Title
Essays on Higher Education Effectiveness and Budget Decisions

Author
Zhou, Quanfeng

Publication Date
2022

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Essays on Higher Education Effectiveness and Budget Decisions

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Economics

by

Quanfeng Zhou

September 2022

Dissertation Committee:

Dr. Michael Bates, Chairperson
Dr. Ozkan Eren
Dr. Ruoyao Shi

The Dissertation of Quanfeng Zhou is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I am incredibly grateful to my advisor, Dr. Michael Bates. This dissertation would not have been made possible without his guidance. He invested much of his time and effort in teaching me how to become a researcher. He has always been supportive and encouraging, especially during many of my difficult times. I feel very fortunate to have him as my advisor.

I would also like to thank my dissertation committee members for their help along the way. I want to sincerely thank Dr. Ruoyao Shi for hearing out my research ideas and for providing insightful and constructive comments on my research. Those comments make significant improvements to this thesis. I am also grateful to Dr. Ozkan Eren for his guidance on many methodological issues that strengthen the credibility of my empirical findings. Beyond their help academically, they have also provided many invaluable tips and opportunities for me to get started in this profession.

The staff members in the department, my classmates, and my friends are also important reasons that I could finish this thesis. I want to thank our graduate student assistant Mr. Gary Kuzas for his constant and continuing help. I also feel very grateful to have met Yaojue Xu and Zhuozhen Zhao in our program. The comments and feedback from my peers and the support and encouragement from my friends were vital in keeping me moving forward.

Lastly, I would like to thank my parents, Ping Chen and Jiahua Zhou, for always believing in me and supporting the choices I make. Thirteen years have been a long time for me to live halfway across the world. I am thrilled that I am finishing this journey and will move on to a new one where I can be closer to my parents.

To my parents for all the support.

ABSTRACT OF THE DISSERTATION

Essays on Higher Education Effectiveness and Budget Decisions

by

Quanfeng Zhou

Doctor of Philosophy, Graduate Program in Economics
University of California, Riverside, September 2022
Dr. Michael Bates, Chairperson

This dissertation consists of three essays that study higher education effectiveness and budget decisions. In the first chapter, I evaluate whether instructional spending in colleges, the main input in the human capital production process, is effective in promoting student earnings after graduation. Using a nationally representative dataset, I find a positive elasticity between instructional spending and student earnings of around 2 percent which is robust to various specifications and potential confounding factors. The effects are mostly driven by private institutions and four-year institutions. The effects are slightly lower for well-established public institutions. Cost-benefit analysis reveals that a student has to work for over 40 years until her increased earnings can cover the cost of the spending, indicating the cost-ineffectiveness of increasing instructional spending from an investment perspective.

In the second chapter, we analyze one driving factor for college budget decisions - the competition for good students. We construct a model to show that colleges choose their share of spending on educational inputs and on amenities to attract good students. Their optimal decision depends on their competitors' choices. Simulation results reveal

that over the range where budget choices are commonly made, colleges respond positively to their competitors' decisions. In addition, when the competitive pressure from competitors weakens, such responsiveness also declines. Empirical evidence generally supports the predictions from the simulation exercise.

In the third chapter, we propose a method that bounds the treatment effect for the general population of interest under randomized controlled trials. Our bounds depend on a set of mild assumptions that are different from existing methods in the literature. Hence, we view our method as an alternative for applied researchers to choose from. Applying our method to analyze the effectiveness of a first-year learning program on retention rates, our bounds suggest a possible negative population effect and restrict the possibility of a large positive population effect, despite the estimated positive effect for the experimental sample. We also show that assumptions required for other methods in the literature to be applicable are unlikely to hold in this context, highlighting the importance of the availability of our new method.

Contents

List of Figures

List of Tables

Chapter 1

Does Higher Instructional Spending in Colleges Promote Student Earnings?

