinking and cannabis use. Consciousness and cognition, 21(1), 292–298.

Schulenberg, J. E., Patrick, M. E., Johnston, L. D., O'Malley, P. M., Bachman, J. G., & Miech, R. A. (2021). Monitoring the future national survey results on drug use, 1975-2020. volume ii, college students & adults ages 19-60. Institute for Social Research.

Scott, J. C., Slomiak, S. T., Jones, J. D., Rosen, A. F., Moore, T. M., & Gur, R. C. (2018). Association of cannabis with cognitive functioning in adolescents and young adults: a systematic review and meta-analysis. JAMA psychiatry, 75(6), 585–595.

Shapiro, J. R., & Williams, A. M. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in stem fields. Sex roles, 66, 175–183.

Sjoquist, D. L., & Winters, J. V. (2015). State merit aid programs and college major: A focus on stem. Journal of Labor Economics, 33(4), 973–1006.

Skinner, B. (2023). rscorecard: A method to download department of education college scorecard data [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=rscorecard (R package version 0.26.0)

Speer, J. D. (2017). The gender gap in college major: Revisiting the role of pre-college factors. Labour Economics, 44, 69–88.

State Higher Education Executive Officers Association. (2023). Grapevine, fy 2023. https://shef.sheeo.org/grapevine/. (Accessed: 2023-6-6)

States reform act. (2021). https://www.congress.gov/bill/117th-congress/house-bill/5977?s=1&r=1. (Accessed: 2022-4-9)

Suerken, C. K., Reboussin, B. A., Egan, K. L., Sutfin, E. L., Wagoner, K. G., Spangler, J., & Wolfson, M. (2016). Marijuana use trajectories and academic outcomes among college students. Drug and alcohol dependence, 162, 137–145.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics, 225(2), 175–199.

Supreme Court of the United States. (2021, June). NCAA v. alston. Retrieved from https://www.supremecourt.gov/opinions/20pdf/20-512_gfbh.pdf (Accessed on 2021-08-17)

Tax Policy Center. (2020). State and local tax policies. https://www.taxpolicycenter.org/briefing-book/what-are-state-balanced-budget-requirements-and-how-do-they-work.

The U.S. Department of Homeland Security. (2020). Dhs stem designated degree program list. https://www.ice.gov/sites/default/files/documents/stem-list.pdf. (Accessed: 2024-01-05)

Tyagi, E., Ueyama, K., & The Urban Institute. (2022). educationdata: Retrieve records from the urban institute's education data portal api [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=educationdata (R package version 0.1.3)

University of Kentucky Center for Poverty Research. (2023). UKCPR National Welfare Data, 1980-2021. http://ukcpr.org/resources/national-welfare-data. (Accessed 2023-09-28)

U.S. Census Bureau. (2024). State characteristics data file: 2000-2020. Available online at U.S. Census Bureau. Washington, D.C.. (2000-2009 Data File (2009) and 2010-2020 Data File (2020))

U.S. Department of Education, National Center for Education Statistics. (2024). Integrated postsecondary education data system (ipeds), 2024. `https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx?year=-1&surveyNumber=-1&sid=ac68b949-876c-439b-abf0-e431b89449a2&rtid=5`. (Retrieved on January, 2024)

U.S. Department of Homeland Security. (2023). Review the updated dhs stem designated degree program list. `https://studyinthestates.dhs.gov/2023/07/review-the-updated-dhs-stem-designated-degree-program-list`. (Accessed: 2024-07-13)

Walker, K., Herman, M., & Eberwein, K. (2023). tidycensus: Load us census boundary and attribute data as 'tidyverse' and 'sf'-ready data frames [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=tidycensus` (R package version 1.4.4)

Webber, D. A. (2017). State divestment and tuition at public institutions. Economics of Education Review, 60, 1–4.

Webber, D. A., & Ehrenberg, R. G. (2010). Do expenditures other than instructional expenditures affect graduation and persistence rates in american higher education? Economics of Education Review, 29(6), 947–958.

Wen, H., Hockenberry, J. M., & Cummings, J. R. (2015). The effect of medical marijuana laws on adolescent and adult use of marijuana, alcohol, and other substances. Journal of health economics, 42, 64–80.

Winston, G. C. (1999). Subsidies, hierarchy and peers: The awkward economics of higher education. Journal of Economic Perspectives, 13(1), 13–36. doi: 10.1257/jep.13.1.13

Wiswall, M., & Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. The Review of Economic Studies, 82(2), 791–824.

Wolniak, G. C., Seifert, T. A., Reed, E. J., & Pascarella, E. T. (2008). College majors and social mobility. Research in social stratification and mobility, 26(2), 123–139.

Wright, A. C., & Krieg, J. M. (2020). Getting into the weeds: Does legal marijuana access blunt academic performance in college? Economic Inquiry, 58(2), 607–623.

Wright, T. J. (2022). Replication of "how much does immigration boost innovation?". Economic Inquiry.

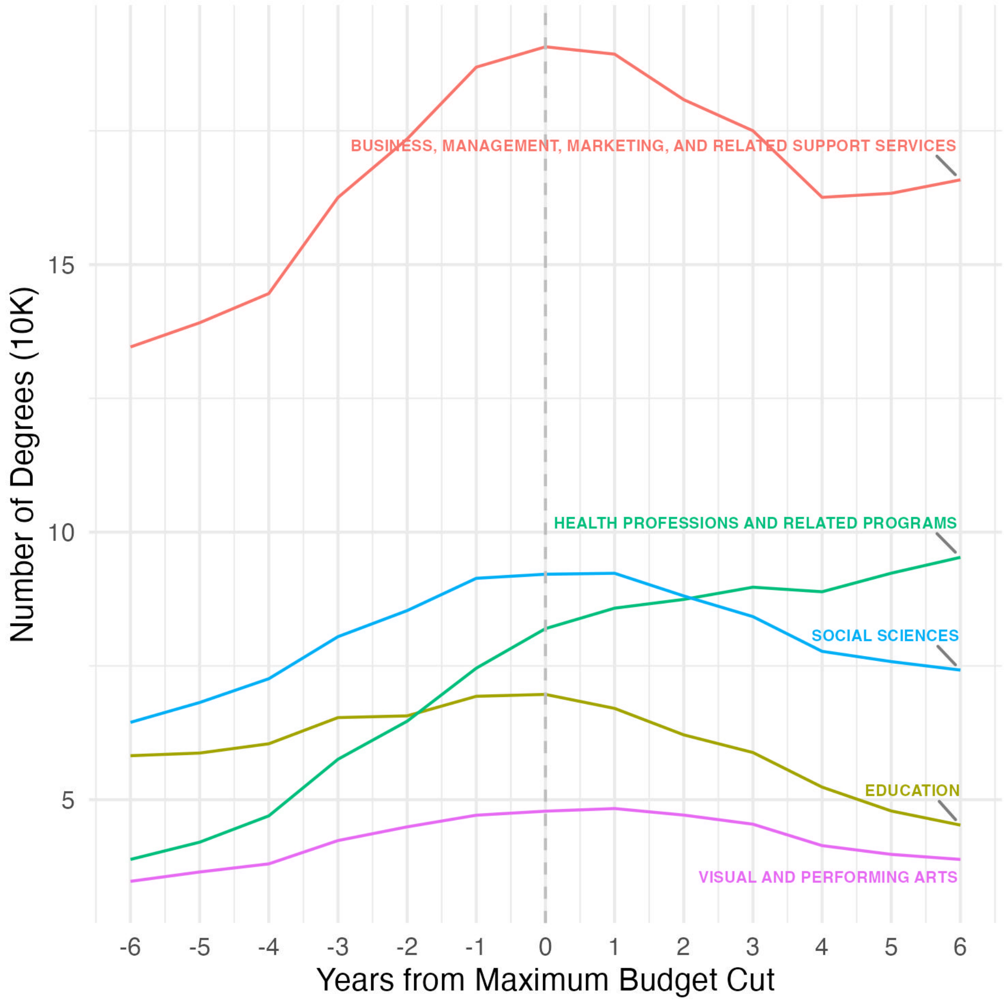
Wu, G., Wen, M., & Wilson, F. A. (2021). Impact of recreational marijuana legalization on crime: Evidence from oregon. Journal of criminal justice, 72, 101742.

Zambiasi, D., & Stillman, S. (2020). The pot rush: Is legalized marijuana a positive local amenity? Economic Inquiry, 58(2), 667–679.

Appendix

A.1 Chapter 1

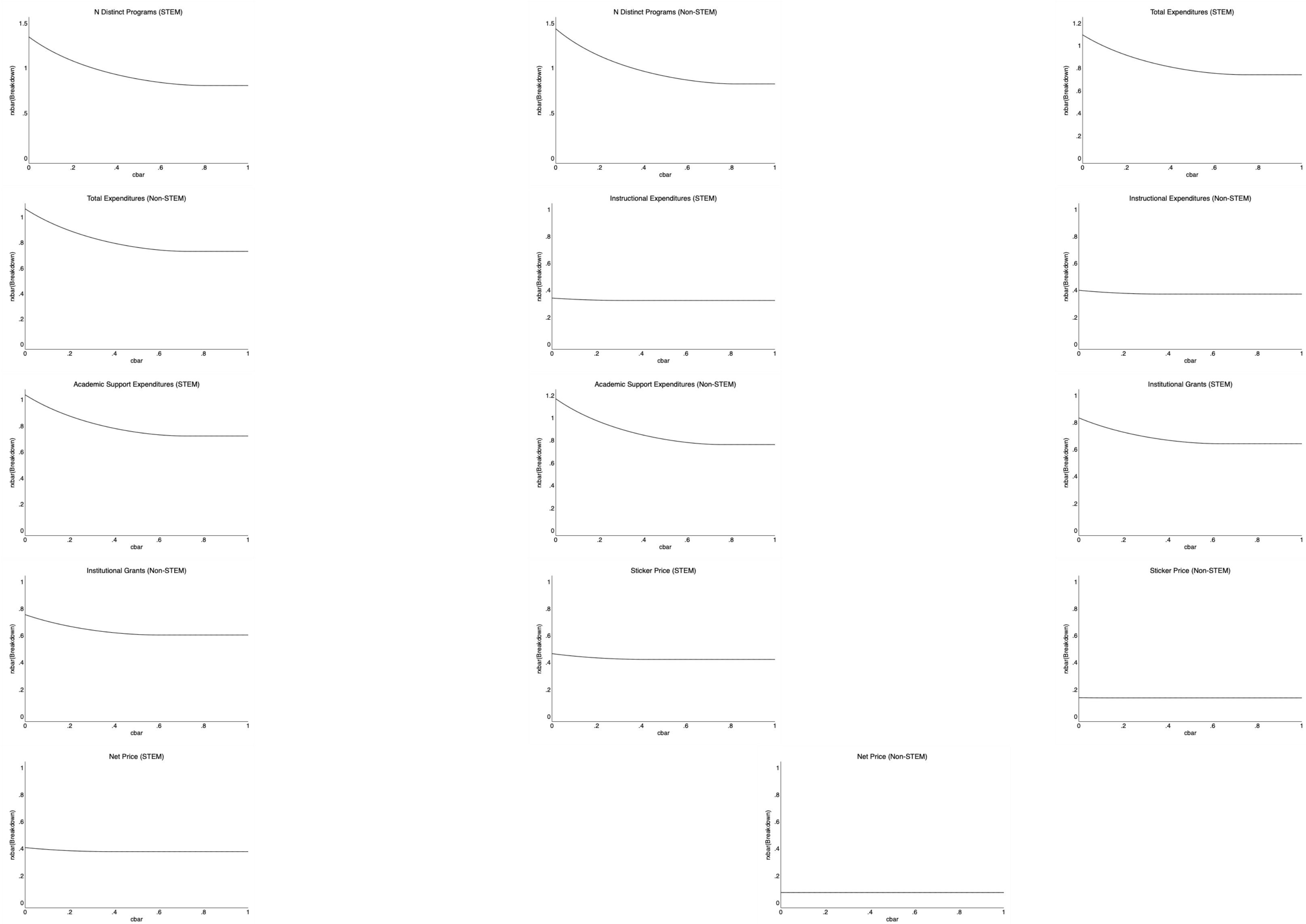
Figure A.1: Number of Non-STEM Bachelor Degrees Vs. Budget Shock



SOURCES.— Data on Number of Degrees (Bachelor) and State Appropriation are sourced from the Integrated Postsecondary Education Data System (IPEDS) as reported in U.S. Department of Education, National Center for Education Statistics (2024). In particular, Number of Degrees is derived from the Completions–Awards/degrees conferred by program (6-digit CIP code), award level, race/ethnicity, and gender, while State Appropriation information is obtained from the Finance Survey. STEM classification is obtained from The U.S. Department of Homeland Security (2020).
All monetary values are adjusted to 2019 dollar terms utilizing the Consumer Price Index (CPI) data provided by the Bureau of Labor Statistics (BLS)

NOTES.— The unit of analysis is college-program type and year. The Figures displays yearly number of STEM bachelor's degrees awarded by all 4-year public institutions for four years before and after a maximum cut in state appropriations for each university in the period 2003 to 2019.

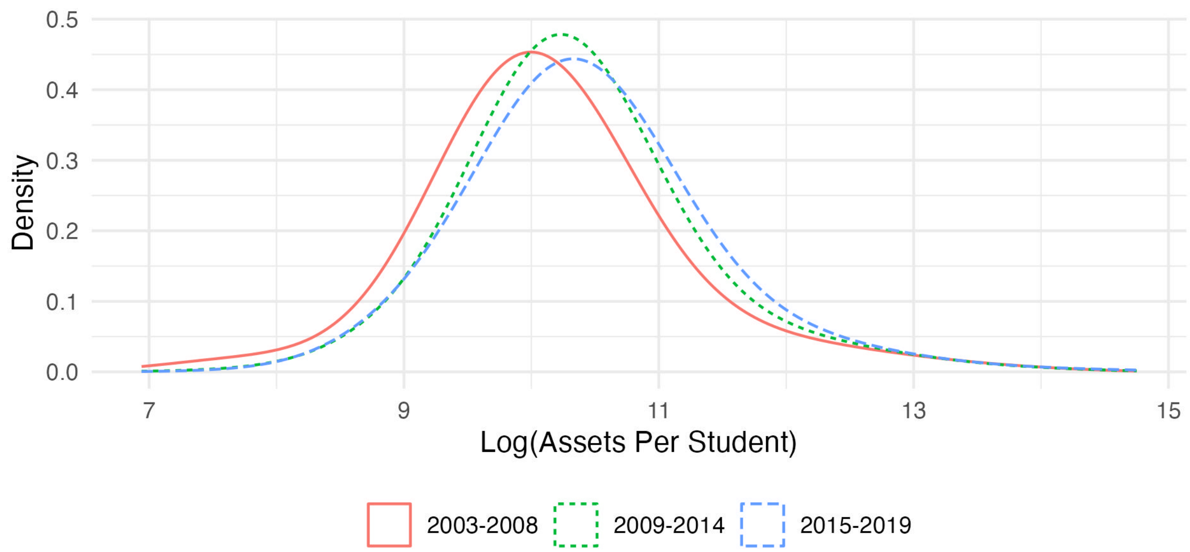
Figure A.2: Sensitivity Analysis (DMP, 2022), Breakdown



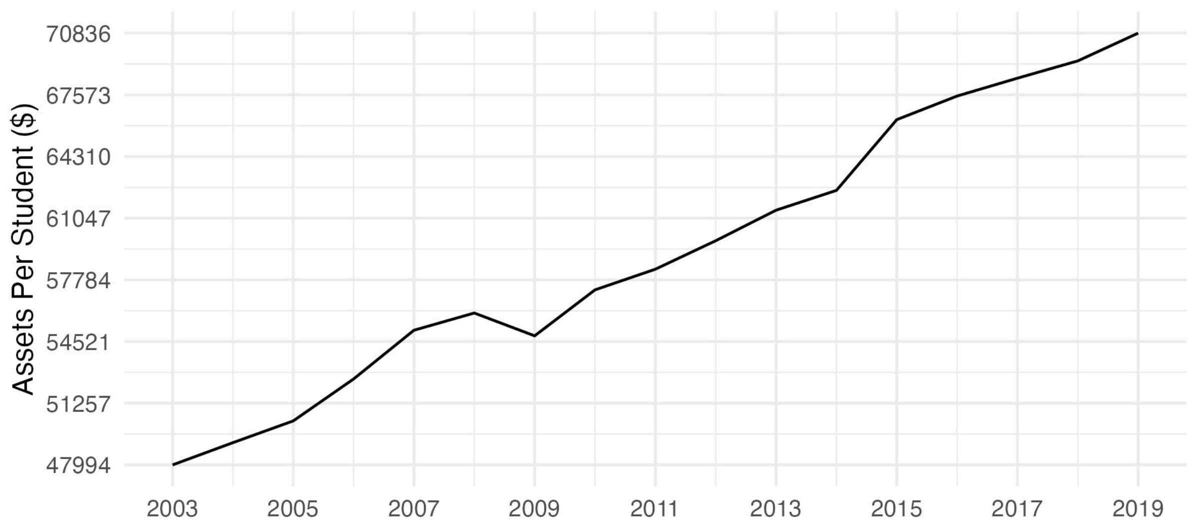
NOTES.— The figures show how the breakdown points ($\bar{r}$), following Diegert et al. (2022), vary with different correlation levels ($\bar{c}$). The title of each subfigure indicates the independent variable and the dependent variable in parentheses, referring to either the number of STEM (Science, Technology, Engineering, and Mathematics) or non-STEM degrees.

Figure A.3: Public University Assets Per Student

(a) Density



(b) Average

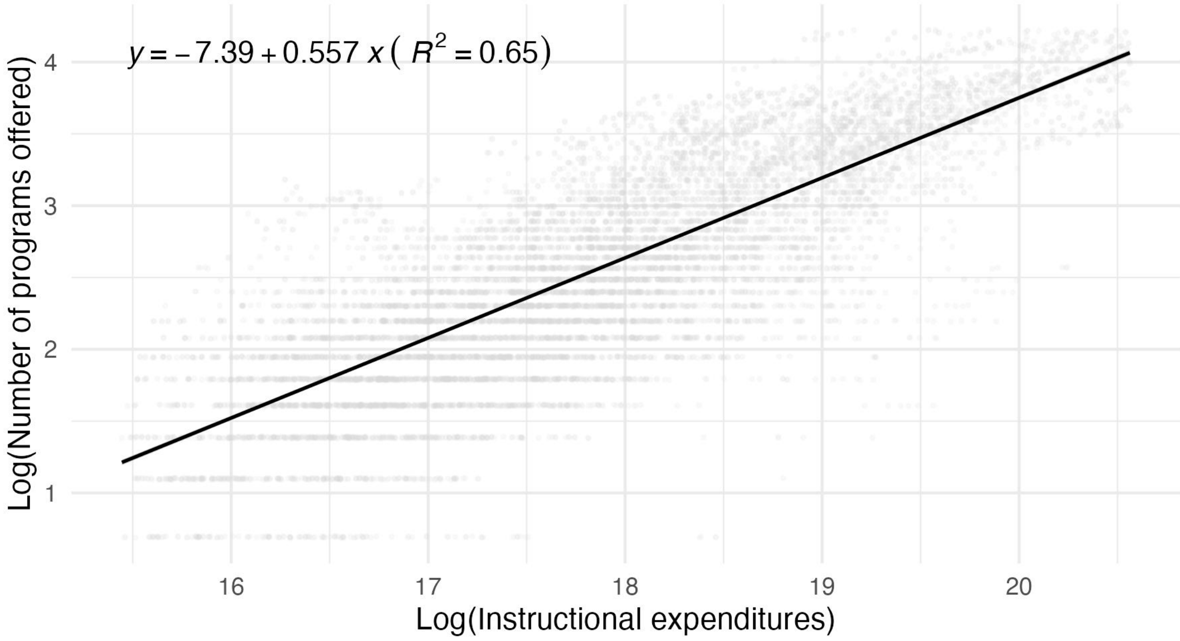


SOURCES.— State Appropriation is extracted from IPEDS finance survey, and universities' headcount or total enrollment from IPEDS 12-Month Enrollment survey.

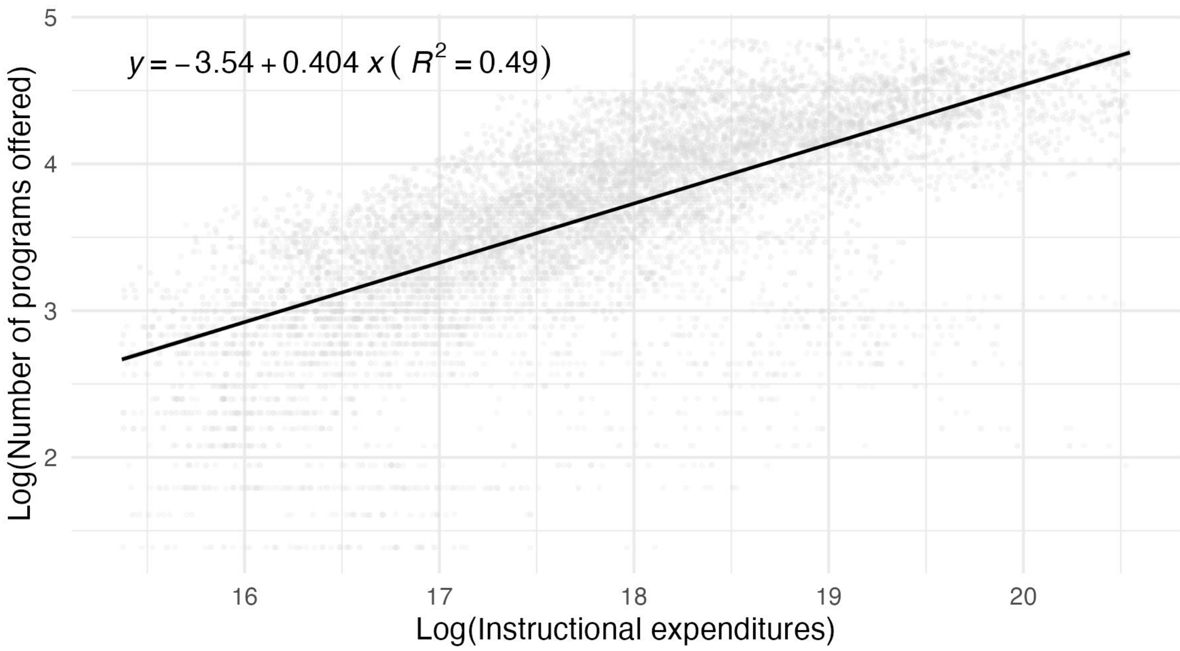
NOTES.— All monetary values are adjusted to 2019 dollar terms utilizing the Consumer Price Index (CPI) data provided by the Bureau of Labor Statistics (BLS). Density plot (a) presents the distribution of log-transformed state appropriations per student. The data spans equally spaced intervals from 2003 to 2019. A kernel density estimate (KDE) is utilized to offer a smooth approximation of the probability density function (PDF). Figure (b) shows the average college assets per student, and tracks its progression over the specified time frame.