1.1 Introduction

More money is spent on fewer college students in the United States now than a decade ago. In real terms, between academic year (AY) 2009-10 and 2018-19, expenditure per full-time equivalent (FTE) student increased by 24.6% for public institutions and 16.4% for private nonprofit institutions.[1] The large increases in per-student spending result from both the growth of total spending and the decline in enrollment. In AY2018-19, all post-

[1] Table 334.10 and Table 334.30, Digest of Education Statistics 2020, National Center for Education Statistic (NCES). In 2019 constant dollar, expenditure per full-time equivalent (FTE) student was $31,076 in AY2009-10, and $38,709 in AY2018-19 for public institutions. In private institutions, expenditure per FTE student was $54,416 in AY2009-10, and $63,321 in AY2018-19.

secondary institutions spent a total of $626 billion on instruction, institutional support, research, academic support, and other services, rising from $535 billion in AY2009-2010. Over the same period, fall enrollment for degree-seeking students declined from 21 million to 19.6 million.[2] With the large increases in per-student spending, it is natural to ask whether these increases are worthwhile.

Instructional spending is the largest spending category for most institutions, while it has been decreasing as a share of total spending.[3] Figure 1.1 depicts the growth of instructional spending and total spending. We see that its growth lagged the growth of total spending, and indeed, its share decreased from 41% in 1986 to 29% in 2018. What might justify that institutions are allocating relatively fewer funds to instruction? This paper investigates the (in)effectiveness of instructional spending in promoting students' earnings after graduation.

Figure 1.2 plots the percentage change in average wage earnings for young workers (age 22-30) with at least an associate degree between 2012 and 2019 against the percentage change in instructional spending per college student between 2009 and 2016 in each state. The dashed line is a linear fit weighted by the population of such young workers in 2019. We observe a clear positive correlation between the larger increase in spending and the larger increase in wage earnings of young workers.[4] Whether such positive correlation is caused by

[2]Total spending is in 2019 constant dollars. Author's calculation based on data from the Integrated Postsecondary Education Data System (IPEDS). Fall enrollments are retrieved from Table 303.10, Digest of Education Statistics 2020, National Center for Education Statistic.

[3]According to the IPEDS, instructional spending consists of expenses related to general academic instruction, occupational and vocational instruction, community education, preparatory and adult basic education, and regular, special, and extension sessions. It may also consist of research, public service, and information technology related expenses if an institution does not separately budget those expenses.

[4]The weighted correlation coefficient is 0.39. This is largely driven by the five largest states: California, Texas, Florida, New York, and Pennsylvania which are almost perfectly correlated with a correlation coefficient above 0.96, either weighted or unweighted. The five largest states also correspond to a much steeper slope. The unweighted correlation coefficient is 0.13 among all states.

any confounding factors or if there is a causal relationship between instructional spending and student earnings remains an open and important question to investigate.

To address the above issue, this paper employs the two-way fixed effects regression using data from the College Scorecard project to estimate the impact of instructional spending on student earnings. In the main specification, I include the institution fixed effects, cohort fixed effects, state-specific time trends, and a rich set of control variables. In addition to the average earnings, I also examine the effects over the 25th, 50th, and 75th percentiles in the earning distribution, for earnings six, eight, and ten years after their first attendance. Therefore, this paper draws a complete picture of the relationship between instructional spending and student earnings in the short term after graduation.

Under the main specification, I find that the elasticities of instructional spending on earning outcomes are 0.5, 1.7, and 1.9 percent for average earnings six, eight, and ten years after first attendance. They are higher on lower percentiles in the earning distribution, ranging from 0.3 to 3 percent. To put the above numbers into perspective, evaluated at the mean values of spending and earnings, a 10% increase of $806 in instructional spending (per year) will lead to an increase in wage earning by $71 eight years after a student's first attendance. After interacting instructional spending with the type of institution, I find that the results are driven mainly by private institutions and four-year institutions. The estimated elasticities are from 0.6 to 3.7 percent for private institutions, while only between 0.3 and 0.8 percent for public institutions. The estimated elasticities are between 0.6 to 5.9 percent for four-year institutions. The data that I use consists of more than 6,700 institutions, representing nearly 95% of all postsecondary institutions in the United States.

Therefore, the above results are very likely to be externally valid.

To provide evidence that the estimates are not driven by confounding factors, I employ the sensitivity analysis proposed in Oster (2019) that examines to what extent the estimates can survive a confounding factor that will bring them towards zero. The results suggest that most of my estimates can survive a confounding factor of 0.5, that is, the selection on unobservables being half as strong as the selection on all observed control variables. A third of my estimates can survive a confounding factor close to 1, a level that 55% of published work using observational data cannot, as studied in Oster (2019).

This paper also uses the instrumental variable (IV) regression to estimate the effect of instructional spending on student earnings. The instrument variable, state-level budget shock multiplied by an institution's historical financial dependence on state appropriations, was constructed in Deming and Walters (2017).[5] I use the public institutions from my analysis sample for this exercise as the instrument is only applicable to public institutions. Although the instrument has high predictive power in the first stage of my treatment variable, instructional spending, the second stage was imprecisely estimated with large standard errors. The result from the IV regression is unsurprising, given that I find small and statistically insignificant effects for public institutions in the two-way fixed effects regressions.

I evaluate the cost-effectiveness of instructional spending by comparing the investment with the discounted series of earning increases, assuming my estimated effects are persistent throughout a student's lifetime. Using a highly forward-looking discount factor

[5]In their paper, they use two instrument variables (1) state-level legislative cap on tuition fee increases, and (2) budget shock multiplied by historical dependence to estimate the impact of two treatment variables (a) tuition fee, and (b) spending on college completion rate.

of 0.9975, a student has to work for 42 years (37 years for private institution and 58 for public institution attendees) until the net present return becomes positive. Alternatively, assuming a student works for 45 years after graduation, for the net present return to be non-negative, the discount factor needs to be 0.9946 (0.9887 for private institution and 1.0066 for public institution attendees). Commonly, we think people and society as a whole are not as forward-looking. Therefore, my results provide a possible explanation for why institutions have been adjusting and lowering the share of spending on instruction.

This paper fits in the extensive literature on estimating the return to school quality. The largest strand of the literature focuses on the return to school spending before college on various outcomes, including test pass rates, college enrollment, years of education, and earnings in the labor market (Grogger (1996); Papke (2005); Lafortune et al. (2018); Hyman (2017); Jackson et al. (2016)), and conclude positive effects, but only recently. Earlier studies found that throwing money at public schools is ineffective in promoting students' performance (e.g., Hanushek (1981)). Another strand of the literature focuses on the earning returns to college quality, examining other dimensions of quality such as peer quality (Dale and Krueger (2002, 2011)) and selectivity of the college (Behrman et al. (1996); Hoekstra (2009)). A set of recent papers estimate the impact of college spending on outcomes such as enrollment, persistence, and completion (Webber and Ehrenberg (2010); Webber (2012); Deming and Walters (2017)).

This paper fills the gap by examining the effect of college spending on students' labor market earnings. We care about other outcomes, including persistence and completion, because we believe they map into earnings. There is an early wave of papers that directly estimate the impact of college spending on earnings, using cross-sectional data from Project Talent that surveys World War II veterans (Solomon (1975); Wachtel (1976); Foster and Rodgers (1980)). They estimate an elasticity coefficient between 10 to 20 percent, which is much higher than I find.

This paper is more closely related to the study by Griffith and Rask (2016). By matching individual-level data to institution characteristics from IPEDS and creating a cross-sectional dataset, they estimate the relationship between college spending and student earnings. The individual-level data are a subset of the students from the Education Longitudinal Study of 2002 (ELS:02) who had valid labor market outcomes and chose to go to four-year colleges. They employ an ordinary least square (OLS) estimation strategy and include a large set of control variables for their identification. An important variable, the average spending at other institutions the student applied to, is argued to capture the unobserved quality of the student and would hence take care of the selection of students into institutions. They use the Heckman selection model to account for attrition and selection into college.

This paper exploits a newly available dataset that allows me to apply a different empirical methodology for identification. Using institution-level panel data, I include institution fixed effects that account for any time-invariant unobserved variables of institutions that could be a confounding factor and unaccounted for when using cross-sectional data. In addition, the subset of the survey participants in the ELS:02 who chose to go to college might not be nationally representative, causing concerns about both the internal and external validity of the results. The data used in this paper cover almost all major postsecondary degree-granting institutions in the United States and is much less concerned with the issue of external validity. I also include cohort fixed effects and state-specific time trends to capture general movements in labor market conditions both nationally and statewide. My identifying variation is exogenous to those confounding factors.