Figure A.4: Number of Programs Offered vs. Instructional Expenditures, by Programs Type

(a) STEM



(b) Non-STEM



SOURCES.— Instructional Expenditures is extracted from IPEDS finance survey, and number of distinct programs offered is from IPEDS completion survey.

NOTES.— These plots display the relationship between log-transformed instructional expenditures and the log-transformed number of programs offered for STEM (top panel) and non-STEM (bottom panel) institutions. Data points represent individual year-institutions, with jittering applied for visual clarity. The solid line shows the linear fit, with the corresponding equation and R-squared value displayed in the upper left corner of each plot. Both axes are log-scaled. The sample is trimmed to exclude the top and bottom 1% of observations for both variables to mitigate the influence of outliers.

Table A.1: Summary Statistics (Full Sample)

	N	Mean	SD
University-Program-Year Level (Program at 6-digit CIP)			
Number of Degrees (Aggregate)			
Total	427,133	42.426	72.345
White	427,133	27.289	44.381
Black	427,133	3.718	9.818
Hispanic	427,133	4.280	15.803
Asian	427,133	3.106	14.276
American Indian	427,133	0.271	1.097
Number of Degrees (Men)			
Men	427,133	18.450	34.586
Men White	427,133	12.112	21.846
Men Black	427,133	1.309	3.479
Men Hispanic	427,133	1.686	6.170
Men Asian	427,133	1.447	7.159
Men American Indian	427,133	0.106	0.489
Number of Degrees (Women)			
Women	427,133	23.976	46.321
Women White	427,133	15.177	29.243
Women Black	427,133	2.409	7.165
Women Hispanic	427,133	2.595	10.841
Women Asian	427,133	1.659	7.851
Women American Indian	427,133	0.165	0.796
University-Year Level			
N Distinct Programs	9,973	67.121	43.363
State Appropriation ($M)	9,973	91.254	114.723
Total Expenditures ($M)	9,973	429.281	763.451
Instructional Expenditures ($M)	9,973	124.185	179.666
Academic Support Expenditures ($M)	9,973	33.734	59.590
Institutional Grants ($M)	9,971	16.828	32.157
Endowment Assets ($M)	9,468	194.044	686.054
Net Price[a]	7,483	12989.120	4063.486
Sticker Price	9,602	7852.819	3032.147
County-Year Level			
Unemployment Rate	8,450	5.834	2.309
Per-capita Personal Income	8,450	39196.304	11903.565
Population	8,450	369946.092	709762.508
Birth Rate	8,450	1196.440	331.174
Death Rate	8,450	819.947	254.754
Share Male Population	8,450	0.493	0.015
Share White Population	8,450	0.805	0.159
Share Black Population	8,450	0.125	0.150
Share Hispanic population	8,450	0.109	0.146
State-Year Level			
State Population (18 to 21)	850	355535.979	394470.157
Per-capita Personal Income	850	42465.868	9268.220
Per 100k SNAP Recipients	850	11578.477	4200.258
Per 100k TANF Recipients	850	1094.994	753.232
Per 100k Medicaid Recipients	850	17707.590	5916.654
Poverty Rate	850	12.612	3.335
State Minimum Wage	850	7.059	1.480
Democratic Governor Dummy	850	0.435	0.496

NOTES.—The unit of analysis is institution-program at the 6-digit CIP (Classification of Instructional Programs) code-year level for the program-level variables. These variables indicate the number of students who graduated in specific majors with bachelor's degrees. The data includes all the programs (STEM and non-STEM) from 2003 to 2019. Other variables are either university-related or pertain to the economic or demographic characteristics of the university's location at the state or county level.

[a] Net Price, which refers to the actual cost of attending college net of any aid or grants, is introduced in IPEDS (Integrated Postsecondary Education Data System) through the Student Financial Aid and Net Price survey beginning from 2007; thus, data for prior years is missing.

Table A.2: Summary Statistics–Number of Bachelor Degrees, by Major Type

	Non-STEM (N=351,306)		STEM (N=124,880)		
	Mean	SD	Mean	SD	Diff. in Means
N Distinct Programs [a]	49.626	31.128	18.397	14.893	-31.229***
Number of Degrees:					
Total	40.735	75.351	30.793	49.707	-9.942***
Women	25.111	49.105	11.514	24.814	-13.597***
Men	15.624	33.707	19.279	31.646	3.655***
White	26.489	46.602	19.060	28.953	-7.429***
Black	3.874	10.381	1.843	5.319	-2.031***
Hispanic	4.295	16.688	2.563	8.661	-1.732***
Asian	2.440	13.239	3.763	14.359	1.323***
American Indian	0.272	1.135	0.163	0.716	-0.109***

SOURCES.— The data are extracted from the Integrated Postsecondary Education Data System (IPEDS) Completions Survey, specifically from the subsurvey Awards/Degrees Conferred by Program (6-digit Classification of Instructional Programs/ CIP code), Award Level, Race/Ethnicity, and Gender (U.S. Department of Education, National Center for Education Statistics, 2024).

NOTES.—The unit of analysis is institution-program at the 6-digit CIP code-year level. STEM refers to Science, Technology, Engineering, and Mathematics programs or majors as designated by the Department of Homeland Security (DHS). Table A.25 presents the complete list of STEM programs. The table shows the mean difference in the number of bachelor's degrees conferred at 6-digit CIP code between STEM and non-STEM majors. See Table A.1 for summary statistics for all variables, including the state and county covariates. Table A.3 extends the summary statistics for the number of degrees by providing a breakdown by gender and race. Table A.4 presents the same summary statistics but compares different periods.

[a] The number of distinct programs offered is at the institution level.

Table A.3: Summary Statistics–Number of Bachelor Degrees, by Major Type, Gender and Race

	Non-STEM (N=351,306)		STEM (N=124,880)		
	Mean	SD	Mean	SD	Diff. in Means
Number of Degrees (Men)					
White	10.396	21.411	12.295	19.738	1.899***
Black	1.257	3.486	0.949	2.784	-0.308***
Hispanic	1.509	6.213	1.523	4.755	0.014
Asian	0.976	6.255	2.206	8.053	1.230***
American Indian	0.096	0.466	0.095	0.46	-0.001
Number of Degrees (Women)					
White	16.093	31.229	6.765	13.972	-9.328***
Black	2.616	7.657	0.893	3.24	-1.723***
Hispanic	2.786	11.604	1.04	4.83	-1.746***
Asian	1.464	7.481	1.557	7.372	0.093***
American Indian	0.176	0.844	0.068	0.408	-0.108***

NOTES.— The unit of analysis is institution-program at the 6-digit CIP code-year level. STEM refers to Science, Technology, Engineering, and Mathematics programs or majors as designated by the Department of Homeland Security (DHS). Table A.25 presents the complete list of STEM programs. See Table A.1 for summary statistics for all variables, including the state and county covariates.

Table A.4: Summary Statistics–Number of Bachelor Degrees, by STEM Type and Period

	Panel (a): STEM Degrees					Panel (b): Non- STEM Degrees				
	2003 to 2007 (N=33,148)		2015 to 2019 (N=40,805)			2003 to 2007 (N=96,717)		2015 to 2019 (N=111,166)		
	Mean	SD	Mean	SD	Diff. in Means	Mean	SD	Mean	SD	Diff. in Means
Number of Degrees (Aggregate)										
Total	25.431	39.336	37.417	59.898	11.986***	38.244	67.464	41.384	83.295	3.140***
White	16.885	25.215	21.49	32.514	4.605***	26.9	45.248	24.592	47.496	-2.307***
Black	1.665	4.853	2.126	6.196	0.461***	3.434	9.092	4.147	11.477	0.713***
Hispanic	1.388	4.486	4.077	12.258	2.689***	2.82	11.134	5.894	22.144	3.074***
Asian	2.981	12.186	4.694	15.945	1.713***	2.251	13.468	2.491	13.282	0.239***
American Indian	0.174	0.706	0.14	0.681	-0.034***	0.306	1.274	0.207	0.907	-0.100***
Number of Degrees (Men)										
Total	16.121	26.078	23.162	38.552	7.041***	14.537	30.853	15.672	35.958	1.136***
White	11.021	17.686	13.663	22.174	2.642***	10.443	21.258	9.53	20.876	-0.913***
Black	0.828	2.476	1.118	3.391	0.290***	1.064	3.096	1.377	3.755	0.314***
Hispanic	0.83	2.609	2.402	6.658	1.572***	0.985	3.92	2.057	8.32	1.072***
Asian	1.768	7.111	2.758	9.378	0.990***	0.877	6.015	0.978	6.255	0.101***
American Indian	0.101	0.458	0.081	0.423	-0.019***	0.106	0.508	0.072	0.379	-0.034***
Number of Degrees (Women)										
Total	9.31	18.88	14.255	29.734	4.944***	23.707	42.979	25.712	55.61	2.005***
White	5.863	11.806	7.827	15.769	1.964***	16.456	29.515	15.063	32.981	-1.394***
Black	0.837	2.917	1.008	3.687	0.172***	2.371	6.62	2.77	8.545	0.399***
Hispanic	0.558	2.395	1.674	6.935	1.117***	1.835	8.056	3.837	15.252	2.003***
Asian	1.213	6.102	1.936	7.941	0.723***	1.374	7.831	1.513	7.568	0.139***
American Indian	0.073	0.395	0.059	0.412	-0.014***	0.2	0.954	0.134	0.668	-0.066***

NOTES.—The unit of analysis is institution-program at the 6-digit CIP code-year level. The table shows the mean difference in the number of degrees between two periods for STEM degrees in panel (a) and for non-STEM degrees in panel (b). STEM refers to Science, Technology, Engineering, and Mathematics programs or majors as designated by the Department of Homeland Security (DHS). Table A.25 presents the complete list of STEM programs. See Table A.1 for summary statistics for all variables, including the state and county covariates.

Table A.5: OLS Estimates of the Effect of State Appropriations on Bachelor's Degrees by STEM Type

	Dependent Variable: Log Number of Bachelor Degrees					
	t+4		t+5		t+6	
Panel (a): STEM						
Log State Appropriations	0.192**	0.206***	0.171**	0.177**	0.101	0.106
	(0.079)	(0.078)	(0.083)	(0.081)	(0.079)	(0.077)
Observations	85,574	84,262	77,206	76,019	69,228	68,157
Adj. R2	0.272	0.271	0.275	0.274	0.275	0.274
Panel (b): Non-STEM						
Log State Appropriations	0.199**	0.178**	0.145*	0.124	0.125*	0.105
	(0.083)	(0.081)	(0.078)	(0.075)	(0.073)	(0.071)
Observations	238,383	235,340	214,454	211,681	191,533	189,026
Adj. R2	0.301	0.301	0.299	0.299	0.297	0.297
Institution FE	✓	✓	✓	✓	✓	✓
Program FE (6-Digit CIP)	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓
County Controls	×	✓	×	✓	×	✓

NOTES.—Each panel and cell refers to a separate OLS regression. The outcome variable in panel (a) is the logged number of undergraduate Science, Technology, Engineering, and Mathematics (STEM) degrees (bachelor's) conferred. In panel (b), the dependent variable is the logged number of non-STEM bachelor's degrees. A lead of "t+i" (ranging from 3 to 6) indicates the ith lead of the outcome variable. For example, the lead t+6 refers to the student cohort that completes bachelor's education in 6 years. CIP refers to the Classification of Instructional Programs, which is used to categorize degree programs by specific majors (see Table A.25 for the subset of STEM majors). The unit of analysis is at the institution, program (6-digit CIP), and year level. The models include year, program, and institution fixed effects, as well as time-varying county and state covariates. The state-level controls include poverty rate, logged population aged 18 to 24, per-capita personal income, minimum wage, number of Supplemental Nutrition Assistance Program (SNAP) recipients per 100,000 population, and number of Temporary Assistance for Needy Families (TANF) recipients per 100,000 population. The county-level controls include logged per-capita personal income, total population, unemployment rate, percentage of population by race (Hispanic, Black, and White), percentage of male population, and birth and death rates per 100,000 population. Standard errors are clustered at the institution level. *p < 0.10, **p < 0.05, ***p < 0.01

Table A.6: The Effect of State Appropriations on Bachelor's Degrees, by STEM Type–Additional Fixed Effects

	Dependent Variable: Log Number of Bachelor Degrees							
	t+3		t+4		t+5		t+6	
Panel (a): STEM								
Log State Appropriations	0.172	0.252**	0.341**	0.260**	0.297**	0.203**	0.235**	0.176**
	(0.176)	(0.126)	(0.169)	(0.115)	(0.130)	(0.090)	(0.108)	(0.082)
Observations	92,804	92,804	84,262	84,262	76,019	76,019	68,157	68,157
Adj. R2	0.257	0.596	0.157	0.590	0.185	0.627	0.221	0.645
F-test (1st stage)	99.742	105.55	114.24	114.96	182.92	168.69	171.02	157.24
Panel (b): Non-STEM								
Log State Appropriations	0.171	0.248	0.206	0.180	0.188*	0.129	0.142	0.099
	(0.219)	(0.153)	(0.170)	(0.116)	(0.111)	(0.082)	(0.091)	(0.070)
Observations	259,969	259,969	235,340	235,340	211,681	211,681	189,026	189,026
Adj. R2	0.301	0.641	0.293	0.659	0.294	0.670	0.301	0.676
F-test (1st stage)	260.50	262.12	363.31	321.88	632.93	534.53	583.50	486.94
Institution FE	✓	✓	✓	✓	✓	✓	✓	✓
Program FE (6-Digit CIP)	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Institution-by-program FE	×	✓	×	✓	×	✓	×	✓
Program-by-year FE	×	✓	×	✓	×	✓	×	✓
State Controls	✓	✓	✓	✓	✓	✓	✓	✓
County Controls	×	✓	×	✓	×	✓	×	✓

NOTES.—Each panel and cell refers to a separate 2SLS regression. The outcome variable in panel (a) is the logged number of undergraduate Science, Technology, Engineering, and Mathematics (STEM) degrees (bachelor's) conferred. In panel (b), the dependent variable is the logged number of non-STEM bachelor's degrees. A lead of "t+i" (ranging from 3 to 6) indicates the ith lead of the outcome variable. For example, the lead t+6 refers to the student cohort that completes bachelor's education in 6 years. CIP refers to the Classification of Instructional Programs, which is used to categorize degree programs by specific majors (see Table A.25 for the subset of STEM majors). The unit of analysis is at the institution, program (6-digit CIP), and year level. The models include year, program, and institution fixed effects, as well as time-varying county and state covariates. The state-level controls include poverty rate, logged population aged 18 to 24, per-capita personal income, minimum wage, number of Supplemental Nutrition Assistance Program (SNAP) recipients per 100,000 population, and number of Temporary Assistance for Needy Families (TANF) recipients per 100,000 population. The county-level controls include logged per-capita personal income, total population, unemployment rate, percentage of population by race (Hispanic, Black, and White), percentage of male population, and birth and death rates per 100,000 population. Standard errors are clustered at the institution level. *p < 0.10, **p < 0.05, ***p < 0.01

Table A.7: The Effect of State Appropriations on Master's Degrees, by STEM Type

	t+1		t+2		t+3		t+4	
	Dependent Variable: Log Number of STEM Master's Degrees							
Log State Appropriations	0.621	0.628	0.782	0.826	0.572**	0.651*	0.573**	0.704*
	(0.465)	(0.536)	(0.640)	(0.765)	(0.281)	(0.357)	(0.272)	(0.365)
Observations	76,970	75,657	69,958	68,751	63,461	62,356	57,373	56,364
F-test (1st stage)	49.79	46.78	38.34	35.75	93.19	79.92	73.48	54.99
	Dependent Variable: Log Number of Non-STEM Master's Degrees							
Log State Appropriations	0.365	0.338	0.378	0.377	0.341	0.349	0.335	0.348
	(0.373)	(0.499)	(0.366)	(0.486)	(0.257)	(0.304)	(0.246)	(0.292)
Observations	187,548	184,639	170,021	167,351	153,629	151,184	138,131	135,916
F-test (1st stage)	95.17	57.14	87.58	52.49	197.80	143.41	208.18	145.24
Institution FE	✓	✓	✓	✓	✓	✓	✓	✓
Program FE (6-Digit CIP)	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓	✓	✓
County Controls	×	✓	×	✓	×	✓	×	✓

NOTES.—Each panel refers to a separate 2SLS regression. The outcome variable in panel (a) is the logged number of graduate Science, Technology, Engineering, and Mathematics (STEM) degrees (Master's) conferred. In panel (b), the dependent variable is the logged number of non-STEM Master's degrees. A lead of "t+i" (ranging from 3 to 6) indicates the ith lead of the outcome variable. For example, the lead t+3 refers to the student cohort that completes Master's education in 3 years. CIP refers to the Classification of Instructional Programs, which is used to categorize degree programs by specific majors (see Table A.25 for the subset of STEM majors). The unit of analysis is at the institution, program (6-digit CIP), and year level. The models include year, program, and institution fixed effects, as well as time-varying county and state covariates. The state-level controls include poverty rate, logged population aged 18 to 24, per-capita personal income, minimum wage, number of Supplemental Nutrition Assistance Program (SNAP) recipients per 100,000 population, and number of Temporary Assistance for Needy Families (TANF) recipients per 100,000 population. The county-level controls include logged per-capita personal income, total population, unemployment rate, percentage of population by race (Hispanic, Black, and White), percentage of male population, and birth and death rates per 100,000 population. Standard errors are clustered at the institution level. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

Table A.8: The Effect of State Appropriations on Doctor's Degrees, by STEM Type

	t+3		t+4		t+5		t+6	
	Dependent Variable: Log Number of STEM Doctor's Degrees							
Log State Appropriations	0.305	0.211	0.254	0.178	0.356	0.355	-0.022	-0.216
	(0.261)	(0.267)	(0.240)	(0.262)	(0.286)	(0.367)	(0.162)	(0.297)
Observations	40,074	39,122	36,233	35,359	32,517	31,724	29,033	28,319
F-test (1st stage)	56.06	44.71	44.79	33.21	43.98	26.49	43.74	22.39
	Dependent Variable: Log Number of Non-STEM Doctor's Degrees							
Log State Appropriations	0.407	0.297	0.335	0.296	0.257	0.230	0.333	0.308
	(0.318)	(0.342)	(0.267)	(0.306)	(0.194)	(0.204)	(0.235)	(0.238)
Observations	45,302	44,338	40,568	39,709	35,990	35,225	31,625	30,946
F-test (1st stage)	69.72	49.54	83.16	49.77	144.53	88.37	132.08	82.03
Institution FE	✓	✓	✓	✓	✓	✓	✓	✓
Program FE (6-Digit CIP)	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓	✓	✓
County Controls	×	✓	×	✓	×	✓	×	✓

Note: —Each panel refers to a separate 2SLS regression. The outcome variable in panel (a) is the logged number of graduate Science, Technology, Engineering, and Mathematics (STEM) degrees (Doctor's) conferred. In panel (b), the dependent variable is the logged number of non-STEM Doctor's degrees. A lead of "t+i" (ranging from 3 to 6) indicates the ith lead of the outcome variable. For example, the lead t+6 refers to the student cohort that completes Doctor's education in 6 years. CIP refers to the Classification of Instructional Programs, which is used to categorize degree programs by specific majors (see Table A.25 for the subset of STEM majors). The unit of analysis is at the institution, program (6-digit CIP), and year level. The models include year, program, and institution fixed effects, as well as time-varying county and state covariates as described in Table A.7. *p < 0.10, **p < 0.05, ***p < 0.01

Table A.9: The Effect of State Appropriations on Spending

	t+0		t+1		t+2		t+3	
	Dependent Variable (a): Log Total Expenditures							
Log State Appropriations	0.130***	0.122***	0.105***	0.101***	0.066***	0.062***	0.041**	0.036**
	(0.033)	(0.032)	(0.026)	(0.025)	(0.023)	(0.022)	(0.019)	(0.018)
N Institutions	663	663	663	663	663	663	663	663
Observations	9,914	9,742	9,248	9,086	8,599	8,447	7,966	7,824
F-test (1st stage)	27.63	25.16	29.34	26.89	28.18	26.31	34.31	33.03
	Dependent Variable (b): Log Total Instructional Expenditures							
Log State Appropriations	0.186***	0.187***	0.156***	0.158***	0.102***	0.102***	0.063***	0.062***
	(0.046)	(0.047)	(0.037)	(0.037)	(0.030)	(0.029)	(0.023)	(0.022)
N Institutions	663	663	663	663	663	663	663	663
Observations	9,914	9,742	9,248	9,086	8,599	8,447	7,966	7,824
F-test (1st stage)	27.63	25.16	29.34	26.89	28.18	26.31	34.31	33.03
	Dependent Variable (c): Log Academic Support Expenditures							
Log State Appropriations	0.272***	0.274***	0.221***	0.224***	0.164***	0.164***	0.120***	0.116***
	(0.067)	(0.069)	(0.054)	(0.053)	(0.046)	(0.045)	(0.037)	(0.035)
N Institutions	663	663	663	663	663	663	663	663
Observations	9,914	9,742	9,248	9,086	8,599	8,447	7,966	7,824
F-test (1st stage)	27.63	25.16	29.34	26.89	28.18	26.31	34.31	33.03
	Dependent Variable (d): Log Total Institutional Grant Aid							
Log State Appropriations	0.235*	0.284**	0.252*	0.320**	0.258**	0.314**	0.253**	0.286**
	(0.138)	(0.140)	(0.129)	(0.132)	(0.126)	(0.129)	(0.112)	(0.112)
N Institutions	663	663	663	663	663	663	663	663
Observations	9,204	9,047	8,604	8,456	8,013	7,875	7,444	7,316
F-test (1st stage)	24.67	22.19	25.33	22.86	23.87	22.14	30.36	28.83
Institution FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓	✓	✓
County Controls	×	✓	×	✓	×	✓	×	✓

NOTES.—Each panel and cell refers to a separate 2SLS regression. "Academic Support Expenditures" includes spending on tutoring, advising, mentoring, and other expenses related to activities and services that support the institution's primary missions. The models include year and institution fixed effects, as well as time-varying county and state covariates, as described in Table 1.2. A lead of "t+i" (ranging from 0 to 3) indicates the ith lead of the outcome variable. The unit of analysis is at the institution and year level. Standard errors are clustered at the institution level. *p < 0.10, **p < 0.05, ***p < 0.01