My results also echo a more recent paper by Mountjoy and Hickman (2021). Using more granular data from the state of Texas, they identify that increase by a standard deviation in a composite index of non-peer inputs, which includes instructional expenditures, full-time faculty share, tenure-track faculty share, and faculty-student ratio, predicts a \$753 additional earnings value-added. My results suggest that a standard deviation ($\$5,935$) increase in instructional spending (on the average of $\$8,066$) leads to an increase in average earnings by \$601, eight years after first attendance.

This paper proceeds as follows. I describe the data and background information in Section 1.2. Section 1.3 contains the empirical specification and discusses the identification strategy. The results are presented in Section 1.4. Section 1.5 considers using an instrumental variable regression approach and discusses its results. Section 1.6 conducts a cost-benefit analysis of the results and concludes.

1.2 Data and Background

I need a dataset containing both students earning outcomes and college education information to answer my research question. Ideally, linking education records to earnings nationwide generates a most comprehensive dataset for the analysis. However, such a project requires enormous resources and has not been accomplished so far to the best of my knowledge.[6] As the independent variable of interest, instructional spending per student, varies only at the institution level, unlike some other variables of interest in related research questions such as student loan status and academic performance that change at the student level, this paper utilizes a nationally representative dataset from the College Scorecard project, where instructional spending per student and earning outcomes are available at the institution level for different cohorts.

The project was created in 2013 under President Barack Obama's administration to make data about colleges more accessible to the general public. It combines data from multiple existing systems and government agencies, including Integrated Postsecondary Education Data System (IPEDS), Department of Treasury, and National Student Loan Data System (NSLDS).[7] The project linked the student-level earnings data from administrative tax records maintained by the Department of the Treasury to the records in NSLDS. The linked records were then aggregated to the institution level and matched to institution characteristics that have been collected annually by the IPEDS.

There are two important advantages of using this data. First, the data are highly representative of institutions in the United States. All postsecondary institutions participating in any federal financial assistance program authorized by Title IV of the Higher Education Act are mandatory to report their data to the IPEDS, making it cover almost all postsecondary institutions in the United States. I observe at least 6,400 unique insti-

[6] Many researchers studying closely related questions collaborated with local governments and universities to match the state Unemployment Insurance (UI) data to the education records. For example, Mountjoy and Hickman (2021) focus on the state of Texas. However, States vary significantly in observed and unobserved ways, raising the issue of external validity. In addition, State UI data cannot capture students who work in a state different from where they attend college.

[7] Other agencies include Department of Education, Office of Postsecondary Education (OPE) and Federal Student Aid (FSA). They facilitated the linking process for the College Scorecard project and provided other variables that are not used in this paper.

tution identifiers each year. Among all institutions, 4,355 appear in all years, of which 3,824 have valid spending information in all years, representing between 84% to 90% of the total student enrollment. My analysis sample represents 89% of all undergraduate student enrollment.[8] Second, this is an institution-level panel dataset starting from 1996 covering topics including cost, enrollment, completion, finance, demographic and socioeconomic backgrounds of students, program offerings, awards and scholarships, and many others, allowing me to control for a rich set of the institution and student characteristics.

This data also have the following complications. First, the earning outcomes are estimated for undergraduate Title IV recipients working and not enrolled in graduate programs when earnings are measured. The share of each institution's entering class represented by Title IV students can vary substantially due to differences in family income of students attending those institutions and state and institutional aid policies. Second, the earning statistics are calculated based on pooled two-year cohorts. On the one hand, this reduces measurement error and can lead to more stable estimates. On the other hand, it shortens the panel, making it harder to identify an effect as there are fewer variations in the independent variable. Third, earning outcomes are suppressed for confidentiality reasons for cohorts less than 30, leading to missing values in the earning outcomes for about one-fifth of the institution-cohort observations. Those are typically small institutions. Last, spending is reported based on fiscal years (Oct 1st to Sept 30th), while expenditures commonly affect students throughout academic years (Sept 1st to Aug 31st). As a check for robustness, I adjust for this difference in calculating the average spending. I complement the College Scorecard project data with the State Higher Education Finance (SHEF) project for my

[8] I discuss in more detail how I construct my analysis sample in Section 1.2.2.

instrumental variable analysis.[9] The variables I use are state-level appropriations for public institutions and the number of FTE enrolled students each year. In addition, I use the variables state appropriations to each public institution and the total revenues from the finance component of IPEDS in 1990.

I use the Consumer Price Index (CPI) retrieved from the Bureau of Labor Statistics to deflate the earning variables and the Higher Education Price Index (HEPI) retrieved from Commonfund Institute to deflate the spending variables.[10] All monetary terms are in 2014 dollars unless otherwise stated.

1.2.1 Data Structure

The original data are a panel of institutions over the years. I transformed it into an institution-cohort panel. For example, in the 2009 data file, the variable average earning ten years after entry is essentially the average earning of the 1999 cohort, measured in 2009, ten years after they entered college. Assuming it is a four-year institution, I can go to the data file from 1999 to 2002 to match the spending information and other institution characteristics to the earning variables of this cohort. In addition, in the 2007 data file, the variable average earning eight years after entry measures the earnings of this same 1999 cohort in 2007.[11] Therefore, I observe multiple earning outcomes for a given cohort. Following the above transformation, I end up with a panel where earnings six, eight, and

[9] The State Higher Education Finance (SHEF) project is under the State Higher Education Executive Officers (SHEEO) Association. Deming and Walters (2017) used the same data.

[10] The HEPI is an inflation index specifically designed to reflect price levels for higher education inputs. All results are qualitatively the same if I use CPI to adjust for price level changes for my spending variable.

[11] The two cohorts may not be identical, though, as the earnings are only measured for those who are working. The employment status could differ between 2007 and 2009. The two cohorts would still overlap significantly.

ten years after entry are measured for cohorts from 1996 to 2007 (with gaps). In addition to the average earnings, I also observe the 25th, 50th, and 75th percentile of the earnings for each cohort. Figure 1.3 shows the availability of the earning variables by cohort. I have a panel of eight, seven, and six cohorts for earning outcomes six, eight, and ten years after entry, respectively.

1.2.2 Sample Restrictions

After the transformation, the panel consists of 51,779 institution-cohort observations where the variable instructional spending and at least one of the earning outcome variables are non-missing. I call this the "Full Sample".

I focus on the institutions that remain consistent (two-year vs. four-year) levels throughout the period to avoid dramatic institutional upheavals that potentially affect spending and student earnings. Change in institution levels also leads to ambiguity in calculating the average instructional spending for a cohort. This restriction leaves out 5,125 institution-cohort observations.

Further, I exclude observations where the spending variable takes on extremely large or small values. For instance, the highest value in the data is 29 million dollars per student per year in instructional spending, which comes from a nursing school in Rhode Island. Therefore, I follow the literature and take the threshold used in Deming and Walters (2017) to admit only observations with spending between \$50 and \$100,000 per year. This restriction takes out another 4,410 institution-cohort observations, many of which are highly specialized, such as training institutions for aviation and performing arts.[12]

[12]Their students most likely work in highly isolated labor markets. Therefore, excluding those institutions

The remaining data with 42,244 institution-cohort observations form my analysis sample, covering a total of 5,360 institutions.[13] I take the average of the instructional spending over the two years (four years) starting from the year of entry of the cohort in two-year (four-year) institutions.[14] Lastly, I average over two adjacent cohorts to correspond to the earning outcomes that are measured for two-year pooled cohorts.

1.2.3 Summary Statistics

Table 1.1 presents summary statistics for the outcome variables. In calculating the summary statistic, I weighed the sample by the undergraduate enrollment. In the full sample, the average earnings for students six, eight, and ten years after entering college are $36,204, $41,459, and $46,090, respectively (in 2014 dollars). The median earning is lower than the average, being $32,922, $37,248, and $40,799, respectively. In the analysis sample, average earnings are $36,626, $41,980, and $46,700, which is about $400-$600 higher than that in the full sample, or by roughly 1 to 1.5%. The numbers suggest that students graduating from institutions that have changed their levels or have reported extremely high or low values of spending per student earn relatively less, though the difference is small.

In Table 1.2, I present summary statistics for the treatment variable and the set of control variables. First, let us focus on instructional spending. The average instructional spending per student in the full sample is $9,722. The extremely high entries of spending lead to not only a higher level of average spending, even after being weighted by undergraduate enrollments, but also an implausibly large standard deviation of $129,516. In the

does not have an enormous impact on the population of general interest.

[13] Not all institutions appear in all years.