Table A.10: The Effect of State Appropriations on Number of Programs Offered

	t+0		t+1		t+2		t+3	
	Dependent Variable (a): Log N Distinct STEM Programs							
Log State appropriations	0.121***	0.126***	0.111***	0.117***	0.084***	0.089***	0.054**	0.058**
	(0.036)	(0.038)	(0.031)	(0.032)	(0.028)	(0.028)	(0.022)	(0.022)
N Institutions	617	617	617	617	617	617	617	617
Observations	9,228	9,061	8,679	8,521	8,129	7,980	7,577	7,437
F-test (1st stage)	25.16	22.52	26.79	24.20	25.64	23.47	32.06	30.33
	Dependent Variable (b): Log N Distinct Non-STEM Programs							
Log State Appropriations	0.071***	0.079***	0.071***	0.078***	0.056***	0.063***	0.037**	0.041**
	(0.026)	(0.027)	(0.023)	(0.024)	(0.021)	(0.022)	(0.017)	(0.017)
N Institutions	663	663	663	663	663	663	663	663
Observations	9,914	9,742	9,248	9,086	8,599	8,447	7,966	7,824
F-test (1st stage)	27.63	25.16	29.34	26.89	28.18	26.31	34.31	33.03
Institution FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓	✓	✓
County Controls	×	✓	×	✓	×	✓	×	✓

NOTES.—Each panel and cell refers to a separate 2SLS regression. "N Distinct STEM/Non-STEM Programs" refers to the institution-level number of distinct programs offered within Science, Technology, Engineering, and Mathematics (STEM) or otherwise (non-STEM). The models include year and institution fixed effects, as well as time-varying county and state covariates, as described in Table 1.2. A lead of "t+i" (ranging from 0 to 3) indicates the ith lead of the outcome variable. The unit of analysis is at the institution and year level. Standard errors are clustered at the institution level. *p < 0.10, **p < 0.05, ***p < 0.01

Table A.11: OLS Estimates for the Effect of Program Offers, Spending, and Prices on Number of Bachelor Degrees

Dept. var.	Log Number of STEM Degrees			Log Number of Non-STEM Degrees		
	t+4	t+5	t+6	t+4	t+5	t+6
Panel (a)						
Log N Distinct Programs	0.171***	0.129***	0.127***	0.066	0.021	-0.019
	(0.041)	(0.034)	(0.035)	(0.045)	(0.052)	(0.045)
Observations	6779	6226	5680	7209	6613	6015
Sensitivity analysis (Diegert et al., 2022)						
$\hat{r}_X^{bp}(\%)$	80.16	79.99	79.64	81.97	81.74	81.46
Panel (b)						
Log Total Expenditures	0.285***	0.227***	0.153***	0.174***	0.102**	0.024
	(0.055)	(0.058)	(0.053)	(0.050)	(0.051)	(0.050)
Observations	6785	6231	5683	7209	6613	6015
Sensitivity analysis (Diegert et al., 2022)						
$\hat{r}_X^{bp}(\%)$	73.72	73.58	73.15	72.85	72.88	72.63
Panel (c)						
Log Academic Support Expenditures	0.124***	0.106***	0.103***	0.039*	0.029	0.011
	(0.030)	(0.029)	(0.028)	(0.023)	(0.021)	(0.021)
Observations	6785	6231	5683	7209	6613	6015
Sensitivity analysis (Diegert et al., 2022)						
$\hat{r}_X^{bp}(\%)$	71.83	71.67	71.39	75.88	75.79	75.51
Panel (d)						
Log Instructional Expenditures	0.263***	0.206***	0.149***	0.115***	0.062	0.016
	(0.047)	(0.050)	(0.051)	(0.039)	(0.039)	(0.037)
Observations	6785	6231	5683	7209	6613	6015
Sensitivity analysis (Diegert et al., 2022)						
$\hat{r}_X^{bp}(\%)$	32.12	32.04	31.80	36.86	36.76	36.52
Panel (e)						
Log Institutional Grants	0.037***	0.023**	0.019**	0.010	0.005	0.003
	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
Observations	6784	6230	5682	7208	6612	6014
Sensitivity analysis (Diegert et al., 2022)						
$\hat{r}_X^{bp}(\%)$	63.84	63.40	62.98	60.08	59.33	58.34
Panel (f)						
Log Sticker Price	-0.134***	-0.115**	-0.046	-0.063	-0.053	-0.008
	(0.052)	(0.052)	(0.052)	(0.046)	(0.049)	(0.049)
Observations	6597	6057	5523	6932	6355	5779
Sensitivity analysis (Diegert et al., 2022)						
$\hat{r}_X^{bp}(\%)$	41.98	41.88	41.74	13.47	13.46	13.74
Panel (g)						
Log Net Price	-0.040	-0.034	0.002	-0.018	-0.001	0.003
	(0.029)	(0.028)	(0.033)	(0.028)	(0.030)	(0.031)
Observations	4593	4054	3527	4871	4295	3725
Sensitivity analysis (Diegert et al., 2022)						
$\hat{r}_X^{bp}(\%)$	37.53	36.98	35.92	7.24	7.75	7.61
N Institutions	663	663	663	617	617	617
Institution FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓
County Controls	✓	✓	✓	✓	✓	✓

NOTES.—Each panel and cell refers to a separate OLS regression. The outcome variable in the first three columns from the left is the logged number of undergraduate Science, Technology, Engineering, and Mathematics (STEM) degrees (bachelor's) conferred. For the last three columns, the dependent variable is the logged number of non-STEM bachelor's degrees. "N Distinct Programs" refers to the institution-level number of distinct programs offered within STEM for the first three columns and within non-STEM for the last three columns. "Academic Support Expenditures" includes spending on tutoring, advising, mentoring, and other expenses related to activities and services that support the institution's primary missions. The models include year and institution fixed effects, as well as time-varying county and state covariates, as described in Table 1.2. A lead of "t+i" (ranging from 3 to 6) indicates the ith lead of the outcome variable. For example, the lead t+6 refers to the student cohort that completes bachelor's education in 6 years. The unit of analysis is at the institution and year level. Standard errors are clustered at the institution level. *p < 0.10, **p < 0.05, ***p < 0.01

Table A.12: The Effect of State Appropriations on Net and Sticker Prices

	t+0		t+1		t+2		t+3	
	Dependent Variable (a): Log Sticker Price							
Log State Appropriations	-0.104***	-0.108***	-0.078***	-0.083***	-0.058***	-0.064***	-0.027**	-0.032**
	(0.029)	(0.030)	(0.022)	(0.022)	(0.018)	(0.019)	(0.013)	(0.013)
N Institutions	663	663	663	663	663	663	663	663
Observations	9,543	9,374	8,909	8,749	8,291	8,141	7,688	7,548
F-test (1st stage)	25.48	23.96	27.43	25.74	26.81	25.59	32.73	32.14
	Dependent Variable (b): Log Net Price							
Log State Appropriations	-0.181***	-0.208***	-0.132***	-0.162***	-0.070**	-0.092**	-0.034	-0.048**
	(0.063)	(0.069)	(0.049)	(0.055)	(0.035)	(0.036)	(0.023)	(0.023)
N Institutions	663	663	663	663	663	663	663	663
Observations	7,429	7,299	7,326	7,195	7,236	7,104	7,154	7,023
F-test (1st stage)	22.18	22.29	19.90	19.33	14.69	15.35	23.66	23.62
Institution FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓	✓	✓
County Controls	×	✓	×	✓	×	✓	×	✓

NOTES.—Each panel and cell refers to a separate 2SLS regression. The models include year and institution fixed effects, as well as time-varying county and state covariates, as described in Table 1.2. "Sticker Price" refers to the institution's published tuition and fees, whereas "Net Price" refers to the total college attendance cost minus all grants and aid. A lead of "t+i" (ranging from 3 to 6) indicates the ith lead of the outcome variable. For example, the lead t+6 refers to the student cohort that completes bachelor's education in 6 years. The unit of analysis is at the institution and year level. Standard errors are clustered at the institution level. *p < 0.10, **p < 0.05, ***p < 0.01

Table A.13: OLS Estimates for the Effect of Program Offers, Spending, and Prices on Bachelor's Degrees (Alternative Sensitivity Analysis)

Dept. var.	Log Number of STEM Degrees			Log Number of Non-STEM Degrees		
	t+4	t+5	t+6	t+4	t+5	t+6
Panel (a)						
Log N Distinct Programs	0.171***	0.129***	0.127***	0.066	0.021	-0.019
	(0.041)	(0.034)	(0.035)	(0.045)	(0.052)	(0.045)
Observations	6779	6226	5680	7209	6613	6015
Sensitivity analysis (Oster, 2019)						
$\hat{\delta}^{bp}_{resid}(\%)$	76.25	75.96	75.29	59.96	59.16	58.75
Panel (b)						
Log Total Expenditures	0.285***	0.227***	0.153***	0.174***	0.102**	0.024
	(0.055)	(0.058)	(0.053)	(0.050)	(0.051)	(0.050)
Observations	6785	6231	5683	7209	6613	6015
Sensitivity analysis (Oster, 2019)						
$\hat{\delta}^{bp}_{resid}(\%)$	62.92	62.80	62.03	58.74	58.76	58.48
Panel (c)						
Log Academic Support Expenditures	0.124***	0.106***	0.103***	0.039*	0.029	0.011
	(0.030)	(0.029)	(0.028)	(0.023)	(0.021)	(0.021)
Observations	6785	6231	5683	7209	6613	6015
Sensitivity analysis (Oster, 2019)						
$\hat{\delta}^{bp}_{resid}(\%)$	62.21	61.85	61.24	66.85	66.48	65.93
Panel (d)						
Log Instructional Expenditures	0.263***	0.206***	0.149***	0.115***	0.062	0.016
	(0.047)	(0.050)	(0.051)	(0.039)	(0.039)	(0.037)
Observations	6785	6231	5683	7209	6613	6015
Sensitivity analysis (Oster, 2019)						
$\hat{\delta}^{bp}_{resid}(\%)$	10.31	10.41	10.47	12.35	12.43	12.45
Panel (e)						
Log Institutional Grants	0.037***	0.023**	0.019**	0.010	0.005	0.003
	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
Observations	6784	6230	5682	7208	6612	6014
Sensitivity analysis (Oster, 2019)						
$\hat{\delta}^{bp}_{resid}(\%)$	50.86	52.03	54.52	48.76	48.83	49.36
Panel (f)						
Log Sticker Price	-0.134***	-0.115**	-0.046	-0.063	-0.053	-0.008
	(0.052)	(0.052)	(0.052)	(0.046)	(0.049)	(0.049)
Observations	6597	6057	5523	6932	6355	5779
Sensitivity analysis (Oster, 2019)						
$\hat{\delta}^{bp}_{resid}(\%)$	158.72	164.00	171.21	-315.21	-304.42	-282.41
Panel (g)						
Log Net Price	-0.040	-0.034	0.002	-0.018	-0.001	0.003
	(0.029)	(0.028)	(0.033)	(0.028)	(0.030)	(0.031)
Observations	4593	4054	3527	4871	4295	3725
Sensitivity analysis (Oster, 2019)						
$\hat{\delta}^{bp}_{resid}(\%)$	157.61	176.26	164.69	-50.64	-52.71	-71.74
Institution FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓
County Controls	✓	✓	✓	✓	✓	✓

NOTES.—Each panel refers to a separate OLS regression. The outcome variable in the first three columns from the left is the logged number of undergraduate Science, Technology, Engineering, and Mathematics (STEM) degrees (bachelor's) conferred. For the last three columns, the dependent variable is the logged number of non-STEM bachelor's degrees. "N Distinct Programs" refers to the institution-level number of distinct programs offered within STEM for the first three columns and within non-STEM for the last three columns. "Academic Support Expenditures" includes spending on tutoring, advising, mentoring, and other expenses related to activities and services that support the institution's primary missions. The models include year and institution fixed effects, as well as time-varying county and state covariates, as described in Table 1.2. A lead of "t+i" (ranging from 3 to 6) indicates the ith lead of the outcome variable. For example, the lead t+6 refers to the student cohort that completes bachelor's education in 6 years. The unit of analysis is at the institution and year level. Standard errors are clustered at the institution level. *p < 0.10, **p < 0.05, ***p < 0.01

Table A.14: The Effect of State Appropriations on Bachelor's Degrees, by Institution Selectivity and Degree Type

	Dependent Variable: Log Number of Bachelor Degrees					
	STEM			Non-STEM		
	t+4	t+5	t+6	t+4	t+5	t+6
Panel (a): Non-Selective						
Log State Appropriations	0.293*	0.303**	0.260**	0.144	0.174	0.144
	(0.174)	(0.147)	(0.129)	(0.163)	(0.117)	(0.100)
Observations	71,632	64,588	57,893	210,512	189,302	169,013
F-test (1st stage)	95.88	155.87	137.46	327.27	547.92	503.42
Panel (b): Selective						
Log State Appropriations	0.225	0.182	0.105	0.281	0.194	0.090
	(0.178)	(0.154)	(0.113)	(0.298)	(0.204)	(0.139)
Observations	12,630	11,431	10,264	24,828	22,379	20,013
F-test (1st stage)	87.57	72.88	45.97	117.25	143.20	118.65
N Institutions	663	663	663	617	617	617
N Programs	339	339	339	783	783	783
Institution FE	✓	✓	✓	✓	✓	✓
Program FE (6-Digit CIP)	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓
County Controls	✓	✓	✓	✓	✓	✓

NOTES.—Each panel and cell refers to a separate 2SLS regression. The outcome variable is the logged number of undergraduate degrees (bachelor's) conferred for Science, Technology, Engineering, and Mathematics (STEM) majors in the first three columns and for non-STEM majors in the last three columns. A lead of "t+i" (ranging from 4 to 6) indicates the ith lead of the outcome variable. The models include year, program, and institution fixed effects, as well as time-varying county and state covariates, as described in Table 1.2. Selectivity is defined based on the Carnegie classification and refers to the institutions that are in the top 20th percentile of selectivity among all baccalaureate institutions with fewer than 20 percent of entering undergraduate transfers. Standard errors are clustered at the institution level. Based on this classification, there are 45 selective 4-year institutions. Table A.19 shows similar results based on the selectivity definition of AAU membership or being listed as competitive in Barron's Profiles of American Colleges. *p < 0.10, **p < 0.05, ***p < 0.01

Table A.15: The Effect of State Appropriations on Non-STEM Bachelor's Degrees, by Program Type

	Dependent Variable: Log Number of Non-STEM Bachelor Degrees							
	t+3		t+4		t+5		t+6	
Panel (a): Education								
Log State Appropriations	1.483	1.436	1.245	1.147	1.206	1.081	0.705	0.524
	(1.045)	(1.062)	(0.865)	(0.843)	(0.739)	(0.721)	(0.560)	(0.537)
Observations	38,730	38,631	34,910	34,819	31,269	31,190	27,880	27,811
F-test (1st stage)	532.71	502.10	423.43	416.51	384.86	389.95	301.21	311.92
Panel (b): Business, Management, Marketing, and Related Support Services								
Log State Appropriations	0.314	0.358	0.340	0.367	0.201	0.211	0.139	0.150
	(0.331)	(0.323)	(0.266)	(0.254)	(0.179)	(0.164)	(0.153)	(0.138)
Observations	34,122	33,699	31,044	30,653	28,085	27,727	25,226	24,902
F-test (1st stage)	29.28	30.85	43.42	47.90	70.38	80.03	66.75	77.21
Panel (c): Visual and Performing Arts								
Log State Appropriations	-0.057	-0.126	-0.012	-0.073	0.023	-0.065	0.063	-0.019
	(0.253)	(0.278)	(0.210)	(0.226)	(0.154)	(0.163)	(0.143)	(0.147)
Observations	31,390	30,968	28,543	28,151	25,802	25,442	23,151	22,822
F-test (1st stage)	43.49	35.14	57.66	47.67	91.83	78.60	79.36	67.99
Panel (d): Social Sciences								
Log State Appropriations	0.052	0.025	0.015	0.0004	0.078	0.062	0.089	0.069
	(0.128)	(0.129)	(0.103)	(0.106)	(0.073)	(0.073)	(0.072)	(0.071)
Observations	27,934	27,322	25,713	25,146	23,547	23,025	21,409	20,934
F-test (1st stage)	39.53	38.69	54.21	49.92	98.16	90.33	90.71	85.18
Panel (e): Health Professions and Related Programs								
Log State Appropriations	-0.452	-0.419	-0.307	-0.301	-0.101	-0.109	0.177	0.148
	(0.325)	(0.319)	(0.248)	(0.247)	(0.187)	(0.184)	(0.220)	(0.207)
Observations	19,735	19,543	17,187	17,018	14,805	14,659	12,584	12,459
F-test (1st stage)	34.93	38.15	53.23	55.77	82.24	82.92	59.55	59.90
Panel (f): Others								
Log State Appropriations	0.293	0.296	0.369**	0.365*	0.308**	0.291**	0.264**	0.231**
	(0.214)	(0.239)	(0.188)	(0.202)	(0.133)	(0.131)	(0.116)	(0.108)
Observations	110,370	108,795	100,113	98,666	90,192	88,871	80,638	79,441
F-test (1st stage)	124.58	103.90	172.15	143.75	310.19	269.71	282.76	255.29
Institution FE	✓	✓	✓	✓	✓	✓	✓	✓
Program FE (6-Digit CIP)	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓	✓	✓
County Controls	×	✓	×	✓	×	✓	×	✓

Note: Each panel refers to a separate 2SLS regression and includes a subset of all sub-majors within the given global major 2-digit CIP code. The outcome variable is the logged number of undergraduate degrees (bachelor's) conferred for non-Science, Technology, Engineering, and Mathematics (non-STEM) majors. See Table A.22 for the included majors in panel (f). A lead of "t+i" (ranging from 4 to 6) indicates the ith lead of the outcome variable. The models include year, program, and institution fixed effects, as well as time-varying county and state covariates, as described in Table 1.2. Standard errors are clustered at the institution level. *p < 0.10, **p < 0.05, ***p < 0.01

Table A.16: Impact of State Appropriations on Bachelor STEM Degrees by Race and Major (Total)

	Dependent Variable: Log Number of Bachelor Degrees											
	ALL STEM Majors			Engineering			Physical Sciences			Bio. & Biomed.[a]		
	t+4	t+5	t+6	t+4	t+5	t+6	t+4	t+5	t+6	t+4	t+5	t+6
Panel A: White												
Log State Appropriations	0.172	0.204*	0.191*	0.144	0.253	0.192	0.156	0.082	0.034	0.436	0.462*	0.488**
	(0.155)	(0.120)	(0.105)	(0.206)	(0.206)	(0.168)	(0.151)	(0.103)	(0.102)	(0.388)	(0.267)	(0.237)
Panel B: Black												
Log State Appropriations	0.245	0.130	0.130	0.840*	0.583*	0.248	0.378	0.194	0.167	0.327	0.336	0.288
	(0.179)	(0.124)	(0.118)	(0.484)	(0.338)	(0.261)	(0.254)	(0.173)	(0.162)	(0.407)	(0.287)	(0.278)
Panel C: Hispanic												
Log State Appropriations	0.248	0.212	0.013	-0.015	0.337	0.044	0.184	0.226	0.038	0.232	0.065	-0.105
	(0.200)	(0.153)	(0.130)	(0.388)	(0.348)	(0.266)	(0.264)	(0.200)	(0.167)	(0.412)	(0.285)	(0.286)
Panel D: Asian												
Log State Appropriations	0.035	0.141	0.147	0.019	0.296	0.285	-0.073	0.052	-0.108	0.412	0.111	0.261
	(0.184)	(0.140)	(0.129)	(0.313)	(0.277)	(0.244)	(0.195)	(0.145)	(0.147)	(0.443)	(0.293)	(0.289)
Panel E: American Indian												
Log State Appropriations	0.073	0.141	0.209**	-0.024	0.148	0.368*	-0.117	-0.0007	-0.003	0.481	0.496*	0.408
	(0.111)	(0.095)	(0.101)	(0.195)	(0.203)	(0.215)	(0.129)	(0.081)	(0.083)	(0.366)	(0.287)	(0.282)
Observations	84,262	76,019	68,157	16,842	15,282	13,786	16,402	15,003	13,619	13,846	12,494	11,203
F-test (1st stage)	105.10	167.60	159.80	16.01	18.90	22.35	28.88	53.17	50.29	13.64	24.21	21.82
Institution FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Program FE (6-Digit CIP)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
County Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

[a] Bio. & Biomed. stands for Biological and Biomedical Sciences.

Notes: Each cell represents a separate 2SLS regression. The dependent variable is the log number of bachelor degrees. A lead of "t+i" (ranging from 4 to 6) indicates the ith lead of the outcome variable. The models include year, program, and institution fixed effects, as well as time-varying county and state covariates, as described in Table 1.2. Standard errors, clustered at the institution level, are in parentheses. ✓ indicates the inclusion of the respective fixed effects or controls. *p < 0.10, **p < 0.05, ***p < 0.01

Table A.17: Impact of State Appropriations on Bachelor STEM Degrees by Race and Major (Men)

	Dependent Variable: Log Number of Bachelor Degrees											
	ALL STEM Majors			Engineering			Physical Sciences			Bio. & Biomed.[a]		
	t+4	t+5	t+6	t+4	t+5	t+6	t+4	t+5	t+6	t+4	t+5	t+6
Panel A: White												
Log State Appropriations	0.195	0.229*	0.172*	0.164	0.206	0.084	0.035	0.006	0.054	0.587	0.605**	0.334
	(0.154)	(0.124)	(0.102)	(0.208)	(0.192)	(0.140)	(0.175)	(0.122)	(0.116)	(0.373)	(0.271)	(0.222)
Panel B: Black												
Log State Appropriations	0.256	0.187	0.147	0.843*	0.676*	0.356	0.381*	0.168	0.128	0.266	0.221	0.319
	(0.170)	(0.122)	(0.114)	(0.498)	(0.367)	(0.279)	(0.230)	(0.161)	(0.144)	(0.375)	(0.283)	(0.270)
Panel C: Hispanic												
Log State Appropriations	0.246	0.223	0.068	-0.028	0.204	-0.102	0.403	0.325	0.051	0.495	0.279	0.199
	(0.192)	(0.143)	(0.123)	(0.358)	(0.313)	(0.242)	(0.274)	(0.197)	(0.160)	(0.445)	(0.291)	(0.310)
Panel D: Asian												
Log State Appropriations	-0.008	0.142	0.204*	0.022	0.304	0.369	-0.094	0.036	-0.058	0.177	0.230	0.270
	(0.175)	(0.130)	(0.123)	(0.302)	(0.284)	(0.258)	(0.199)	(0.139)	(0.138)	(0.429)	(0.306)	(0.289)
Panel E: American Indian												
Log State Appropriations	0.065	0.130	0.205**	0.068	0.193	0.410*	-0.058	0.018	0.042	0.117	0.106	0.282
	(0.088)	(0.081)	(0.089)	(0.187)	(0.212)	(0.229)	(0.105)	(0.064)	(0.075)	(0.255)	(0.196)	(0.200)
Observations	84,262	76,019	68,157	16,842	15,282	13,786	16,402	15,003	13,619	13,846	12,494	11,203
F-test (1st stage)	105.10	167.60	159.80	16.01	18.90	22.35	28.88	53.17	50.29	13.64	24.21	21.82
Institution FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Program FE (6-Digit CIP)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
County Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

[a] Bio. & Biomed. stands for Biological and Biomedical Sciences.

Notes: Each cell represents a separate 2SLS regression. The dependent variable is the log number of bachelor degrees. A lead of "t+i" (ranging from 4 to 6) indicates the ith lead of the outcome variable. The models include year, program, and institution fixed effects, as well as time-varying county and state covariates. Standard errors, clustered at the institution level, are in parentheses. *p < 0.10, **p < 0.05, ***p < 0.01

Table A.18: Impact of State Appropriations on Bachelor STEM Degrees by Race and Major (Women)

	Dependent Variable: Log Number of Bachelor Degrees											
	ALL STEM Majors			Engineering			Physical Sciences			Bio. & Biomed.[a]		
	t+4	t+5	t+6	t+4	t+5	t+6	t+4	t+5	t+6	t+4	t+5	t+6
Panel A: White												
Log State Appropriations	-0.129	-0.044	-0.038	0.207	0.223	0.224	-0.134	-0.106	-0.176	0.179	0.252	0.432*
	(0.170)	(0.118)	(0.106)	(0.291)	(0.275)	(0.237)	(0.217)	(0.159)	(0.150)	(0.371)	(0.262)	(0.251)
Panel B: Black												
Log State Appropriations	0.049	0.022	0.053	0.034	0.024	0.137	0.099	0.090	0.167	0.490	0.356	0.175
	(0.147)	(0.104)	(0.097)	(0.288)	(0.238)	(0.228)	(0.181)	(0.138)	(0.145)	(0.413)	(0.294)	(0.278)
Panel C: Hispanic												
Log State Appropriations	0.031	0.107	-0.042	-0.284	0.309	0.073	-0.026	-0.012	-0.101	0.073	0.019	-0.237
	(0.163)	(0.128)	(0.122)	(0.329)	(0.304)	(0.260)	(0.216)	(0.164)	(0.167)	(0.431)	(0.286)	(0.314)
Panel D: Asian												
Log State Appropriations	0.078	0.094	0.023	0.033	-0.054	-0.192	0.032	0.089	-0.035	0.654	0.376	0.303
	(0.177)	(0.130)	(0.117)	(0.370)	(0.265)	(0.262)	(0.178)	(0.130)	(0.132)	(0.455)	(0.315)	(0.299)
Panel E: American Indian												
Log State Appropriations	0.025	0.037	0.009	-0.128	-0.008	-0.068	-0.060	-0.025	-0.056	0.537*	0.448*	0.162
	(0.067)	(0.051)	(0.050)	(0.126)	(0.077)	(0.098)	(0.082)	(0.065)	(0.061)	(0.323)	(0.245)	(0.228)
Observations	84,262	76,019	68,157	16,842	15,282	13,786	16,402	15,003	13,619	13,846	12,494	11,203
F-test (1st stage)	105.10	167.60	159.80	16.01	18.90	22.35	28.88	53.17	50.29	13.64	24.21	21.82
Institution FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Program FE (6-Digit CIP)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
County Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

[a] Bio. & Biomed. stands for Biological and Biomedical Sciences.

Notes: Each cell represents a separate 2SLS regression. The dependent variable is the log number of bachelor degrees. A lead of "t+i" (ranging from 4 to 6) indicates the ith lead of the outcome variable. The models include year, program, and institution fixed effects, as well as time-varying county and state covariates. Standard errors, clustered at the institution level, are in parentheses. *p < 0.10, **p < 0.05, ***p < 0.01

Table A.19: Impact of State Appropriations on Bachelor's Degrees by Institution Selectivity and Degree Type (Alternative Selectivity Definition)

	Dependent Variable: Log Number of Bachelor Degrees					
	STEM			Non-STEM		
	t+4	t+5	t+6	t+4	t+5	t+6
Panel A: Non-Selective						
Log State Appropriations	0.365**	0.293**	0.274**	0.260	0.220*	0.188*
	(0.182)	(0.141)	(0.131)	(0.166)	(0.121)	(0.106)
Observations	68,003	61,331	54,970	202,727	182,291	162,710
F-test (1st stage)	101.61	163.15	143.24	350.29	574.95	526.24
Panel B: Selective						
Log State Appropriations	0.123	0.307	0.105	-0.359	-0.058	-0.279
	(0.271)	(0.328)	(0.190)	(0.779)	(0.318)	(0.267)
Observations	16,259	14,688	13,187	32,613	29,390	26,316
F-test (1st stage)	32.02	28.62	21.56	20.53	45.135	43.43
Institution FE	✓	✓	✓	✓	✓	✓
Program FE (6-Digit CIP)	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓
County Controls	✓	✓	✓	✓	✓	✓

Notes: Each panel and cell represent a separate 2SLS regression. The dependent variable is the log number of bachelor's degrees conferred for STEM majors in the first three columns and for non-STEM majors in the last three columns. A lead of "t+i" (ranging from 4 to 6) indicates the ith lead of the outcome variable. The models include year, program, and institution fixed effects, as well as time-varying county and state covariates, as described in Table 1.2. Selectivity is defined based on the membership in the Association of American Universities (AAU) or Barron's Selectivity (competitive universities). Based on this classification, there are 54 selective 4-year institutions. *p < 0.10, **p < 0.05, ***p < 0.01

Table A.20: Impact of State Appropriations on Bachelor's Degrees by Institution Doctoral Offering and Degree Type

	Dependent Variable: Log Number of Bachelor Degrees					
	STEM			Non-STEM		
	t+4	t+5	t+6	t+4	t+5	t+6
Panel A: Non-Doctoral Institutions						
Log State Appropriations	0.368**	0.262**	0.238**	0.226	0.189*	0.181*
	(0.167)	(0.113)	(0.102)	(0.144)	(0.107)	(0.100)
Observations	40,057	36,028	32,209	132,885	119,213	106,169
F-test (1st stage)	94.250	168.96	153.91	324.45	578.64	496.91
Panel B: Doctoral Institutions						
Log State Appropriations	0.139	0.254	0.147	-0.047	0.058	-0.059
	(0.295)	(0.279)	(0.208)	(0.349)	(0.217)	(0.169)
Observations	44,205	39,991	35,948	102,455	92,468	82,857
F-test (1st stage)	34.018	41.759	38.514	90.323	133.81	133.33
Institution FE	✓	✓	✓	✓	✓	✓
Program FE (6-Digit CIP)	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓
County Controls	✓	✓	✓	✓	✓	✓

Notes: Each panel and cell represent a separate 2SLS regression. The dependent variable is the log number of bachelor's degrees conferred for STEM majors in the first three columns and for non-STEM majors in the last three columns. A lead of "t+i" (ranging from 4 to 6) indicates the ith lead of the outcome variable. The models include year, program, and institution fixed effects, as well as time-varying county and state covariates, as described in Table 1.2. Doctoral institutions are defined by Carnegie classification as either Doctoral/Research Universities–Intensive or Doctoral/Research Universities–Extensive. *p < 0.10, **p < 0.05, ***p < 0.01

Table A.21: Impact of State Appropriations on Bachelor's Degrees by Institution Land-Grant Status and Degree Type

	Dependent Variable: Log Number of Bachelor Degrees					
	STEM			Non-STEM		
	t+4	t+5	t+6	t+4	t+5	t+6
Panel A: Non-Land-Grant Institutions						
Log State Appropriations	0.176	0.148*	0.128*	0.110	0.094	0.085
	(0.127)	(0.088)	(0.076)	(0.154)	(0.090)	(0.077)
Observations	60,637	54,635	48,912	189,753	170,522	152,096
F-test (1st stage)	106.36	195.61	175.03	314.10	605.22	539.00
Panel B: Land-Grant Institutions						
Log State Appropriations	1.19**	1.76*	1.46*	0.848	1.38*	0.889
	(0.553)	(0.923)	(0.839)	(0.691)	(0.724)	(0.614)
Observations	23,625	21,384	19,245	45,587	41,159	36,930
F-test (1st stage)	21.626	10.862	8.0561	76.348	44.162	33.367
Institution FE	✓	✓	✓	✓	✓	✓
Program FE (6-Digit CIP)	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓
County Controls	✓	✓	✓	✓	✓	✓

Notes: Each panel and cell represent a separate 2SLS regression. The dependent variable is the log number of bachelor's degrees conferred for STEM majors in the first three columns and for non-STEM majors in the last three columns. A lead of "t+i" (ranging from 4 to 6) indicates the ith lead of the outcome variable. The models include year, program, and institution fixed effects, as well as time-varying county and state covariates, as described in Table 1.2. As defined in IPEDS documentation, Land-grant institutions are colleges or universities designated by state legislatures or Congress to receive benefits from the Morrill Acts of 1862 and 1890. These institutions were originally tasked with providing practical education in agriculture, military tactics, mechanic arts, and classical studies to the working class. *p < 0.10, **p < 0.05, ***p < 0.01

Table A.22: Frequency of Other Non-STEM Fields

Category	Count
FOREIGN LANGUAGES, LITERATURES, AND LINGUISTICS	23394
COMMUNICATION, JOURNALISM, AND RELATED PROGRAMS	14507
ENGLISH LANGUAGE AND LITERATURE/LETTERS	11770
AREA, ETHNIC, CULTURAL, GENDER, AND GROUP STUDIES	11370
PARKS, RECREATION, LEISURE, AND FITNESS STUDIES	9406
PSYCHOLOGY	8895
HISTORY	8301
LIBERAL ARTS AND SCIENCES, GENERAL STUDIES AND HUMANITIES	8203
PHILOSOPHY AND RELIGIOUS STUDIES	7633
PUBLIC ADMINISTRATION AND SOCIAL SERVICE PROFESSIONS	7277
FAMILY AND CONSUMER SCIENCES/HUMAN SCIENCES	7092
MULTI/INTERDISCIPLINARY STUDIES	7020
HOMELAND SECURITY, LAW ENFORCEMENT, FIREFIGHTING AND RELATED PROTECTIVE SERVICES	6526
AGRICULTURE, AGRICULTURE OPERATIONS, AND RELATED SCIENCES	5426
ARCHITECTURE AND RELATED SERVICES	3471
NATURAL RESOURCES AND CONSERVATION	1993
LEGAL PROFESSIONS AND STUDIES	1275
TRANSPORTATION AND MATERIALS MOVING	828
COMMUNICATIONS TECHNOLOGIES/TECHNICIANS AND SUPPORT SERVICES	742
MECHANIC AND REPAIR TECHNOLOGIES/TECHNICIANS	352
COMPUTER AND INFORMATION SCIENCES AND SUPPORT SERVICES	339
PERSONAL AND CULINARY SERVICES	214
CONSTRUCTION TRADES	186
LIBRARY SCIENCE	152
ENGINEERING TECHNOLOGIES AND ENGINEERING-RELATED FIELDS	93
PRECISION PRODUCTION	63
THEOLOGY AND RELIGIOUS VOCATIONS	32
MILITARY TECHNOLOGIES AND APPLIED SCIENCES	13

Note: The count is based on the number of observations within 6-digit CIP, institution, and year for each shown 2-digit CIP program category.

Table A.23: Top 10 Rotemberg Weight for Institutions

Weight	Institution
0.2835	COLORADO STATE UNIVERSITY
0.2172	MESA STATE COLLEGE
0.1940	UNIVERSITY OF COLORADO AT BOULDER
0.1924	METROPOLITAN STATE COLLEGE OF DENVER
0.1924	UNIVERSITY OF NORTHERN COLORADO
0.1904	UNIVERSITY OF COLORADO AT COLORADO SPRINGS
0.1889	ADAMS STATE COLLEGE
0.1882	FORT LEWIS COLLEGE
0.1877	WESTERN STATE COLLEGE OF COLORADO
0.0604	LINCOLN UNIVERSITY

Note: The Rotemberg weights identify the 4-year public institutions that carry the highest weight in the instrument for the identification (Goldsmith-Pinkham et al., 2020). Since the data does not include industries or countries of origin as replications work in Goldsmith-Pinkham et al. (2020), I use similar approach and compute the institution weights as follow: $\alpha_k = \frac{\gamma_k \beta}{\sum_k \gamma_k \beta}$ where α_k refers to the Rotemberg weight for observation k, γ_k to the covariance between the endogenous variable x_k and the instrument z_k, and β to the first-stage coefficient on the Bartik instrument.

Table A.24: Impact of State Appropriations on Bachelor's Degrees (Excluding Top 10 Rotemberg Weight Institutions)

	Dependent Variable: Log Number of Bachelor Degrees					
	t+4		t+5		t+6	
Panel (a): STEM						
Log State Appropriations	0.467**	0.504**	0.595**	0.642**	0.459**	0.476**
	(0.232)	(0.256)	(0.240)	(0.273)	(0.207)	(0.235)
Observations	84,193	82,881	75,963	74,776	68,113	67,042
F-test (1st stage)	127.45	104.81	105.04	78.781	110.16	76.334
Panel (a): Non-STEM						
Log State Appropriations	0.290	0.269	0.371*	0.330	0.328*	0.263
	(0.237)	(0.257)	(0.213)	(0.226)	(0.199)	(0.209)
Observations	235,122	232,079	211,523	208,750	188,913	186,406
F-test (1st stage)	789.16	687.35	686.48	590.32	563.23	472.38
Institution FE	✓	✓	✓	✓	✓	✓
Program FE (6-Digit CIP)	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
State Controls	✓	✓	✓	✓	✓	✓
County Controls	×	✓	×	✓	×	✓

NOTES.—Each panel and cell refers to a separate 2SLS regression. The outcome variable in panel (a) is the logged number of undergraduate Science, Technology, Engineering, and Mathem
atics (STEM) degrees (bachelor's) conferred. In panel (b), the dependent variable is the logged number of non-STEM bachelor's degrees. A lead of "t+i" (ranging from 3 to 6) indicates the ith lead of the outcome variable. For example, the lead t+6 refers to the student cohort that completes bachelor's education in 6 years. CIP refers to the Classification of Instructional Programs, which is used to categorize degree programs by specific majors (see Table A.25 for the subset of STEM majors). The Table replicates the main findings in Table 1.2 after omitting the top 10 institutions that have the highest Rotemberg weights shown in Table A.23. Standard errors are clustered at the institution level. *p < 0.10, **p < 0.05, ***p < 0.01

Table A.25: DHS STEM Designated Degree Programs

CIP Code Two-Digit Series	2010 CIP Code	CIP Code Title
01	01.0308	Agroecology and Sustainable Agriculture.
01	01.0901	Animal Sciences, General.
01	01.0902	Agricultural Animal Breeding.
01	01.0903	Animal Health.
01	01.0904	Animal Nutrition.
01	01.0905	Dairy Science.
01	01.0906	Livestock Management.
01	01.0907	Poultry Science.
01	01.0999	Animal Sciences, Other.
01	01.1001	Food Science.
01	01.1002	Food Technology and Processing.
01	01.1099	Food Science and Technology, Other.
01	01.1101	Plant Sciences, General.
01	01.1102	Agronomy and Crop Science.
01	01.1103	Horticultural Science.
01	01.1104	Agricultural and Horticultural Plant Breeding.
01	01.1105	Plant Protection and Integrated Pest Management.
01	01.1106	Range Science and Management.
01	01.1199	Plant Sciences, Other.
01	01.1201	Soil Science and Agronomy, General.
01	01.1202	Soil Chemistry and Physics.
01	01.1203	Soil Microbiology.
01	01.1299	Soil Sciences, Other.
03	03.0101	Natural Resources/Conservation, General.
03	03.0103	Environmental Studies.
03	03.0104	Environmental Science.

Table A.25: (continued)

03	03.0199	Natural Resources Conservation and Research, Other.
03	03.0205	Water, Wetlands, and Marine Resources Management.
03	03.0502	Forest Sciences and Biology.
03	03.0508	Urban Forestry.
03	03.0509	Wood Science and Wood Products/Pulp and Paper Technology.
03	03.0601	Wildlife, Fish and Wildlands Science and Management.
04	04.0902	Architectural and Building Sciences/Technology.
09	09.0702	Digital Communication and Media/Multimedia.
10	10.0304	Animation, Interactive Technology, Video Graphics and Special Effects.
11	11.0101	Computer and Information Sciences, General.
11	11.0102	Artificial Intelligence.
11	11.0103	Information Technology.
11	11.0104	Informatics.
11	11.0199	Computer and Information Sciences, Other.
11	11.0201	Computer Programming/Programmer, General.
11	11.0202	Computer Programming, Specific Applications.
11	11.0203	Computer Programming, Vendor/Product Certification.
11	11.0299	Computer Programming, Other.
11	11.0301	Data Processing and Data Processing Technology/Technician.
11	11.0401	Information Science/Studies.
11	11.0501	Computer Systems Analysis/Analyst.
11	11.0701	Computer Science.
11	11.0801	Web Page, Digital/Multimedia and Information Resources Design.
11	11.0802	Data Modeling/Warehousing and Database Administration.
11	11.0803	Computer Graphics.
11	11.0804	Modeling, Virtual Environments and Simulation.
11	11.0899	Computer Software and Media Applications, Other.
11	11.0901	Computer Systems Networking and Telecommunications.
11	11.1001	Network and System Administration/Administrator.
11	11.1002	System, Networking, and LAN/WAN Management/Manager.

Table A.25: (continued)

11	11.1003	Computer and Information Systems Security/Information Assurance.
11	11.1004	Web/Multimedia Management and Webmaster.
11	11.1005	Information Technology Project Management.
11	11.1006	Computer Support Specialist.
11	11.1099	Computer/Information Technology Services Administration and Management, Other.
13	13.0501	Educational/Instructional Technology.
13	13.0601	Educational Evaluation and Research.
13	13.0603	Educational Statistics and Research Methods.
14	14.XXXX	Engineering.
15	15.0000	Engineering Technology, General.
15	15.0101	Architectural Engineering Technology/Technician.
15	15.0201	Civil Engineering Technology/Technician.
15	15.0303	Electrical, Electronic and Communications Engineering. Technology/Technician.
15	15.0304	Laser and Optical Technology/Technician.
15	15.0305	Telecommunications Technology/Technician.
15	15.0306	Integrated Circuit Design.
15	15.0399	Electrical and Electronic Engineering Technologies/Technicians, Other.
15	15.0401	Biomedical Technology/Technician.
15	15.0403	Electromechanical Technology/Electromechanical Engineering Technology.
15	15.0404	Instrumentation Technology/Technician.
15	15.0405	Robotics Technology/Technician.
15	15.0406	Automation Engineer Technology/Technician.
15	15.0499	Electromechanical and Instrumentation and Maintenance Technologies/Technicians, Other.
15	15.0501	Heating, Ventilation, Air Conditioning and Refrigeration Engineering Technology/Technician.
15	15.0503	Energy Management and Systems Technology/Technician.
15	15.0505	Solar Energy Technology/Technician.

Table A.25: (continued)

15	15.0506	Water Quality and Wastewater Treatment Management and Recycling Technology/Technician.
15	15.0507	Environmental Engineering Technology/Environmental Technology.
15	15.0508	Hazardous Materials Management and Waste Technology/Technician.
15	15.0599	Environmental Control Technologies/Technicians, Other.
15	15.0607	Plastics and Polymer Engineering Technology/Technician.
15	15.0611	Metallurgical Technology/Technician.
15	15.0612	Industrial Technology/Technician.
15	15.0613	Manufacturing Engineering Technology/Technician
15	15.0614	Welding Engineering Technology/Technician.
15	15.0615	Chemical Engineering Technology/Technician.
15	15.0616	Semiconductor Manufacturing Technology.
15	15.0699	Industrial Production Technologies/Technicians, Other.
15	15.0701	Occupational Safety and Health Technology/Technician.
15	15.0702	Quality Control Technology/Technician.
15	15.0703	Industrial Safety Technology/Technician.
15	15.0704	Hazardous Materials Information Systems Technology/Technician.
15	15.0799	Quality Control and Safety Technologies/Technicians, Other.
15	15.0801	Aeronautical/Aerospace Engineering Technology/Technician.
15	15.0803	Automotive Engineering Technology/Technician.
15	15.0805	Mechanical Engineering/Mechanical Technology/Technician.
15	15.0899	Mechanical Engineering Related Technologies/Technicians, Other.
15	15.0901	Mining Technology/Technician.
15	15.0903	Petroleum Technology/Technician.
15	15.0999	Mining and Petroleum Technologies/Technicians, Other.
15	15.1001	Construction Engineering Technology/Technician.
15	15.1102	Surveying Technology/Surveying.
15	15.1103	Hydraulics and Fluid Power Technology/Technician.
15	15.1199	Engineering-Related Technologies, Other.
15	15.1201	Computer Engineering Technology/Technician.

Table A.25: (continued)

15	15.1202	Computer Technology/Computer Systems Technology.
15	15.1203	Computer Hardware Technology/Technician.
15	15.1204	Computer Software Technology/Technician.
15	15.1299	Computer Engineering Technologies/Technicians, Other.
15	15.1301	Drafting and Design Technology/Technician, General.
15	15.1302	CAD/CADD Drafting and/or Design Technology/Technician.
15	15.1303	Architectural Drafting and Architectural CAD/CADD.
15	15.1304	Civil Drafting and Civil Engineering CAD/CADD.
15	15.1305	Electrical/Electronics Drafting and Electrical/Electronics CAD/CADD.
15	15.1306	Mechanical Drafting and Mechanical Drafting CAD/CADD.
15	15.1399	Drafting/Design Engineering Technologies/Technicians, Other.
15	15.1401	Nuclear Engineering Technology/Technician.
15	15.1501	Engineering/Industrial Management.
15	15.1502	Engineering Design.
15	15.1503	Packaging Science.
15	15.1599	Engineering-Related Fields, Other.
15	15.1601	Nanotechnology.
15	15.9999	Engineering Technologies and Engineering-Related Fields, Other.
26	26.XXXX	Biological and Biomedical Sciences.
27	27.XXXX	Mathematics and Statistics.
28	28.0501	Air Science/Airpower Studies.
28	28.0502	Air and Space Operational Art and Science.
28	28.0505	Naval Science and Operational Studies.
29	29.0201	Intelligence, General.
29	29.0202	Strategic Intelligence.
29	29.0203	Signal/Geospatial Intelligence.
29	29.0204	Command & Control (C3, C4I) Systems and Operations.
29	29.0205	Information Operations/Joint Information Operations.
29	29.0206	Information/Psychological Warfare and Military Media Relations.
29	29.0207	Cyber/Electronic Operations and Warfare.