[14] As noted in Section 1.2.1, I average over an additional year to check for robustness due to the slight misalignment between the fiscal year and academic year. Results are almost identical and are not presented.

analysis sample, average instructional spending per student is $8,066, with a much more reasonable standard deviation of $5,935.[15] In Figure 1.4, I plot the kernel density of the log instructional spending in the full sample and the analysis sample. The log values of the two samples are centered at the same point. A small amount of the mass in the tails in the full sample is moved closer to the mean in the analysis sample, as I exclude extreme values. Overall, they align nicely.

The first set of control variables in the analysis sample is measured for the entering cohort. They include the average SAT score (1,076), the average family income ($40,244), the average median household income ($82,677), and the average age at entry (23.9). Another set of the control variables are demographic characteristics averaged over the region where students of the entering cohort come from. They include the percent of White, Black, Asian, and Hispanic from students' zip codes through the Census data, as well as the percent holding Bachelor's degree, Graduate or Professional degree, born in the U.S., and the local unemployment rate and poverty rate. I also control for characteristics of all the enrolled students in the institution, namely, the fraction that is female, married, dependent, veteran, and first-generation college students. For all control variables, the analysis sample has almost identical means to the full sample.

From the above comparisons, the analysis sample is representative of the characteristics of postsecondary institutions and college students in the United States. It covers 5,360 institutions, representing 89% of all undergraduate enrollment during its period. To

[15]I compare $8,066 (in 2014 dollars) to similar statistics found in Table 373 in the Digest of Education Statistics 2010, published by the National Center for Education Statistics. In the Digest, instructional spending per student from AY2003-04 to AY2007-08 ranges between $7,792 and $8,221, after adjusting to 2014 dollars. The average $8,066 lies perfectly within the range, while $9,722 is unreasonably too high. I note the difference that my average is calculated over AY1996-97 to AY2007-08, a more extended period than that reported in the Digest.

avoid losing observations due to missing values resulting from item non-response, I recoded the missing values of the control variables to negative ones. I then added indicator variables that equal one if the corresponding data element is missing and zero otherwise.

1.3 Model and Specification

1.3.1 Model

I consider the relationship between spending and student earning as a production process, as discussed in Black and Smith (2006). Following their formulation, consider the education production function

$$Y = f(x_1, ..., x_k, U, \varepsilon), \tag{1.1}$$

where Y denotes student log earnings. I denote different educational inputs such as spending, peer quality, student-to-faculty ratio, and others with $x_1, ..., x_k$. U represents other observed factors affecting student earnings, and ε is an idiosyncratic error.

I am interested in estimating the parameter

$$\beta = \frac{\partial E(Y|x_1, ..., x_k, U)}{\partial x_1}, \tag{1.2}$$

where x_1 is the instructional spending per student. This parameter is of particular interest to policymakers in making spending decisions, especially when facing budget shocks.

It is implausible to argue that one can exhaust the list and control for all educational inputs $x_1, ..., x_k$, nor do I make this claim. Black and Smith (2006) lament that most existing empirical work that (implicitly) claims to have estimated the parameter β

only controls for a single input x_1. They point out that controlling for only x_1 leads to the estimation of a different parameter

$$\beta' = \frac{\partial E(Y|x_1, U)}{\partial x_1}, \tag{1.3}$$

that is of no clear economic interpretation and empirical relevance. In addition, omitted variable bias arises when the researcher fails to control for another input x_j that is correlated with x_1. For instance, they showed that five measures or proxies of input (faculty-student ratio, rejection rate, freshman retention rate, mean SAT score, and mean faculty salaries) are positively correlated with the coefficient of correlation ranging from 0.3 to 0.7. Therefore, β' will likely pick up part of the effects of other inputs and hence be greater than β. Indeed, I find similar correlations in my data. Using three variables from my data: average SAT score of the entering class, instructional spending per student, and first-year retention rate, I demonstrate their correlation coefficients in Table 1.3. In the top panel, they range from 0.4 to 0.7, very close to what was found in Black and Smith (2006).