Table A.25: (continued)

29	29.0299	Intelligence, Command Control and Information Operations, Other.
29	29.0301	Combat Systems Engineering.
29	29.0302	Directed Energy Systems.
29	29.0303	Engineering Acoustics.
29	29.0304	Low-Observables and Stealth Technology.
29	29.0305	Space Systems Operations.
29	29.0306	Operational Oceanography.
29	29.0307	Undersea Warfare.
29	29.0399	Military Applied Sciences, Other.
29	29.0401	Aerospace Ground Equipment Technology.
29	29.0402	Air and Space Operations Technology.
29	29.0403	Aircraft Armament Systems Technology.
29	29.0404	Explosive Ordinance/Bomb Disposal.
29	29.0405	Joint Command/Task Force (C3, C4I) Systems.
29	29.0406	Military Information Systems Technology.
29	29.0407	Missile and Space Systems Technology.
29	29.0408	Munitions Systems/Ordinance Technology.
29	29.0409	Radar Communications and Systems Technology.
29	29.0499	Military Systems and Maintenance Technology, Other.
29	29.9999	Military Technologies and Applied Sciences, Other.
30	30.0101	Biological and Physical Sciences.
30	30.0601	Systems Science and Theory.
30	30.0801	Mathematics and Computer Science.
30	30.1001	Biopsychology.
30	30.1701	Behavioral Sciences.
30	30.1801	Natural Sciences.
30	30.1901	Nutrition Sciences.
30	30.2501	Cognitive Science.
30	30.2701	Human Biology.
30	30.3001	Computational Science.

Table A.25: (continued)

30	30.3101	Human Computer Interaction.
30	30.3201	Marine Sciences.
30	30.3301	Sustainability Studies.
40	40.XXXX	Physical Sciences.
41	41.0000	Science Technologies/Technicians, General.
41	41.0101	Biology Technician/Biotechnology Laboratory Technician.
41	41.0204	Industrial Radiologic Technology/Technician.
41	41.0205	Nuclear/Nuclear Power Technology/Technician.
41	41.0299	Nuclear and Industrial Radiologic Technologies/Technicians, Other.
41	41.0301	Chemical Technology/Technician.
41	41.0303	Chemical Process Technology.
41	41.0399	Physical Science Technologies/Technicians, Other.
41	41.9999	Science Technologies/Technicians, Other.
42	42.2701	Cognitive Psychology and Psycholinguistics.
42	42.2702	Comparative Psychology.
42	42.2703	Developmental and Child Psychology.
42	42.2704	Experimental Psychology.
42	42.2705	Personality Psychology.
42	42.2706	Physiological Psychology/Psychobiology.
42	42.2707	Social Psychology.
42	42.2708	Psychometrics and Quantitative Psychology.
42	42.2709	Psychopharmacology.
42	42.2799	Research and Experimental Psychology, Other.
43	43.0106	Forensic Science and Technology.
43	43.0116	Cyber/Computer Forensics and Counterterrorism.
45	45.0301	Archeology.
45	45.0603	Econometrics and Quantitative Economics.
45	45.0702	Geographic Information Science and Cartography.
49	49.0101	Aeronautics/Aviation/Aerospace Science and Technology, General.
51	51.1002	Cytotechnology/Cytotechnologist.

Table A.25: (continued)

51	51.1005	Clinical Laboratory Science/Medical Technology/Technologist.
51	51.1401	Medical Scientist.
51	51.2003	Pharmaceutics and Drug Design.
51	51.2004	Medicinal and Pharmaceutical Chemistry.
51	51.2005	Natural Products Chemistry and Pharmacognosy.
51	51.2006	Clinical and Industrial Drug Development.
51	51.2007	Pharmacoeconomics/Pharmaceutical Economics.
51	51.2009	Industrial and Physical Pharmacy and Cosmetic Sciences.
51	51.2010	Pharmaceutical Sciences.
51	51.2202	Environmental Health.
51	51.2205	Health/Medical Physics.
51	51.2502	Veterinary Anatomy.
51	51.2503	Veterinary Physiology.
51	51.2504	Veterinary Microbiology and Immunobiology.
51	51.2505	Veterinary Pathology and Pathobiology.
51	51.2506	Veterinary Toxicology and Pharmacology.
51	51.2510	Veterinary Preventive Medicine, Epidemiology, and Public Health.
51	51.2511	Veterinary Infectious Diseases.
51	51.2706	Medical Informatics.
52	52.1301	Management Science.
52	52.1302	Business Statistics.
52	52.1304	Actuarial Science.
52	52.1399	Management Science and Quantitative Methods, Other.

NOTES.—The table lists all the STEM (Science, Technology, Engineering, and Mathematics) programs or majors as designated by the Department of Homeland Security (The U.S. Department of Homeland Security, 2020).

A.2 Background on The Framework

The classic production function for education relates a function of inputs to student achievements, such as degree completion. For instance, the following simple production function demonstrates that student achievement (A^S for STEM related fields and A^{NS} for non-STEM) is a function of various inputs, including expenditures on instruction, academic support, institutional grants, and other factors for both STEM fields (S) and non-STEM majors (NS).[4]

$$A^j = f(x_1^j, x_2^j, \ldots, x_{n^j}^j) \quad j \in \{S, NS\}$$

This hypothetical production function is then subject to a budget constraint, which requires costs to be less than total revenue from all sources, as illustrated in the following equation:

$$\sum_{k=1}^{m} R_k \geq \sum_{i=1}^{n^S} p_i^S x_i^S + \sum_{i=1}^{n^{NS}} p_i^{NS} x_i^{NS}$$

Where x_i and p_i refer to input and input price, respectively. For example, an input x_1 could be the number of faculty hours, with the corresponding price p_1 being the faculty hourly wage. Similarly, an input x_2 could be the number of academic support staff hours, with the corresponding price p_2 being the academic support hourly wage. Public institutions generate revenue (R_k) from various sources, including local, state, and federal funds, as well as tuition, fees, and individual donations. In fiscal year 2019-20, the top five sources of revenue for 4-year public institutions were tuition and fees (20.29%), state appropriations (16.69%), hospital revenue (15.98%), federal grants and contracts (8.48%), and federal nonoperating grants (4.75%) (National Center for Education Statistics, 2024).[5]

In this hypothetical scenario, the objective of public institutions is to minimize their costs. The solution involves equalizing the marginal product per dollar of input. The proposed objective

[4]Handel and Hanushek (2022) discusses a similar objective function relevant to K-12 education. There are many other possible outputs besides A^j, such as promoting diversity and equity, advancing sustainability and environmental stewardship, and driving innovation through research (Chan, 2016). Additionally, public universities employ a distinctive production technology wherein students act both as customers and as inputs due to the influence of peer effects (Winston, 1999).

[5]In fiscal year 2007-08, the top five sources of revenue for 4-year public institutions were state appropriations (23.83%), tuition and fees (17.93%), hospital revenue (11.27%), federal grants and contracts (10.51%), and sales and services of auxiliary enterprises (8.28%) (National Center for Education Statistics, 2024).

function is useful in illustrating that STEM achievement (measured as the number of degrees conferred) may require a different set or a greater quantity of resources than non-STEM majors. For instance, instructional expenditures per student credit hour are 92% higher for electrical engineering than for English-related courses (Hemelt et al., 2021). This implies that different majors (e.g., STEM vs. non-STEM) necessitate different levels of inputs. The engineering department, for example, may require more resources per student than the business or english departments, including additional support for specialized software and laboratory supervision.

The resource-intensive nature of STEM majors motivates the hypothesis that the marginal product of STEM majors exceeds that of non-STEM majors. Consequently, state appropriations may have a more pronounced effect on STEM major completion rates.

A.3 Optimization Problem of a Public 4-Year University

In this section, I develop a model of a college's optimization problem for producing degrees (STEM or Non-STEM). The college aims to maximize the number of degrees produced (Y) through a production function that depends on instruction spending (I), aid spending (D), and academic support spending (S). The production function takes the form of a Cobb-Douglas function with total factor productivity A and output elasticities α, β, and γ for I, D, and S, respectively. The college faces a budget constraint where the sum of these expenditures must not exceed the total available revenue, which consists of state appropriations (R_s) and revenue from non-state sources (R_n).

$$\begin{aligned} \max_{I,D,S} \quad & Y = AI^{\alpha}D^{\beta}S^{\gamma} \\ \text{s.t.} \quad & I + D + S = R_s + R_n \end{aligned}$$

I focus on three spending production inputs (I, S, D) because they are the core expenditures most likely to affect the production of degrees, as highlighted in the literature (Deming & Walters, 2018; Dynarski, 2003; Hinrichs, 2022). I assume that other expenditures such as public service (i.e., non-instructional services to the community), student services (e.g., registrar activities), and auxiliary enterprise expenses (e.g., dining services and bookstores) have minimal to no effect on colleges' ability to convert enrollments into degree completions.[6]

This formulation allows me to analyze how changes in state funding affect the optimal allocation of resources and, consequently, the production of degrees. By solving this constrained optimization problem using the Lagrangian method as shown in subsection C.1, I derive the following closed-form solutions for the optimal levels of each type of spending and the resulting degree production.

$$Y^* = A(R_s + R_n)^{\alpha+\beta+\gamma} \left(\frac{\alpha^\alpha \beta^\beta \gamma^\gamma}{(\alpha+\beta+\gamma)^{\alpha+\beta+\gamma}} \right) \tag{6}$$

The marginal effect of state appropriations on degree production is given by the partial derivative $\frac{\partial Y}{\partial R_s} = (\alpha+\beta+\gamma)\frac{Y}{R_s+R_n}$. This expression decomposes into two components: the sum of output elasticities $(\alpha+\beta+\gamma)$, representing the overall responsiveness of degree production to changes in total resources, and the ratio of current degree production to total revenue $\frac{Y}{R_s+R_n}$. An increase in R_s, ceteris paribus, directly augments total revenue $(R_s + R_n)$, leading to a proportional increase in Y that is $(\alpha+\beta+\gamma)$ times the percentage increase in $(R_s + R_n)$. However, this effect exhibits diminishing marginal returns as the denominator $(R_s + R_n)$ grows with increasing R_s. Despite this, the absolute number of degrees produced (Y) continues to increase with R_s, albeit at a decreasing rate.

The elasticity of Y with respect to R_s is $(\alpha+\beta+\gamma)\frac{R_s}{R_s+R_n}$, implying that when $(\alpha+\beta+\gamma) = 1$, a 1% increase in R_s leads to a $\frac{R_s}{R_s+R_n}\%$ increase in Y. The model allows for various scale effects: if $(\alpha+\beta+\gamma) > 1$, there are increasing returns to scale, amplifying the impact of additional state funding, while $(\alpha+\beta+\gamma) < 1$ implies decreasing returns to scale.

University spending can indirectly increase the number of degrees awarded, particularly in STEM fields. Conversely, reduced spending may limit new program offerings that align with market needs, thereby affecting student persistence and degree completion.

To model the indirect effects of program offerings on degree completion through spending, let $P = \phi I^\theta$. The production function is adjusted to $Y = AI^\alpha D^\beta S^\gamma (\phi I^\theta)^\eta$, where η is the elasticity of output with respect to programs offered, θ is the elasticity of programs with respect

[6]Appendix A.2 discusses an alternative related objective function that minimizes the cost of production, which is more relevant to K-12 education. It is worth noting that the objective function of public universities is more nuanced, with no consensus in the literature on its precise formulation.

to instructional spending, and ϕ is a scaling parameter for the number of programs. Solution steps are detailed in subsection C.2.[7]

The resultant closed-form is:

$$Y^* = A\phi^\eta (R_s + R_n)^{\alpha+\beta+\gamma+\eta\theta} \left(\frac{(\alpha+\eta\theta)^{\alpha+\eta\theta}\beta^\beta\gamma^\gamma}{(\alpha+\beta+\gamma+\eta\theta)^{\alpha+\beta+\gamma+\eta\theta}} \right) \tag{7}$$

When $\eta = 0$, the model is silent on the number of programs and the reduced form turns back to equation 1. The introduction of program offerings into the production function ($\eta > 0$) yields several important insights. First, it leads to a larger share of resources being allocated to instruction ($\frac{\alpha+\eta\theta}{\alpha+\beta+\gamma+\eta\theta} > \frac{\alpha}{\alpha+\beta+\gamma}$), reflecting the additional benefit of instructional spending through its effect on program diversity. Second, the elasticity of output with respect to total revenue increases ($\alpha + \beta + \gamma + \eta\theta > \alpha + \beta + \gamma$), implying that changes in funding have a more pronounced impact on degree production when program offerings are considered. Finally, the model predicts an amplification effect of instructional spending on output: a 1% increase in instructional expenditure leads to an $(\alpha + \eta\theta)\%$ increase in degree production, rather than just $\alpha\%$.

This framework provides a theoretical foundation for understanding the mechanisms through which state appropriations influence degree completion in public universities, aligning with and informing the empirical analysis in my study. Specifically, I estimate α, β, γ, and η for both STEM and non-STEM production functions and then examine the counterfactual effects of state appropriations in the counterfactual analysis section.

[7] Figure A.4 illustrates the strong linear correlation between instructional expenditures and the number of programs, particularly for STEM majors. A limitation of this model is that it does not account for all factors influencing public universities' ability to expand their program offerings (e.g., accreditation requirements, state regulations, and approval processes).

C.1 Optimization solution without indirect effect of number of programs offered

$$\max_{I,D,S} \quad Y = AI^{\alpha}D^{\beta}S^{\gamma} \tag{8}$$

$$\text{s.t.} \quad I + D + S = R_s + R_n \tag{9}$$

where:

$$\begin{aligned}
Y &= \text{Number of degrees (STEM or Non-STEM) produced} \\
A &= \text{Total factor productivity} \\
I &= \text{Instruction spending} \\
D &= \text{Aid spending (institutional grants)} \\
S &= \text{Academic support spending} \\
R_s &= \text{Revenue from state appropriations} \\
R_n &= \text{Revenue from non-state sources} \\
\alpha, \beta, \gamma &= \text{Output elasticities}
\end{aligned}$$

I form the Lagrangian:

$$\mathcal{L} = AI^{\alpha}D^{\beta}S^{\gamma} + \lambda(R_s + R_n - I - D - S) \tag{10}$$

The first-order conditions are:

$$\frac{\partial \mathcal{L}}{\partial I} = \alpha AI^{\alpha-1}D^{\beta}S^{\gamma} - \lambda = 0 \tag{11}$$

$$\frac{\partial \mathcal{L}}{\partial D} = \beta AI^{\alpha}D^{\beta-1}S^{\gamma} - \lambda = 0 \tag{12}$$

$$\frac{\partial \mathcal{L}}{\partial S} = \gamma AI^{\alpha}D^{\beta}S^{\gamma-1} - \lambda = 0 \tag{13}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = R_s + R_n - I - D - S = 0 \tag{14}$$

From the first three conditions, I can derive:

$$\frac{\alpha Y}{I} = \frac{\beta Y}{D} = \frac{\gamma Y}{S} = \lambda \tag{15}$$

This implies:

$$I = \frac{\alpha Y}{\lambda} \tag{16}$$
$$D = \frac{\beta Y}{\lambda} \tag{17}$$
$$S = \frac{\gamma Y}{\lambda} \tag{18}$$

Substituting these into the budget constraint:

$$\frac{\alpha Y}{\lambda} + \frac{\beta Y}{\lambda} + \frac{\gamma Y}{\lambda} = R_s + R_n \tag{19}$$

Solving for λ:

$$\lambda = \frac{(\alpha + \beta + \gamma)Y}{R_s + R_n} \tag{20}$$

Substituting back, I get the optimal allocation:

$$I^* = \frac{\alpha(R_s + R_n)}{\alpha + \beta + \gamma} \tag{21}$$
$$D^* = \frac{\beta(R_s + R_n)}{\alpha + \beta + \gamma} \tag{22}$$
$$S^* = \frac{\gamma(R_s + R_n)}{\alpha + \beta + \gamma} \tag{23}$$

Substituting these into the production function yields:

$$Y^* = A\left(\frac{\alpha(R_s + R_n)}{\alpha + \beta + \gamma}\right)^{\alpha}\left(\frac{\beta(R_s + R_n)}{\alpha + \beta + \gamma}\right)^{\beta}\left(\frac{\gamma(R_s + R_n)}{\alpha + \beta + \gamma}\right)^{\gamma} \tag{24}$$

This can be simplified to:

$$Y^* = A(R_s + R_n)^{\alpha+\beta+\gamma}\left(\frac{\alpha^{\alpha}\beta^{\beta}\gamma^{\gamma}}{(\alpha + \beta + \gamma)^{\alpha+\beta+\gamma}}\right) \tag{25}$$

C.2 Optimization solution with indirect effect of number of programs offered

Multiple factors influence public universities' ability to expand their program offerings, including funding and budget constraints, accreditation requirements, and state regulations and approval processes. For simplicity, and based on the observed high correlation between instructional spending and the number of programs offered (see Figure A.4), I assume that the number of programs is a function of instructional expenditures and update the college's production function for degrees as follow:

$$Y = AI^{\alpha}D^{\beta}S^{\gamma}P^{\eta}$$

Substitute $P = \phi I^{\theta}$ into the production function:

$$Y = AI^{\alpha}D^{\beta}S^{\gamma}(\phi I^{\theta})^{\eta}$$

Simplified updated production function:

$$Y = A\phi^{\eta}I^{\alpha+\eta\theta}D^{\beta}S^{\gamma}$$

Hence the optimization problem becomes:

$$\max_{I,D,S} \quad Y = A\phi^{\eta}I^{\alpha+\eta\theta}D^{\beta}S^{\gamma} \tag{26}$$

$$\text{s.t.} \quad I + D + S = R_s + R_n \tag{27}$$

where:

$P = \phi I^{\theta}$ (Number of programs offered)

$\eta =$ Elasticity of output with respect to programs offered

$\theta =$ Elasticity of programs with respect to instruction spending

$\phi =$ Scaling parameter for the number of programs

All other variables and parameters are as described in the previous optimization.

1) The Lagrangian function is now given by:

$$\mathcal{L} = A\phi^{\eta} I^{\alpha+\eta\theta} D^{\beta} S^{\gamma} + \lambda(R_s + R_n - I - D - S)$$

2) First-order conditions: differentiating the Lagrangian with respect to I, D, S, and λ:

$$\frac{\partial \mathcal{L}}{\partial I} = A\phi^{\eta}(\alpha+\eta\theta) I^{\alpha+\eta\theta-1} D^{\beta} S^{\gamma} - \lambda = 0 \tag{28}$$

$$\frac{\partial \mathcal{L}}{\partial D} = A\phi^{\eta}\beta I^{\alpha+\eta\theta} D^{\beta-1} S^{\gamma} - \lambda = 0 \tag{29}$$

$$\frac{\partial \mathcal{L}}{\partial S} = A\phi^{\eta}\gamma I^{\alpha+\eta\theta} D^{\beta} S^{\gamma-1} - \lambda = 0 \tag{30}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = R_s + R_n - I - D - S = 0 \tag{31}$$

3) Solving for optimal allocations: from the first-order conditions, we can derive:

$$\frac{(\alpha+\eta\theta)Y}{I} = \frac{\beta Y}{D} = \frac{\gamma Y}{S} = \lambda \tag{32}$$

This implies:

$$D = \frac{\beta I}{(\alpha+\eta\theta)} \tag{33}$$

$$S = \frac{\gamma I}{(\alpha+\eta\theta)} \tag{34}$$

Substituting these into the budget constraint:

$$I + \frac{\beta I}{(\alpha+\eta\theta)} + \frac{\gamma I}{(\alpha+\eta\theta)} = R_s + R_n \tag{35}$$

Solving for I:

$$I = \frac{(\alpha+\eta\theta)(R_s + R_n)}{(\alpha+\beta+\gamma+\eta\theta)} \tag{36}$$

And consequently:

$$D = \frac{\beta(R_s + R_n)}{(\alpha + \beta + \gamma + \eta\theta)} \tag{37}$$

$$S = \frac{\gamma(R_s + R_n)}{(\alpha + \beta + \gamma + \eta\theta)} \tag{38}$$

4) Substituting these optimal allocations back into the production function:

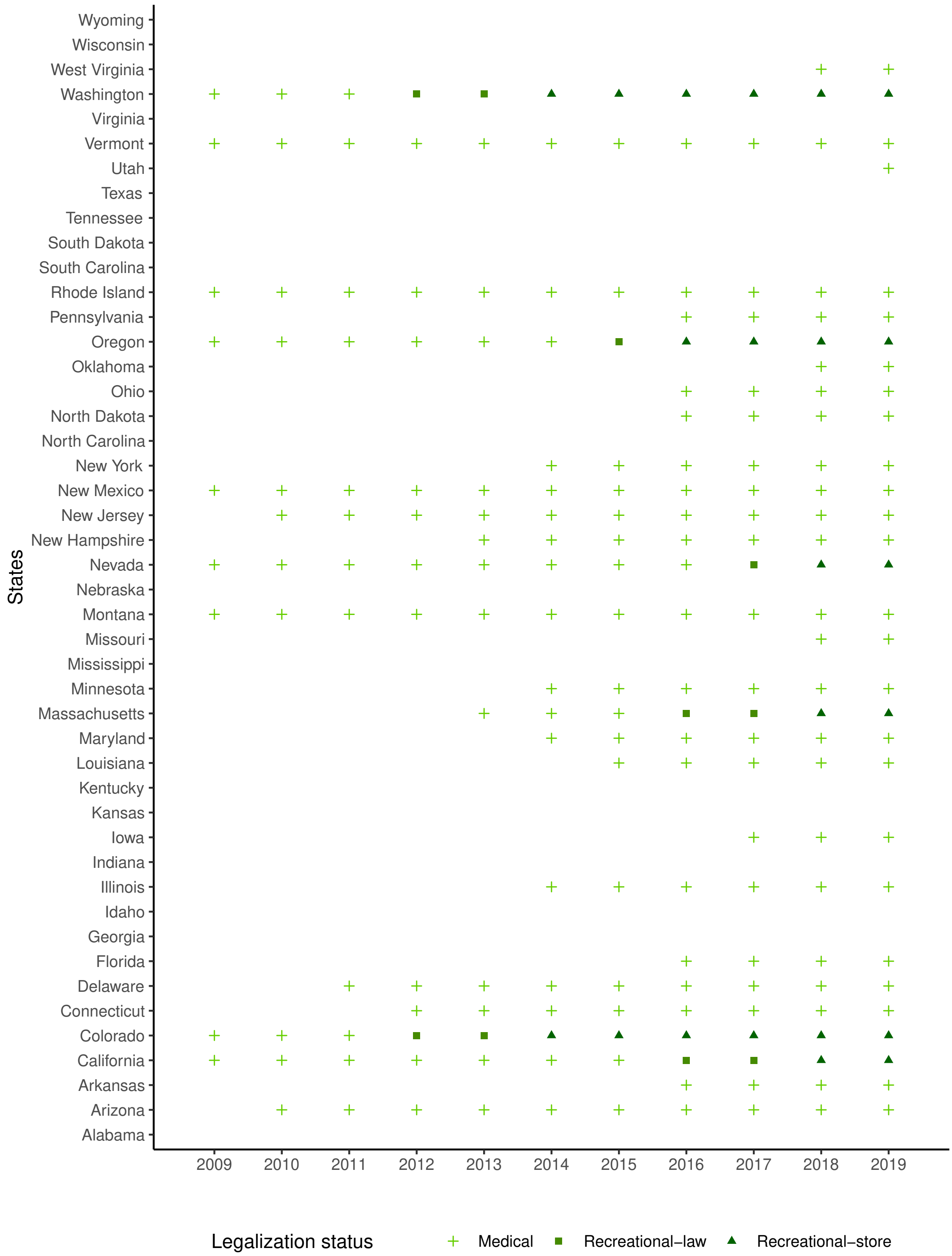
$$Y^* = A\phi^\eta \left(\frac{(\alpha + \eta\theta)(R_s + R_n)}{(\alpha + \beta + \gamma + \eta\theta)}\right)^{\alpha+\eta\theta} \left(\frac{\beta(R_s + R_n)}{(\alpha + \beta + \gamma + \eta\theta)}\right)^{\beta} \left(\frac{\gamma(R_s + R_n)}{(\alpha + \beta + \gamma + \eta\theta)}\right)^{\gamma} \tag{39}$$

This can be simplified to:

$$Y^* = A\phi^\eta (R_s + R_n)^{\alpha+\beta+\gamma+\eta\theta} \left(\frac{(\alpha + \eta\theta)^{\alpha+\eta\theta}\beta^\beta\gamma^\gamma}{(\alpha + \beta + \gamma + \eta\theta)^{\alpha+\beta+\gamma+\eta\theta}}\right) \tag{40}$$

B.1 Chapter 2

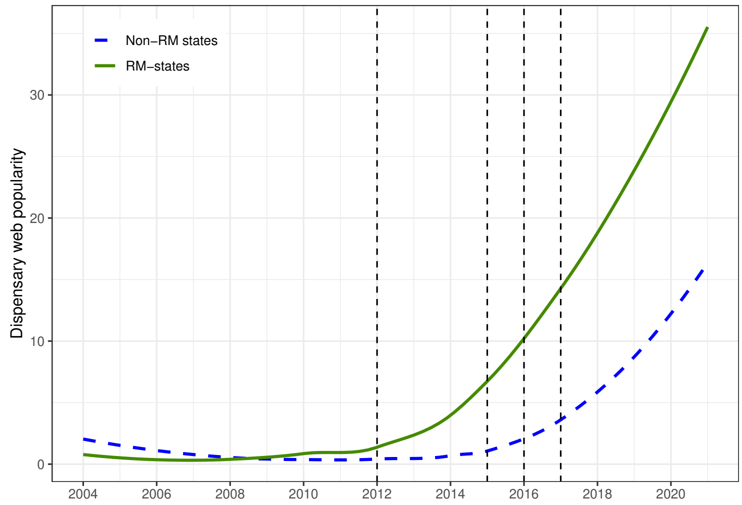
Figure B.1: Marijuana Legalization Timeline



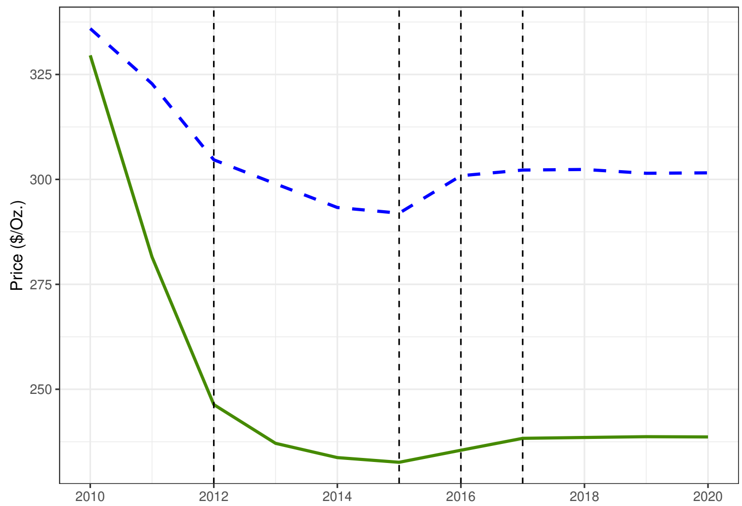
Note: the figure shows the marijuana legalization timeline by state and the type of legalization. Recreational law means that recreational marijuana is legal by means of law passage or approval (i.e., dispensaries not yet open). Recreational store means that recreational marijuana is available for sale by authorized retailers (i.e., dispensaries open). DC, Alaska, Maine, Michigan, and all US territories are excluded due to either being non-contiguous to other states or the unavailability of at least two years of post-policy observations within sample periods.

Figure B.2: Marijuana Consumption and Price by Legalization Status

Panel (a): Google Trends for "dispensary"

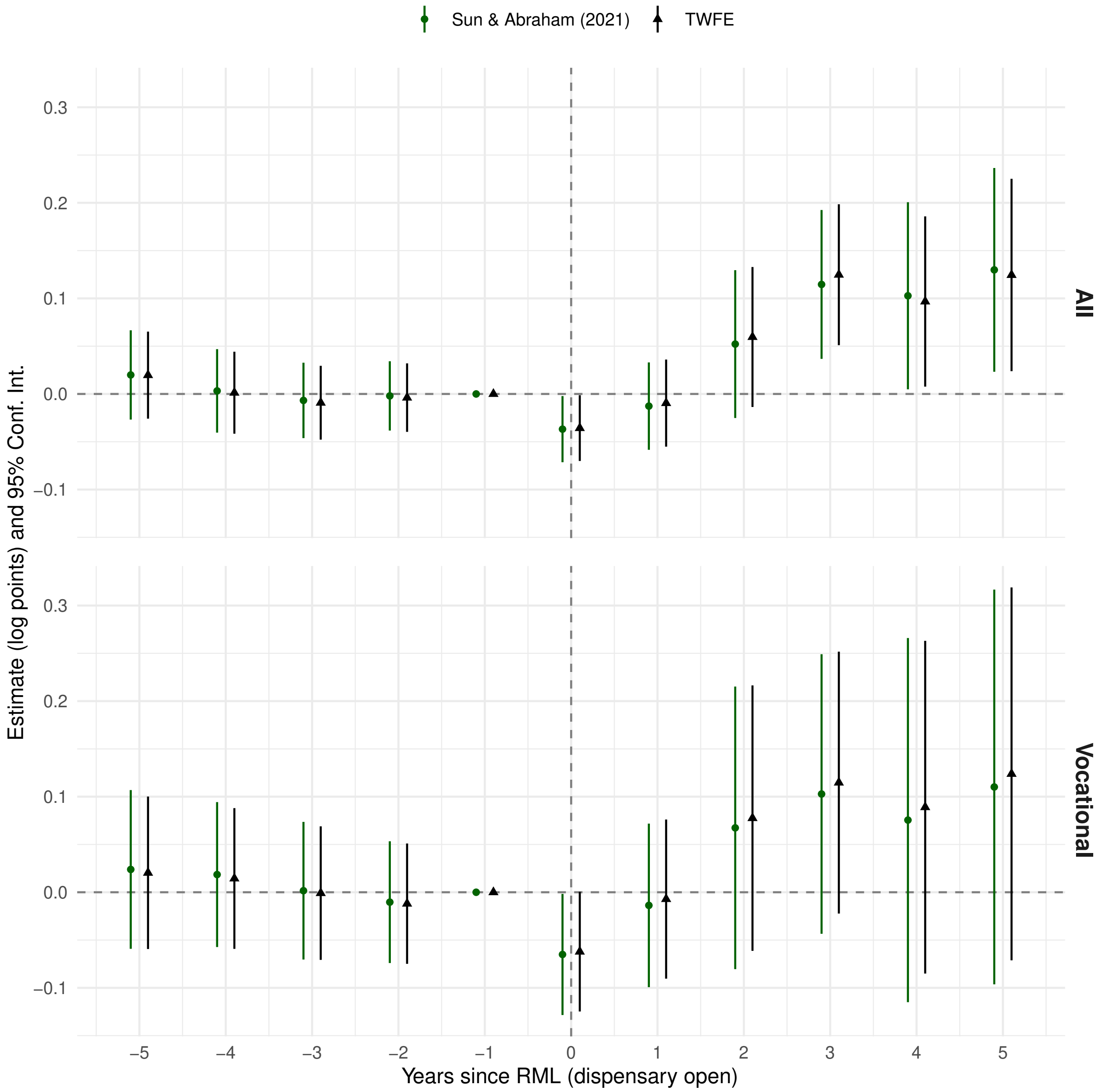


Panel (b): Marijuana Price



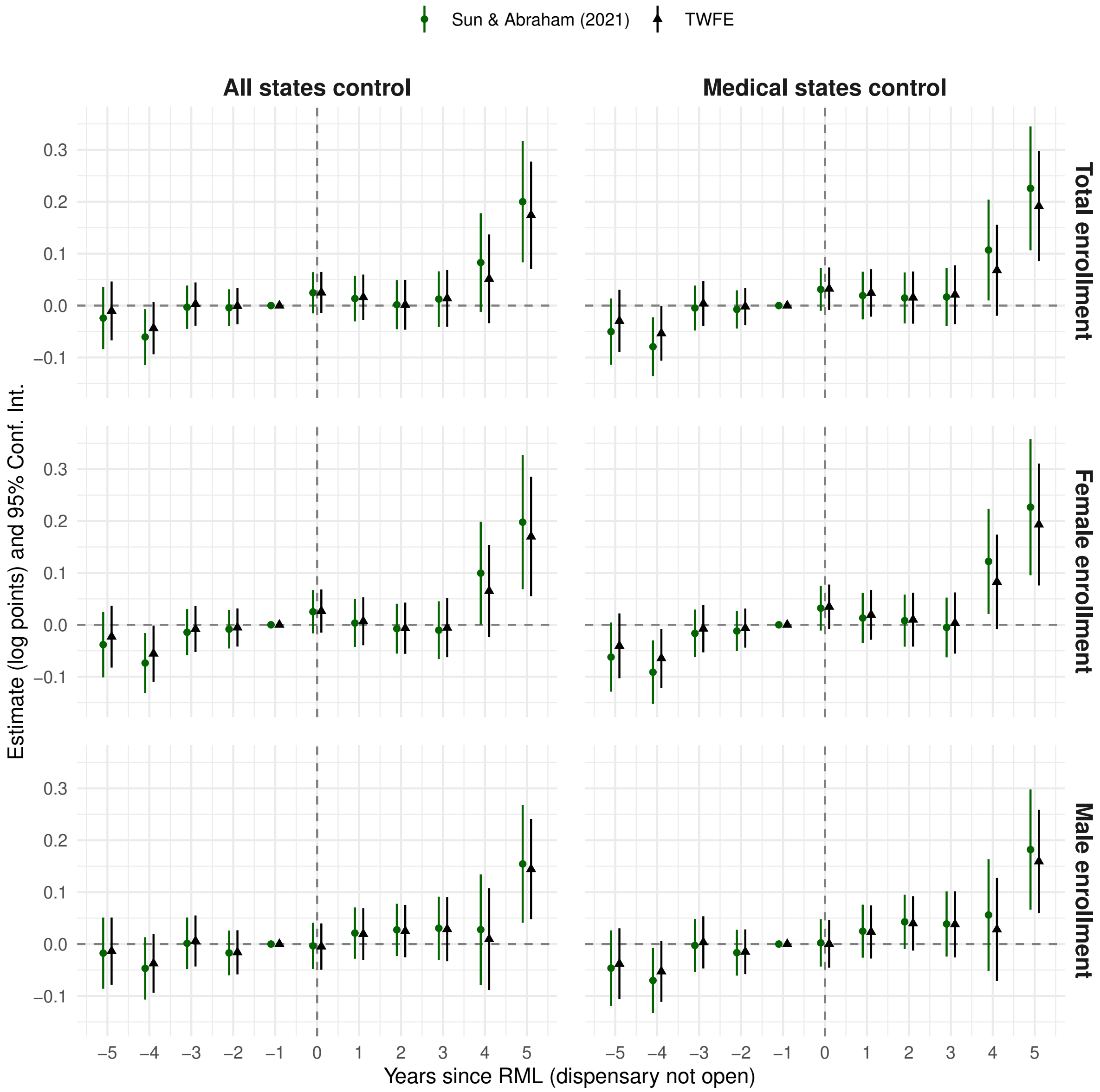
Note: Google web popularity (refer to as hits in Google Trends) is a number between 0 and 100 that measures the popularity of Google search terms or words, showing how popular a word is among all other searches in a particular region and time. Panel (a) compares the average popularity of the word "dispensary" between states that legalized recreational marijuana (RM-states) and those that did not legalize it (Non-RM states). Marijuana prices are extracted from PriceOfWeed.com via Wayback Machine. Panel (b) depicts the average of the medium and high-quality marijuana prices (low-quality prices are not available for most years) for each state and year. The figure compares the average yearly cannabis prices between states that legalized marijuana practically for recreational use as of 2019 and those that did not. The vertical lines refer to the year in which different states legalized marijuana–Colorado and Washington in 2012, Oregon in 2015, Massachusetts and California in 2016, and Nevada in 2017.

Figure B.3: Event Study Graph of RML Treatment by Institution Type



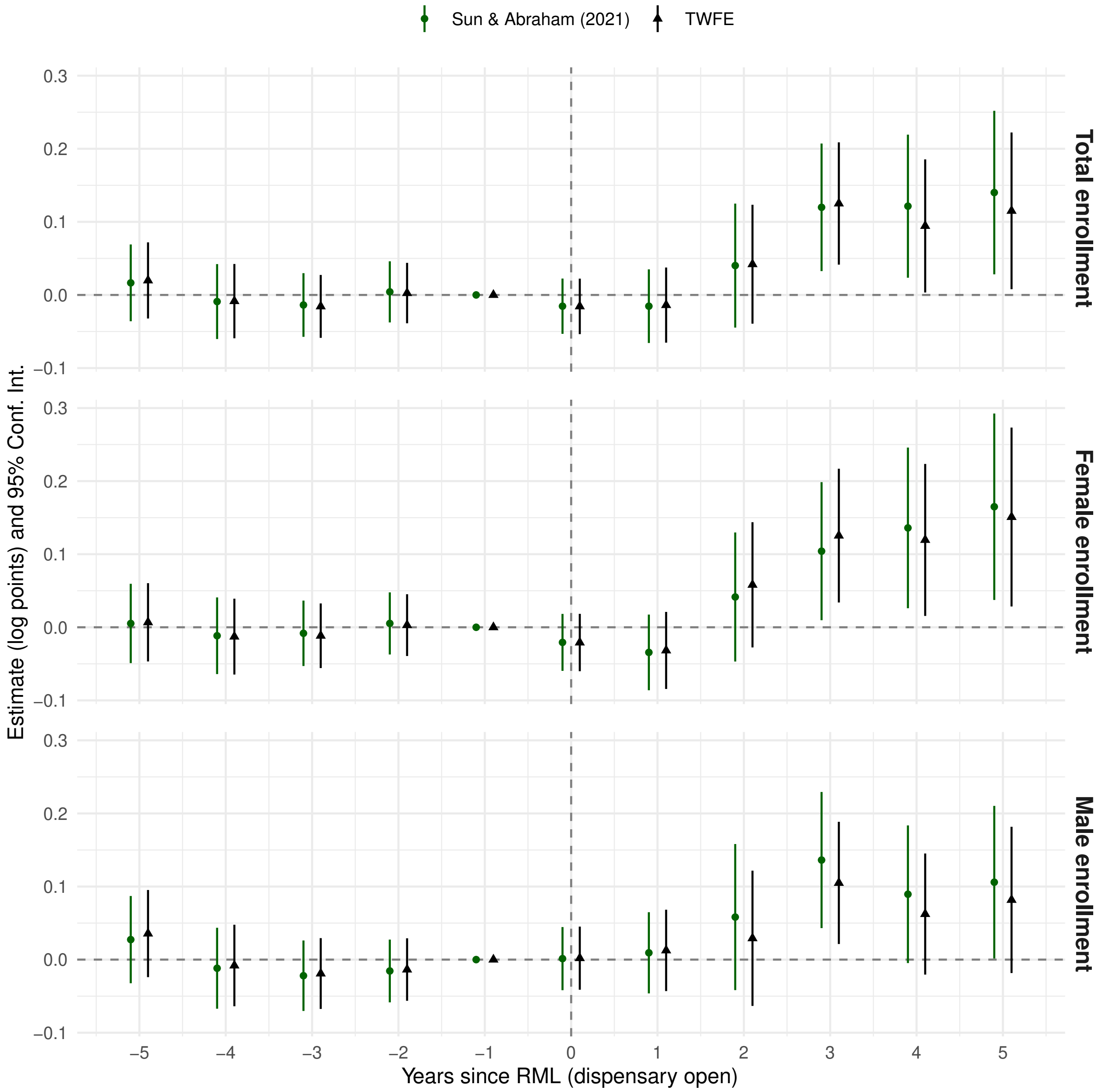
Note: The plots depict the RML effect estimates on first-time enrollments over time based on both the TWFE difference-in-differences and Sun and Abraham (2021) models. All outcome variables are logged to simplify the interpretation of the coefficients. The zero value on the x-axis refers to the first year of legalization. The control group consists of all the states that legalized medical marijauna (see Figure B.7). RML is defined by the dispensary opening time. All Control variables and fixed effects are included as described in Table B.2. Standard errors are clustered at the state level (see Figure 1).

Figure B.4: Event Study Graph of RML (Dispensary Not Open) Treatment



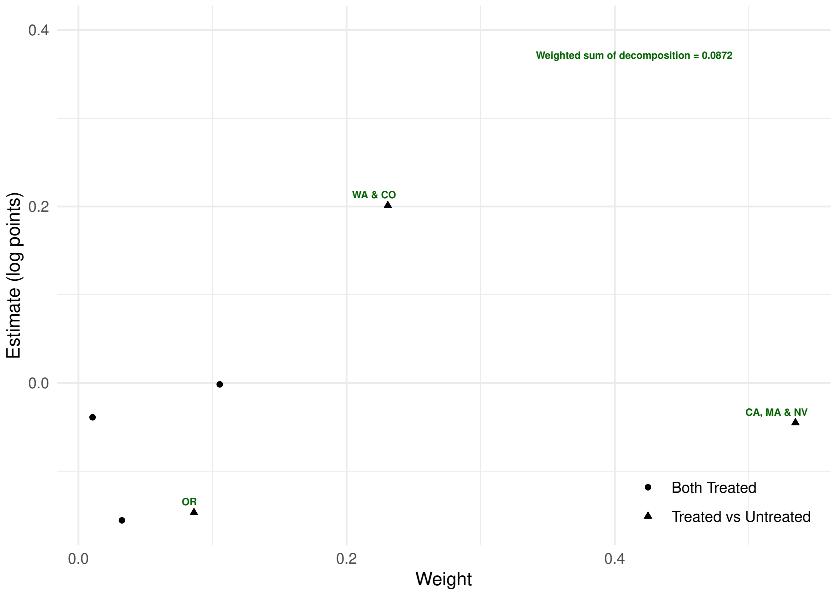
Note: The plots depict the RML effect estimates on first-time enrollments over time using both TWFE difference-in-differences and Sun and Abraham (2021) models. All outcome variables are logged to simplify the interpretation of the coefficients. Medical states control refers to the use of states that legalized marijuana for medical use as a control group (see Figure B.7). The zero value on the x-axis refers to the first year of legalization. RML is defined utilizing the law passage timing as shown in Figure 1. All Control variables and fixed effects are included as described in Figure B.3. Standard errors are clustered at the state level.

Figure B.5: Event Study Graph of RML (Dispensary Open) Treatment Using All States Control Group



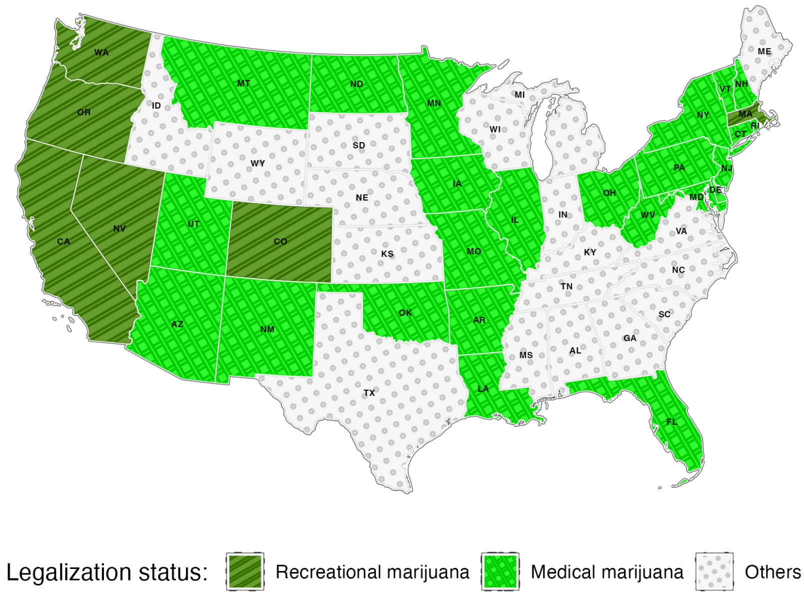
Note: The plots depict the RML effect estimates on first-time enrollments in academic institutions over time based on both the TWFE and Sun and Abraham (2021) models. The zero value on the x-axis refers to the first year of legalization. RML policy is defined by the first dispensary opening time as shown in figure 1. All Control variables and fixed effects are included as described in Figure B.3. Standard errors are clustered at the state level.