I make two improvements that significantly ameliorate this issue. First, I control for another input: the average SAT score of the entering class. Consider the first-year retention rate as an omitted variable. In the top panel of Table 1.3, the correlation coefficient between instructional spending and retention rate is 0.45. I show in the mid-panel that conditional on the average SAT score, the remaining variation in instructional spending and retention rate has a correlation coefficient of only 0.03. Controlling the average SAT score could greatly reduce bias caused by the omission of other inputs that are correlated with the SAT score. Second, due to the panel nature of my data, any inputs that are not time-varying are controlled by including institution fixed effects, even if unobserved or unquantifiable, such as the location of the institution. For time-varying inputs, again, consider the first-year retention rate as the omitted variable. The retention rate and instructional spending are both correlated to the level of prestigious of the institution, which is unobserved and cannot be controlled. By subtracting from the two variables their institution average, in the bottom panel of Table 1.3, the correlation coefficient between the demeaned values is only 0.05, much smaller than 0.45 in the top panel. Therefore, including institution fixed effects could also greatly reduce the potential bias caused by time-varying inputs that are not controlled in the regression.

Therefore, I consider the parameter I estimate in the reduced form a close approximation to the parameter of interest β.

1.3.2 Two-way Fixed Effects Regression

I estimate the following reduced form equation using the analysis sample defined in Section 1.2.

$$\ln(Earning)_{ic} = \alpha + \beta \cdot \ln(InExp)_{ic} + \boldsymbol{X_{ic}} \cdot \boldsymbol{\gamma} + \boldsymbol{\kappa_i} + \boldsymbol{\eta_c} + \boldsymbol{\theta_{s(i)}} \cdot c + u_{ic} \tag{1.4}$$

where for institution i cohort c, $\ln(Earning)_{ic}$ is the log of one of the earning outcomes and $\ln(InExp)_{ic}$ is the log of the average instructional spending per student defined in Section 1.2.2. $\boldsymbol{X_{ic}}$ denotes the set of control variables described in Section 1.2.3. In addition, I also control for the program offering in each institution by including the share of degrees awarded to each Classification of Instructional Programs (CIP) code.

Exploiting the advantages of this panel data set, $\boldsymbol{\kappa_i}$ controls for the institution fixed effects. It is well understood that estimating using cross-sectional data without controlling for institution fixed effects is likely to suffer from omitted variable bias as I discussed in Section 1.3.1.

School spending and student earnings are likely to be simultaneously affected by business cycles. I include $\boldsymbol{\eta_c}$ as controls for the cohort fixed effects and $\boldsymbol{\theta_{s(i)}} \cdot c$ as controls for linear state-specific time trends to address the issue. The cohort fixed effects control for cohort-specific events at the national level, such as recessions. But that alone may not be enough, as labor market conditions can vary differently across the nation. I include the state-specific time trends to better address this problem. I do so at the state level for the following two reasons. First, for public institutions, spending changes are primarily affected by state legislation and state appropriations, which happens at the state level. Second, the time trends should ideally capture the movement in economic conditions in different labor markets. Each state can be viewed as a large and relatively concentrated labor market, though not perfectly isolated. It is inappropriate to consider each institution as a separate labor market, and hence I do not replace state-specific time trends with institution-specific time trends. The error term u_{ic} is assumed to be independent of $\ln(InExp)_{ic}$ after conditioning on all the control variables, fixed effects, and state-specific time trends.

I allow for arbitrary correlation in error terms within an institution by conducting inference using institution-clustered standard errors. If not taken into account, such correlations often vastly underestimate the standard errors of the estimated coefficients, leading to over-rejection of the null hypothesis. As a robust exercise, I also calculate the cluster-robust bootstrap standard errors that maintain the error structure within an institution by resampling at the institution level.

I start by estimating a regression that does not exploit the panel feature of my data. Pooling all the institution-cohort observations, I regress $\ln(Earning)_{ic}$ on $\ln(InExp)_{ic}$ and the controls only to mimic the estimation of early studies. I then add the fixed effects and eventually the state-specific time trends to see how the estimated coefficients evolve. The preferred specification is the complete Equation 1.4. Lastly, I interact the treatment with the type of institution to analyze effect heterogeneity.

1.3.3 Parallel Trends Assumption

The two-way fixed effects specification implicitly makes a parallel trends assumption for identification. Conditioning on all the control variables had institutions that experienced changes in instructional spending not undergo such changes; the change in their average earnings would have behaved in the same way as the institutions that did not experience changes in instructional spending, on average.