Figure B.6: Goodman-Bacon Decomposition



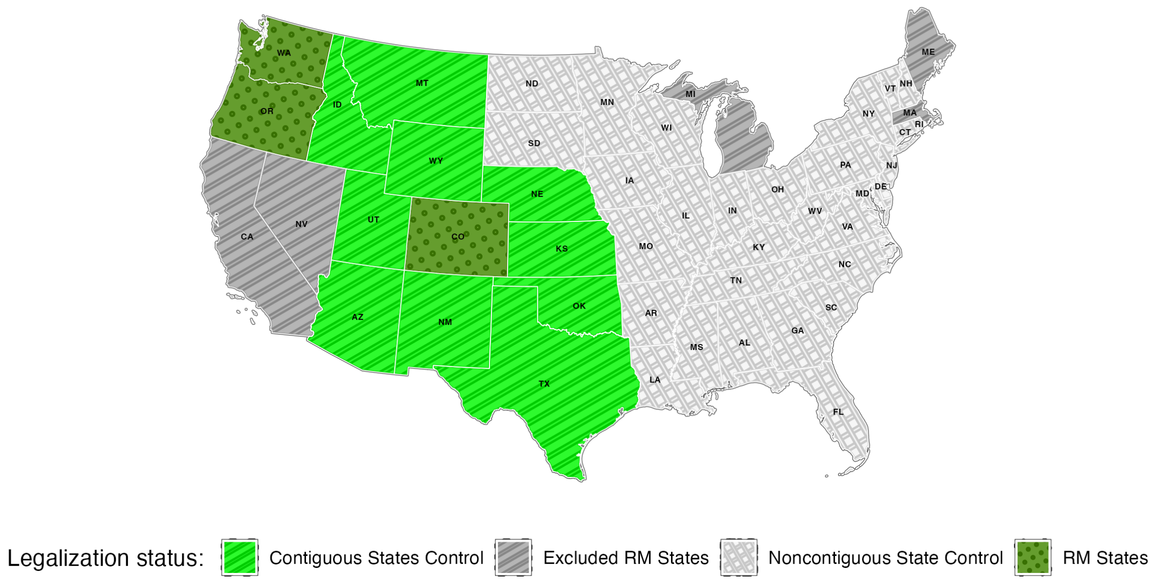
Note: The figures illustrate the point estimate subparts according to Goodman-bacon decomposition. The outcome variable is the logged total first-time enrollments in academic institutions. The decomposition shows that the point estimates are mainly derived from RM and non-RM state pairs–"both treated" pairs have minimal weights. Control variables and fixed effects are as described in Figure B.3.

Figure B.7: Medical Marijuana Control Group



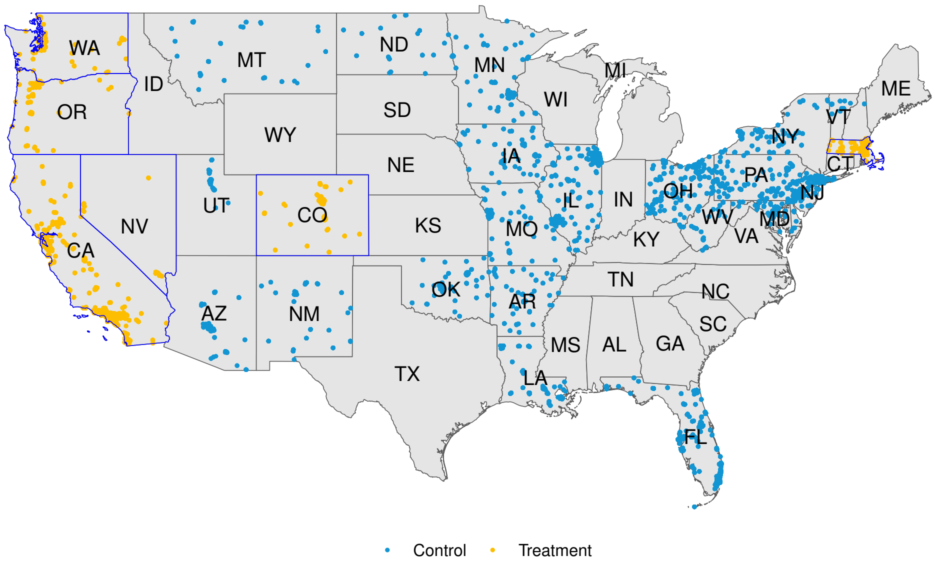
Note: the map shows the states which are included in the medical control group–the states that legalized medical marijuana as of 2019. Others refer to the states which are excluded (Maine, Michigan) because of a lack of post-policy observations or states that have not yet legalized marijuana for medical use.

Figure B.8: Contiguous and Noncontiguous Control States



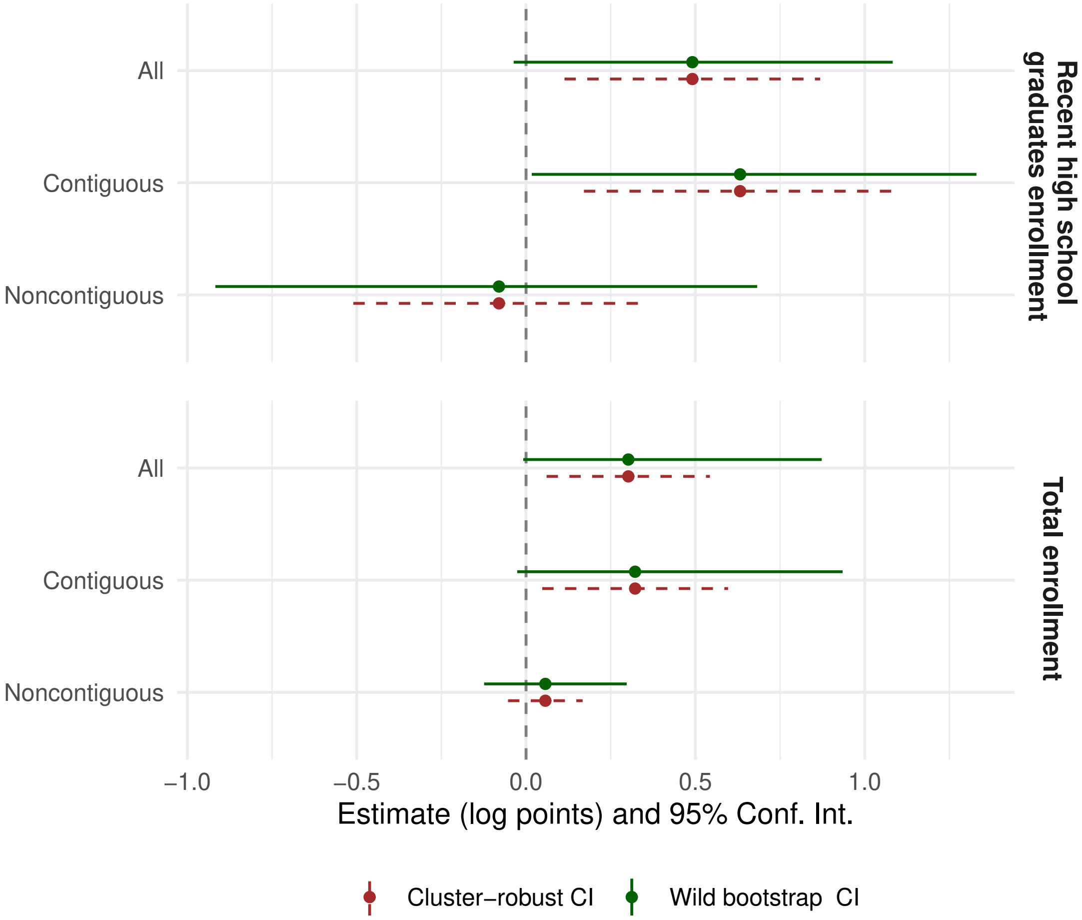
Note: the map shows the states which are included in the contiguous and noncontiguous control groups.

Figure B.9: The Exclusion of Control Colleges Within 80 Miles of Treated Borders



Note: The figure depicts colleges in the states that are included in the control and treatment groups for spillover analysis. Colleges within 80 miles are excluded from the control group, which includes all the states that legalized marijuana for medical use.

Figure B.10: RML (Dispensary Not Open) Effect on Out-of-State Enrollment, by Contiguity Status



Note: The plots depict the RML effect estimates on total and recent high school graduates' out-of-state first-time enrollments based on TWFE difference-in-differences. Contiguous refers to states that are contiguous or within close proximity to RM-states as shown in figure B.8. The treatment group includes early movers (Colorado, Washington, and Oregon) because out-of-state data is available only biennially from 2008 to 2018. RML is defined utilizing the law passage timing as shown in Figure 1. All Control variables and fixed effects are included as described in Figure B.3. Standard errors are clustered at the state level.

Table B.1: Summary Statistics–All States Control Group

	Non-RM States N Observations=28710 N institutions=3114		RM States N Observations=6534 N institutions=733			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	p
Log total first-time enrollment	5.672	1.655	5.765	1.844	0.092	<0.001
Log male first-time enrollment	4.693	1.907	4.775	2.110	0.082	0.004
Log female first-time enrollment	5.065	1.684	5.179	1.814	0.114	<0.001
Log number of degrees	6.361	1.632	6.518	1.733	0.157	<0.001
Aggregate enrollment over 20k dummy	0.059	0.236	0.099	0.299	0.040	<0.001
Offering medical degree dummy	0.038	0.192	0.029	0.168	-0.009	<0.001
Offering ROTC program dummy	0.320	0.466	0.226	0.418	-0.094	<0.001
Offering distance programs dummy	0.794	0.405	0.762	0.426	-0.032	<0.001
Log age 18 to 24 population	10.091	1.498	11.495	1.464	1.404	<0.001
Age 18 to 24 population female share	0.492	0.027	0.487	0.022	-0.005	<0.001
Unemployment rate	6.328	2.521	7.267	3.397	0.939	<0.001
Log per capita income	10.657	0.259	10.793	0.258	0.136	<0.001
Log Net migration	11.296	0.119	11.099	1.284	-0.197	<0.001
Log total out-of-state enrollment	6.444	1.677	6.507	1.852	0.063	0.088
Log RHG out-of-state enrollment	5.743	2.305	5.652	2.465	-0.091	0.064
Tuition and fees revenue	16.504	1.741	16.586	1.611	0.082	0.069
Tuition and fees revenue per student	8.842	1.174	8.497	1.683	-0.345	<0.001
Log number of admissionss	7.130	1.597	7.131	1.960	0.001	0.973
Log number of ACT and SAT submissions	6.307	1.249	6.681	1.345	0.374	<0.001
Bachelor's Graduation Rate (100% normal time)	35.009	22.241	41.507	25.451	6.498	<0.001
Bachelor's Graduation Rate (150% normal time)	49.919	20.732	57.255	22.455	7.336	<0.001
Bachelor's Graduation Rate (200% normal time)	51.969	20.282	59.888	21.538	7.919	<0.001
Associate's Graduation Rate (100% normal time)	20.153	20.007	16.618	18.001	-3.535	<0.001
Associate's Graduation Rate (150% normal time)	30.597	20.456	32.238	19.697	1.641	<0.001
Associate's Graduation Rate (200% normal time)	34.939	19.756	38.473	18.752	3.535	<0.001
Retention rate	66.252	17.022	70.376	16.816	4.124	<0.001
ACT 75th Math score	24.866	3.362	26.528	3.863	1.662	<0.001
SAT 75th Math score	585.438	70.585	611.573	77.858	26.135	<0.001
ACT 75th English score	25.504	3.913	27.042	4.437	1.538	<0.001
ACT 75th Crititcal Reading score	582.072	68.650	605.316	72.090	23.243	<0.001

Note: enrollment refers to the number of first-time degree-seeking undergraduate students. RM states refer to Colorado, Washington, Oregon, California, Nevada, and Massachusetts which respectively opened the first dispensary in 2014, 2014, 2016, 2018, 2018, and 2018. Territories, Hawaii, Alaska, Maine, and Michigan are excluded either because they are noncontiguous to other states or because they open their first dispensary in 2019 or after (see Figure 1). The unit of observation is university-year. Bachelor's and associate's degrees make up the total number of degrees conferred. RHG stands for Recent High School Graduates and ROTC for Reserve Officers' Training Corps. The control group in this table consists of all the states that did not legalize marijuana for recreational use (see Figure B.1).

Table B.2: RML Effect on First-Time Enrollment

	Total enrollment				Female enrollment				Male enrollment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel (a): RM policy by the law approval												
RM	0.073+	0.089+	0.071+	0.087*	0.068	0.081+	0.066+	0.079+	0.072	0.093+	0.070	0.091+
	(0.043)	(0.045)	(0.037)	(0.039)	(0.044)	(0.047)	(0.038)	(0.041)	(0.051)	(0.052)	(0.046)	(0.047)
Wild bootstrap p	0.098	0.061	0.070	0.039	0.147	0.102	0.101	0.073	0.162	0.094	0.138	0.072
Panel (b): RM policy by the retailer stores opening												
RM	0.066	0.079	0.061	0.073	0.054	0.066	0.049	0.060	0.088	0.103+	0.083	0.097+
	(0.055)	(0.056)	(0.053)	(0.054)	(0.060)	(0.062)	(0.057)	(0.059)	(0.058)	(0.056)	(0.055)	(0.053)
Wild bootstrap p	0.252	0.190	0.258	0.200	0.390	0.323	0.415	0.349	0.139	0.084	0.146	0.092
N Obs.	23 527	23 527	23 526	23 526	23 527	23 527	23 526	23 526	23 527	23 527	23 526	23 526
N college	2549	2549	2549	2549	2549	2549	2549	2549	2549	2549	2549	2549
Pre-RML mean enrollment	1002	1002	1002	1002	531	531	531	531	471	471	471	471
County controls	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
College controls	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
College FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Note: The table displays the results of the TWFE difference-in-differences. The outcome variables refer to the logged first-time enrollments. The sample control units include the states which legalized marijuana for medical use. The included states are depicted in figure B.7. Pre-RML mean enrollment refers to the treated states (CO, WA, OR, CA, NV, and MA) average first-time enrollment before legalizing marijuana for recreational use. The models include year and institution fixed effects as well as baseline covariates. The institution-level controls are as follows: medical degree dummy, college size dummy (having over 20k total enrollments), student-to-faculty ratio, distance program dummy, and Reserve Officers' Training Corps (ROTC) program dummy. The county-level controls include the age 18 to 24 population, age 18 to 24 female population share, per-capita income, unemployment rate, and net migration. Using state-year subclusters and 10,000 wild-cluster bootstrap iterations, I follow MacKinnon and Webb (2018) to show the wild bootstrap p values. Standard errors are clustered at the state level. + $p < 0.1$,* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.3: RML Effect on First-Time Enrollment Using All States Control Group

	Total enrollment				Female enrollment				Male enrollment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel (a): RM policy by the law approval												
RM	0.023	0.036	0.039	0.049	0.022	0.029	0.038	0.042	0.021	0.039	0.034	0.050
	(0.035)	(0.036)	(0.031)	(0.032)	(0.038)	(0.040)	(0.035)	(0.036)	(0.040)	(0.040)	(0.037)	(0.037)
Wild bootstrap p	0.526	0.334	0.231	0.141	0.580	0.486	0.299	0.266	0.620	0.351	0.359	0.196
Panel (b): RM policy by the retailer stores opening												
RM	0.022	0.035	0.036	0.045	0.015	0.024	0.030	0.034	0.034	0.052	0.047	0.060
	(0.048)	(0.049)	(0.046)	(0.047)	(0.053)	(0.055)	(0.052)	(0.053)	(0.051)	(0.049)	(0.048)	(0.047)
Wild bootstrap p	0.663	0.505	0.464	0.375	0.790	0.680	0.589	0.549	0.509	0.320	0.346	0.231
N Obs.	35 244	34 752	35 241	34 749	35 244	34 752	35 241	34 749	35 244	34 752	35 241	34 749
N college	3851	3851	3851	3851	3851	3851	3851	3851	3851	3851	3851	3851
Pre-RML mean enrollment	1002	1002	1002	1002	531	531	531	531	471	471	471	471
County controls	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
College controls	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
College FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Note: the table displays the results of the TWFE difference-in-differences. The outcome variables refer to the logged first-time enrollments. All Control variables and fixed effects are included as described in Figure B.3. Pre-RML mean enrollment refers to the treated states (CO, WA, OR, CA, NV, and MA) average first-time enrollment before legalizing marijuana for recreational use. Standard errors are clustered at the state level. Using state-year subclusters and 10,000 wild-cluster bootstrap iterations, I follow MacKinnon and Webb (2018) to show the wild bootstrap p values. + $p < 0.1$,* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.4: RML Effect on Completion Using All States Control Group

	Lead 1	Lead 2	Lead 3	Lead 4	Lead 5	Lead 6
Panel (a): Log Number of Undergraduate Degrees						
	Bachelor Degrees					
RM	−0.002	0.043	0.062+	0.128***	0.118***	0.093**
	(0.065)	(0.049)	(0.036)	(0.024)	(0.023)	(0.027)
Wild bootstrap p	0.976	0.519	0.239	0.015	0.011	0.022
N Obs.	18 951	16 603	14 326	12 127	10 058	8102
	Associate Degrees					
RM	−0.027	−0.007	0.037	0.069	0.058	0.031
	(0.091)	(0.077)	(0.070)	(0.060)	(0.060)	(0.055)
Wild bootstrap p	0.786	0.932	0.643	0.346	0.419	0.653
N Obs.	22 345	19 409	16 595	13 937	11 454	9118
Panel (b): Bachelor Degrees Graduation Rate						
	100% of Normal Time (4-years)					
RM	−0.517	−0.063	1.277	−0.148	−0.635	0.115
	(0.737)	(0.905)	(0.976)	(0.597)	(1.219)	(1.479)
Wild bootstrap p	0.616	0.961	0.356	0.859	0.702	0.948
	150% of Normal Time (6-years)					
RM	0.419	1.413+	2.037+	0.632	0.035	−0.659
	(0.949)	(0.704)	(1.101)	(0.591)	(1.747)	(1.218)
Wild bootstrap p	0.742	0.216	0.221	0.467	0.988	0.693
N Obs.	14 887	14 997	13 660	12 317	10 963	9597
Panel (c): Associate Degrees Graduation Rate						
	100% of Normal Time (2-years)					
RM	2.005	0.654	1.496	2.262	3.615+	3.283*
	(1.878)	(2.580)	(1.937)	(1.748)	(2.033)	(1.576)
Wild bootstrap p	0.417	0.861	0.632	0.389	0.286	0.252
	150% of Normal Time (3-years)					
RM	1.839	0.973	−0.986	−1.374	−0.866	−1.604
	(1.451)	(1.615)	(1.764)	(1.775)	(1.913)	(2.071)
Wild bootstrap p	0.289	0.617	0.675	0.530	0.722	0.577
N Obs.	8102	8009	7189	6378	5585	4810
Controls	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
College FE	Y	Y	Y	Y	Y	Y

Note: The outcome variable in panel (a) refers to the logged number of all undergraduate degrees (associates and bachelors) conferred. In panels (b) and (c), the dependent variable is the graduation rate within the standard degree completion time (i.e., 4 years for bachelor) or or within 150% of the normal time (i.e., 6 years for bachelor). The graduation rate refers the percentage of students in a cohort who completed their degrees within the specified times (100% or 150% of normal degree time). The policy is defined by law approval timing (see Figure 1) to allow the use of 6 leads. This is in contrast to prior exhibits which are based on the dispensary opening. A lead of "i" (ranging from 1 to 6) indicates that the ith lead of the outcome variable was used in the analysis. Colorado and Washington are the only treated states. Nevada, California, Oregon, and Massachusetts are dropped from this analysis due to the nonexistence of observations for leads of the outcome variable. The control group includes states that legalized marijuana for medical use (see figure B.7 for more details). Using state-year subclusters and 10,000 wild-cluster bootstrap iterations, I follow MacKinnon and Webb (2018) to show the wild bootstrap p values. Control variables and fixed effects are as described in Table B.2. Standard errors are clustered at the state level. + $p < 0.1$,* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.5: RML Effect on Tuition Revenue and Retention Rate–RML by Dispensaries Not Open

	Tuition Revenue		Tuition Revenue Per Student		Retention Rate	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel (a): All state control group						
RM	0.070	0.061	0.057	0.024	−0.006	−0.013
	(0.041)	(0.033)	(0.036)	(0.018)	(0.018)	(0.020)
Wild bootstrap p	0.089	0.077	0.128	0.274	0.742	0.519
N Obs.	33 587	33 587	33 584	33 584	32 729	32 729
N college	3851	3851	3851	3851	3851	3851
Panel (b): medical states control group						
RM	0.106*	0.087*	0.085	0.033	−0.024	−0.035
	(0.050)	(0.035)	(0.045)	(0.019)	(0.021)	(0.027)
Wild bootstrap p	0.031	0.019	0.056	0.238	0.252	0.203
N Obs.	22 558	22 558	22 556	22 556	21 998	21 998
N college	2549	2549	2549	2549	2549	2549
Controls	Y	Y	Y	Y	Y	Y
College FE	Y	Y	Y	Y	Y	Y
Year FE	N	Y	N	Y	N	Y

Note: In Panel (b), the sample control units include the states which legalized marijuana for medical use. The included states are depicted in figure B.7. The tuition and fees variables are derived by binding all the IPEDS finance surveys. Since these surveys depend on the institution types (public or private) and on the type of accounting standards used (GASP and FASB), I controlled for these types in the tuition models. Tuition refers to all the tuition and fees (logged) collected during a year period. Retention rate (logged) is defined on IPEDS's survey as "the percent of the (fall full-time cohort from the prior year minus exclusions from the fall full-time cohort) that re-enrolled at the institution as either full- or part-time in the current year". All Control variables and fixed effects are included as described in Figure B.3. Standard errors are clustered at the state level. I report wild bootstrap p values following MacKinnon and Webb (2018) using 10,000 bootstrap iterations at state-year subcluster. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.6: RML Effect on Admissions and Test Scores

	N Admission (1)	N Tests Submissions (2)	ACT Math (3)	SAT Math (4)	ACT English (5)	SAT Reading (6)
Panel (a): All state controls						
RM	-0.031	-0.036	0.039	0.067	0.127^{*}	-0.006
	(0.032)	(0.038)	(0.041)	(0.038)	(0.048)	(0.028)
Wild bootstrap p	0.377	0.461	0.384	0.146	0.036	0.841
N Obs.	16 749	13 111	11 217	12 152	11 217	12 056
N college	2007	1458	1345	1400	1345	1400
Panel (b): Medical state controls						
RM	-0.015	-0.010	0.040	0.077	0.181^{**}	0.043
	(0.032)	(0.049)	(0.038)	(0.042)	(0.055)	(0.045)
Wild bootstrap p	0.689	0.866	0.336	0.133	0.010	0.408
N Obs.	11 258	8395	6862	7938	6863	7881
N college	1329	933	841	909	841	909
Controls	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
College FE	Y	Y	Y	Y	Y	Y

Note: The dependent variables in columns (1) and (2) refer respectively to the logged number of admitted students and the logged number of SAT and ACT test scores received for admission purposes to an undergraduate program. In columns (3) to (6), the outcome variables are the standardized 75th average scores for each test type. In Panel (b), the sample control units include the states which legalized marijuana for medical use. The included states are depicted in Figure B.7. All Control variables and fixed effects are included as described in Figure B.3. Standard errors are clustered at the state level. I report wild bootstrap p values following MacKinnon and Webb (2018) using 10,000 bootstrap iterations at state-year subcluster. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.7: RML Effect on In-State Enrollment

	Total in-state				Recent in-state high school graduates			
RM	0.063	0.023	0.025	-0.003	0.099^{*}	0.077	0.029	0.020
	(0.035)	(0.038)	(0.030)	(0.034)	(0.045)	(0.052)	(0.038)	(0.048)
Wild bootstrap p	0.313	0.680	0.643	0.941	0.246	0.471	0.660	0.781
N Obs.	10 683	8898	15 765	12 436	10 683	8898	15 765	12 436
N college	2503	2088	3786	3004	2503	2088	3786	3004
Control group	MM	MM	All	All	MM	MM	All	All
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
College FE	Y	Y	Y	Y	Y	Y	Y	Y
Contiguous states excluded	N	Y	N	Y	N	Y	N	Y

Note: Total in-state refers to the logged number of local/ in-state first-time undergraduate students. Recent in-state high school graduates refer to logged local first-time undergraduate students who graduated from high school in the past 12 months. MM refers to the control group consisting of states that legalized marijuana for medical use (see Figure B.7) and All to all non-RM states. The included states are depicted in Figures B.8. All Control variables and fixed effects are included as described in Figure B.3. Standard errors are clustered at the state level. I report wild bootstrap p values following MacKinnon and Webb (2018) using 10,000 bootstrap iterations at state-year subcluster. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.8: Heterogeneity of RML Effect by the Maximum Allowed Cultivation Thresholds

	None (WA)		4 plants (OR)		6 plants (CA, MA, NV, & CO)	
RM	0.142**	0.225***	−0.028	−0.027	−0.078	−0.131
	(0.039)	(0.034)	(0.042)	(0.049)	(0.071)	(0.085)
Wild bootstrap p	0.249	0.070	0.768	0.765	0.457	0.302
N Obs.	23 526	23 526	23 526	23 526	23 526	23 526
N college	2549	2549	2549	2549	2549	2549
Year FE	N	Y	N	Y	N	Y
College FE	Y	Y	Y	Y	Y	Y

Note: The outcome variable is log first-time enrollments. The columns designated as "None," "4 plants," and "6 plants" pertain to different types of interaction terms with the primary policy dummy. For instance, in Washington (WA), where cannabis cultivation is prohibited, the term "RM" denotes the interaction between the RML policy dummy and another binary variable set to one if the state is WA. The sample control units include the states which legalized marijuana for medical use. The included states are depicted in Figure B.7. All Control variables and fixed effects are included as described in Figure B.3. Standard errors are clustered at the state level. I report wild bootstrap p values following MacKinnon and Webb (2018) using 10,000 bootstrap iterations at state-year subcluster.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.9: Effects of RML on Enrollment Sensitivity to Proximity Spillover with Institution Clustering

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel (a): All Institutions								
RM	0.078**	0.079**	0.082**	0.078**	0.076*	0.075*	0.076*	0.075*
	(0.029)	(0.029)	(0.029)	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)
N Obs.	23,325	23,112	22,919	22,698	22,526	22,409	22,290	22,082
N colleges	2,529	2,507	2,486	2,466	2,451	2,439	2,427	2,409
Panel (b): Public and Non-research Institutions								
RM	0.106***	0.104***	0.106***	0.096**	0.087**	0.086**	0.087**	0.087**
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.032)	(0.032)
N Obs.	5,635	5,559	5,528	5,418	5,354	5,319	5,257	5,230
N colleges	602	594	591	581	576	572	565	562
N Miles removed	10	20	30	40	50	60	70	80
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
College FE	Y	Y	Y	Y	Y	Y	Y	Y

Note: This table presents the results of the TWFE difference-in-differences analysis. The outcome variable is log first-time enrollments. Panel (a) is based on the main results sample, which contains all academic institutions. Panel (b) sample consists of public institutions that do not offer graduate degrees (i.e., offering only bachelor or associate degrees). To assess potential bias in the estimates caused by spillover effects from students commuting or migrating to policy-affected colleges, each column shows the results after excluding control colleges located within a certain distance of the policy-affected border. For instance, Column 8 displays the results obtained after excluding all untreated colleges within 80 miles of the border of the treated states. Figure B.9 illustrates the geographical distribution of control and treatment institutions. Standard errors are clustered at the institution level. The policy dummy variable is defined at the dispensary opening time of legalization. Control variables and fixed effects are as described in Figure B.3. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

B.2 Description of IPEDS surveys

legally required to file IPEDS surveys. All higher education institutions that participate in any federal student financial assistance program must file the surveys (National Center for Education Statistics, 2019).

To restrict the sample to only colleges, community colleges, and universities, I use string filtration based on the name of the institutions after looking up the mission of each. For instance, I filtered out beauty, barber, welding academies, and other entities whose primary mission is to provide training programs. Students that target vocational entities for their postsecondary education are less likely to relocate to other states perhaps due to their socioeconomic status or to the transiency of the vocational education and training (VET) programs (i.e, the low cost of vocational entities is most likely a major factor influencing VET program choice). Each survey is joined by year and institution identification number with a directory survey which provides important geographical information such as location coordinates, whether the institution offers medical degrees, and, importantly the latitude and longitude coordinates where each entity is located. The latter coordinates were available on IPEDS as of 2009 and are necessary for detailed geographical analysis.

For the main outcome of interest, I use enrollment by race/ethnicity, gender, attendance status, and level of student IPEDS survey,[8] but I focus only on gender and overall fall first-time enrollment.[9] One advantage of using a fall-enrollment survey is that it also includes other important measures such as retention rate.

To study possible channels driving improvements in first-time enrollments due to RML, I use other IPEDS surveys. The admissions survey is limited to colleges that have admission requirements such as submission of ACT or SAT scores. The variables of interest are the number of admissions, the 75th percentile scores of SAT and ACT, and the number of these test score submissions. It should be noted that admission may include other requirements besides test scores, such as a high school GPA or completion of a preparatory program.

[8]This survey is a subpart of multiple surveys under Fall Enrollments.

[9]Race enrollments most likely suffer from sample selection due to the existence of so-called historical colleges for particular races (National Center for Education Statistics, 2022b). Also, many colleges do not report the race enrollments in each period.

Another useful IPEDS survey is that of residence and migration of first-time undergraduate students; the latter allows me to study the effect of RML on out-of-state enrollments. The survey provides state residency data on first-time undergraduate students who graduated from high school in the past 12 months (referred to as RHS later) and the total. This survey, though, does not provide information on the gender of students and is available on a biennial basis only. Further, the finance survey includes financial information for each academic institution. Since the survey is filed separately by public and private colleges that follow different accounting systems (GASP and FASP), I use all tuition and fees observations and control for the filing type.

The completion and graduation rates survey provides information about the number of awards or degrees by program type and level of the degree. Similar to other surveys, I restrict the sample to undergraduate academic institutions and calculate the total number of degrees granted by each institution for each academic year, from July 1st to June 30th. This survey is critical in determining whether the effect on enrollment also translates to a similar effect on the number of degrees conferred. Similarly, graduation rates survey provides data on the graduation proportions within cohorts that finish the program 100% of normal time (4 years for bachelor degree), 150%, or 200%. These variables are available in NCES IPEDS Data Center by selecting all institutions and choosing graduation rate variables for 4-year and 2-year institutions.

B.3 Marijuana legalization, use, and prices

This section provides an overview of recreational marijuana legalization policy and investigates the trends in marijuana use and prices across states that have adopted this policy and those that have not. The timing of legalization varies because it depends on the time gap between passing the law and the actual availability of cannabis in retail stores. Furthermore, this section explores the significant disparities in marijuana use and prices among states according to their legalization status.

Policy definition

I use different sources to establish the time upon which each state legalized RM. I also take into account the year that the law is implemented, which refers to the opening year of the first dispensary stores. Based on Carnevale Associates (2022), Marijuana Policy Project (2022c), Kim, O'Connor, and Norwood (2020), and online searches, Figure B.1 depicts the timeline of marijuana legalization for each state. I distinguish between the year when the legalization became legal (law) and the year in which the law went into effect by the local government distributing retail licenses (first store or dispensary evidence). States that legalized marijuana by law but did not implement the law are excluded from the data.[10] Hence, this study focuses on states that legalized marijuana not just legally but also practically by opening dispensaries. Among all the states that legalized marijuana within our data period (2009-2019), six states satisfy the practical legalization condition: Colorado, Washington, Oregon, California, Massachusetts, and Nevada. Figure 1 shows that it takes these treated states one to two years after legalization to open the first retail store.

Marijuana use and prices, by legalization status

The upswing in marijuana use in RM jurisdictions is not surprising as it is already well documented based on survey data that marijuana legalization results in higher levels of consumption

[10]Maine and Michigan are excluded from the analysis due to the timing of their marijuana retail store openings, which occurred in or close to the year 2020 (Marijuana Policy Project, 2019, 2022b).

and addiction (Cerdá et al., 2017, 2012; Dave et al., 2023; Martins et al., 2016; Wen et al., 2015). To further support this literature, I utilize Google Trends data to depict differences in online searches for the word "dispensary" between RM and non-RM states. Google Trends refers to a normalized number (hits) that is between zero and one hundred, depicting the popularity of a search term among other searches at each state and time.[11] Google Trends measures are based on unbiased samples of all Google searches. I use the Google Trends measure related to the word "dispensary" as a proxy to gauge the demand for marijuana because people usually look up the location of marijuana stores by searching the word "dispensary". Figure B.2 (a) shows that the search for the word "dispensary" is most popular in RM states as opposed to non-RM states.

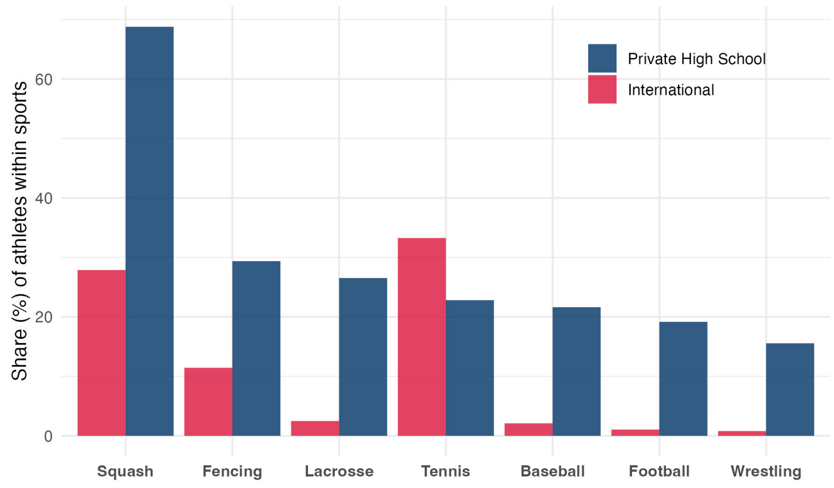
The price of marijuana in Figure B.2 (b), on the other hand, is less expensive in RM states. The prices are lower in RM states most likely due to the transition of the industry toward a more competitive market structure. Some of the mechanisms deriving this price decrease in RM states include the reduction of the risk premium and increased supply (Hall & Lynskey, 2016). Hence, colleges located in RM states could gain a competitive advantage if students view marijuana as another consumption amenity, especially due to its availability and low prices.

One valid argument against the RM policy effects is that it may not be binding if people can still purchase cannabis in states that deem marijuana illegal. In response, non-RM states have harsh punishments for small possession of marijuana. Further, possession of marijuana is still considered a federal offense with punishment on the first conviction that goes up to one-year imprisonment with a fine (Norml, 2020). Another inhibitor of seeking cannabis illegally is related to its high price in non-RM states. Figure B.2 panel (b) shows that the price of marijuana is about two times higher in non-RM states than in RM states.

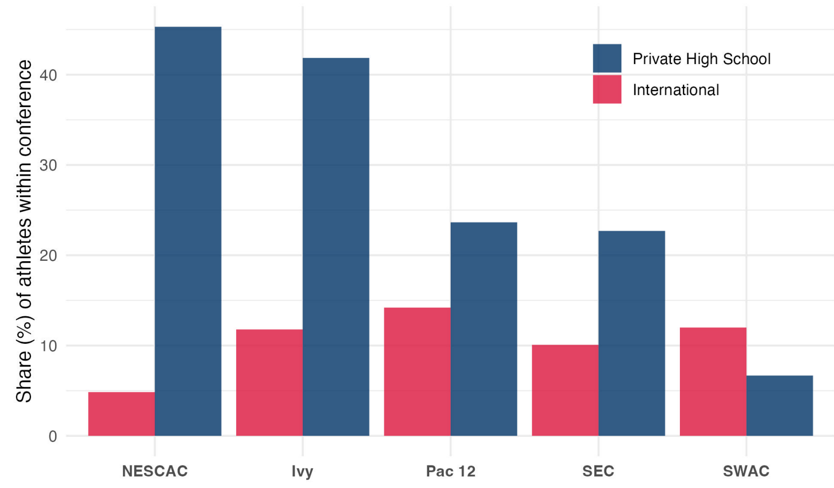
[11] Rogers, Simon (2022) discusses details about Google Trends measurement.

C.1 Chapter 3

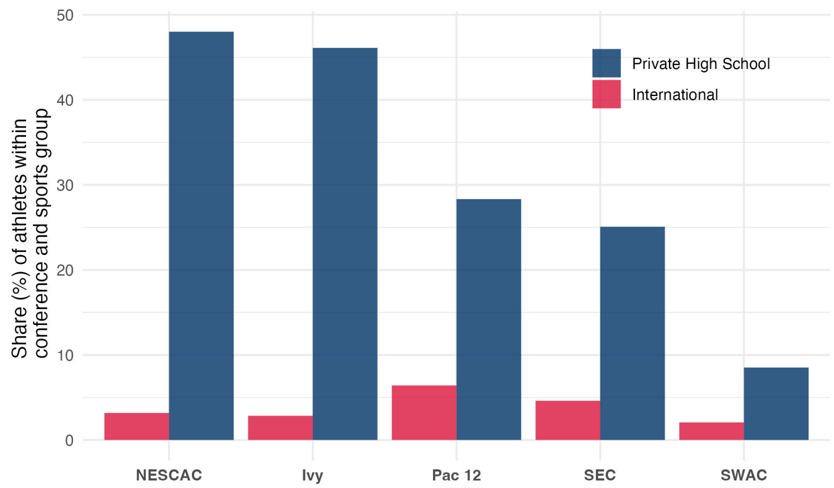
Figure C.1: Share of Athletes from Private High Schools and International Backgrounds by Sport and Institution Type



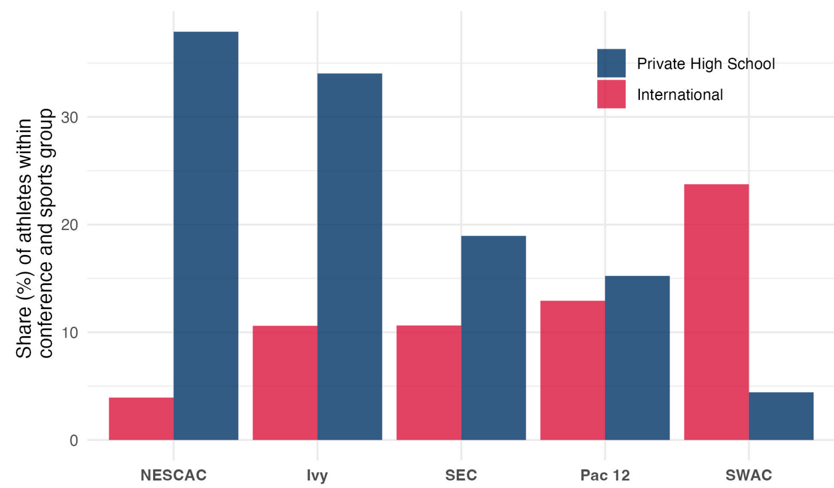
(a) By sport, all athletes



(b) By athletic conference, all athletes



(c) By athletic conference, football and basketball players only



(d) By athletic conference, cross-country and soccer players only

NOTES.— Private high school and international representation varies substantially across sports. Elite sports like squash and fencing show the highest proportion of private school athletes, while wrestling and football draw more heavily from public schools. Elite conferences (Ivy, NESCAC) recruit significantly higher percentages of private school students than other major conferences. International recruitment also varies across conference types.

Table C.1: Summary Statistics for High Schools, by School Sector and Sending an Athlete Status

	Private					Public				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Panel A: Schools that sent an athlete										
Number of students	4098	384.79	379.47	1.00	4789.00	12745	1022.05	794.93	1.00	9654.00
Share of white students	4098	67.41	27.18	0.00	100.00	12745	58.74	31.84	0.00	100.00
Share of African Amr. students	4098	12.17	19.42	0.00	100.00	12745	14.58	22.76	0.00	100.00
Share of Hispanic Amr. students	4098	9.30	14.91	0.00	100.00	12745	18.35	23.21	0.00	100.00
Share of Asian Amr. students	4098	6.19	9.88	0.00	100.00	12745	3.79	7.73	0.00	99.15
Student to faculty ratio	4098	10.78	20.12	0.20	812.50	12617	16.73	8.41	0.00	434.00
Large city location dummy	4098	0.19	0.39	0.00	1.00	12745	0.10	0.31	0.00	1.00
Small or midsize city location dummy	4098	0.19	0.40	0.00	1.00	12745	0.11	0.31	0.00	1.00
Suburb location dummy	4098	0.34	0.48	0.00	1.00	12745	0.31	0.46	0.00	1.00
Town location dummy	4098	0.08	0.27	0.00	1.00	12745	0.15	0.36	0.00	1.00
Rural location dummy	4098	0.19	0.39	0.00	1.00	12745	0.32	0.47	0.00	1.00
HS zip code average salary	3932	72.58	53.25	17.79	974.30	12454	56.94	30.51	18.11	588.73
Top 100 private high school dummy	4098	0.02	0.15	0.00	1.00	0				
Catholic affiliation	4098	0.26	0.44	0.00	1.00	0				
Other religious affiliation	4098	0.45	0.50	0.00	1.00	0				
Nonsectarian affiliation	4098	0.29	0.45	0.00	1.00	0				
Number of free or reduced price lunch recipients	0					11856	452.15	469.19	0.00	7503.00
Panel B: Schools That Did Not Send an Athlete										
Number of students	8810	134.32	264.97	1.00	7701.00	12471	275.85	417.55	1.00	14 319.00
Share of white students	8810	65.87	32.65	0.00	100.00	12471	50.35	35.12	0.00	100.00
Share of African Amr. students	8810	15.23	24.49	0.00	100.00	12471	15.58	24.40	0.00	100.00
Share of Hispanic Amr. students	8810	11.16	19.02	0.00	100.00	12471	24.59	28.96	0.00	100.00
Share of Asian Amr. students	8809	3.93	11.21	0.00	100.00	12471	1.77	4.94	0.00	100.00
Student to faculty ratio	8810	10.11	31.28	0.22	2073.68	11539	15.04	15.56	0.00	665.00
Large city location dummy	8355	0.16	0.37	0.00	1.00	12471	0.17	0.37	0.00	1.00
Small or midsize city location dummy	8355	0.15	0.36	0.00	1.00	12471	0.12	0.32	0.00	1.00
Suburb location dummy	8355	0.33	0.47	0.00	1.00	12471	0.20	0.40	0.00	1.00
Town location dummy	8355	0.11	0.31	0.00	1.00	12471	0.14	0.34	0.00	1.00
Rural location dummy	8355	0.25	0.43	0.00	1.00	12471	0.38	0.48	0.00	1.00
Number of free or reduced price lunch recipients	0					11249	174.90	298.66	0.00	10 848.00
Top 100 private high school dummy	8810	0.00	0.00	0.00	0.00	0				
Catholic affiliation	8810	0.05	0.23	0.00	1.00	0				
Other religious affiliation	8810	0.62	0.49	0.00	1.00	0				
Nonsectarian affiliation	8810	0.32	0.47	0.00	1.00	0				

Table C.2: Variable Descriptions

Variable	Description
Predicted Revenues	Expected athletic revenues for university i if it selects bundle j
Predicted Expenses	Expected athletic expenses for university i if it selects bundle j
Predicted ZIP Code Salary	Average salary in the ZIP codes where recruited athletes attended high school
Predicted % International	Percentage of athletes who are international (non-US citizens)
Predicted % Private HS	Percentage of athletes who attended private high schools
Has Football	Binary indicator for whether the bundle includes football
Men's/Women's Standards Categories	Classification of sports quality/competitiveness standards (Categories 1-3)
Men's/Women's Regional Categories	Classification of sports by regional representation (Categories 1-3)
Men's/Women's Niche Categories	Classification of sports by popularity/niche status (Categories 1-3)

Table C.3: Data Sources and Variable Descriptions

Dataset	ID Variables	Source	Description
Athletic Rosters	University ID, Sport ID, Athlete Name	University Athletic Department Websites	Individual-level data on NCAA athletes including name, sport, academic year, hometown, and high school for 2019-20 academic year
CCD	NCES School ID, ZIP Code	National Center for Education Statistics	Public secondary school characteristics including location, enrollment, student demographics, and staffing for 2017-18 academic year
PSS	NCES School ID, ZIP Code	National Center for Education Statistics	Private secondary school characteristics including location, enrollment, student demographics, religious affiliation for 2017-18 academic year
IRS SOI	ZIP Code	Internal Revenue Service	ZIP code level tax return data including income distribution, number of returns, and types of income for tax year 2017
ACS	ZCTA	U.S. Census Bureau	ZIP Code Tabulation Area (ZCTA) level demographic and socioeconomic characteristics from 2013-2017 5-year estimates
IPEDS	University ID	National Center for Education Statistics	University-level enrollment, financial, and institutional characteristics for 2019-20 academic year
NCAA Directory	University ID	NCAA	Sports offered and athletic division/conference membership for 2019-20 academic year
EADA	University ID, Sport ID	U.S. Department of Education Office of Postsecondary Education	Sport-specific revenues, expenses, participation, and coaching staff data at university level for 2019-20 academic year
Opportunity Atlas	University ID	Opportunity Insights	University-level data on student and parent income distributions, student outcomes, and mobility rates based on 1980-1991 birth cohorts

Notes: CCD refers to Common Core of Data; PSS refers to Private School Survey; IRS SOI refers to Statistics of Income; ACS refers to American Community Survey; IPEDS refers to Integrated Postsecondary Education Data System; EADA refers to Equity in Athletics Data Analysis. ZCTA refers to ZIP Code Tabulation Area, which is the Census Bureau's statistical equivalent of ZIP codes. Opportunity Atlas data were constructed using population-level federal income tax returns and other administrative data as described in Chetty et al. (2020).