In standard difference-in-difference estimation, researchers adopting the parallel trends assumption typically plot a graph showing that the control and treatment groups' average outcomes are two parallel lines before the treatment starts (or in the pre-period). As instructional spending is a continuous variable and continuously changing, it is unclear which institutions are the control group and the treated group, and what periods are the pre-period and post-period. I attempt to mimic the idea and provide a figure evaluating whether the parallel trends assumption holds, while I bear in mind that this is at best suggestive. To do so, I regress the instructional spending on all the control variables and predict the residuals, which are my identifying variation. I cut my panel by the midpoint (the year 2002) into two artificial "pre-period" and "post-period".[16] I select the institutions that have a small standard deviation (below the median) in the identifying variation in the "pre-period". Among those institutions, I divide them into three groups: whether the instructional spending increased, decreased, or experienced little to no change ("control group") between 2002 and 2003, and plot the average earnings eight years after entry over time for each group.

The top panel of Figure 1.8 plots the histogram of the standard deviation in the identifying variation in the "pre-period". The distribution is highly skewed towards zero with a median of 0.103, indicated by the dashed line. Institutions below the median are included for analysis in the bottom panel. In the bottom panel, the red line (increased spending) and the green line (decreased spending) are shifted so that their levels match the blue line (no change, or "control group") in 2002. We see that there is a decreasing pre-trend for institutions with increased spending from 2002 to 2003 compared to the "control group". But instead of continuing to decrease, the average earnings increased relative to the "control group" when their spending increased in 2003. The other group of institutions where spending decreased from 2002 to 2003 does not exhibit an obvious pre-trend. There is a slight negative relative difference in earnings to the "control group" in 2003 when the spending decreased.

Although the above evidence is imperfect, it is suggestive that the parallel trends assumption can hold in my data.

[16]Realistically, the changes in treatment happen all the time, and it is not clear what pre-period and post-period should be.

1.3.4 Oster's Sensitivity Analysis

Even with both sets of fixed effects and state-specific time trends, other time-varying variables may exist that correlate with both spending and earnings. I follow the method first proposed in Altonji et al. (2005) and later formalized in Oster (2019) to analyze the sensitivity of the coefficients estimated using the two-way fixed effects regression to potential confounding factors.

The method involves two sensitivity parameters. The first parameter δ measures the ratio of selection on unobserved confounders to the selection on observed variables. The second parameter R_{max} represents the amount of variation in the outcome variable that would have been explained if all confounding variables had been included. If we choose $R_{max} = 1$, we assume that there is no idiosyncratic variation in student earnings that is uncorrelated with instructional spending.[17] I describe the approach in more detail in Appendix A.1.

With a choice of R_{max}, a researcher can ask two questions under this framework. First, how "strong" do the confounding factors need to be relative to the existing control variables in bringing the estimated coefficients toward zero? A "strong" confounding factor is one that highly correlates with the treatment variable. Second, for a given value of δ, what will the estimate of β be after adjusting for such a level of confounding?[18] An intuitive choice considered in Altonji et al. (2005) is $\delta = 1$, which represents the case where

[17]By choosing $R_{max} = 1$, we assume that all the variation in student earnings can be explained by either (1) instructional spending or (2) observed control variables including the fixed effects and state-specific time trends, or (3) unobserved variables that are orthogonal to the observed control variables and are correlated to instructional spending (confounders).

[18]Oster assumes the case where controlling for the unobserved confounders will cause the estimated coefficient to move in the same direction as controlling for the observed variables does. In this paper, they move towards zero.

the selection on unobserved confounders is equally important as the selection on observed variables.

A choice of R_{max} also has to be made. Oster (2019) studies 76 results in 27 papers published in top journals. She finds that a choice of R_{max} following the rule $R_{max} = \min(1.3\tilde{R}, 1)$ will allow 90 percent of the results using randomized data to survive a $\delta = 1$. The $\tilde{R}$ here is the R-squared in the controlled regression. In my case, it is the within R-squared in Equation 1.4. Choosing R_{max} using this rule has an intuitive understanding. When choosing control variables, researchers usually choose the ones that are most relevant to explaining the outcomes according to either theory or previous knowledge. Therefore, adding the remaining confounders may only explain a small additional proportion of the variation in the outcome variables. I follow Oster (2019) and set the $R_{max} = \min(1.3\tilde{R}, 1)$ in the analysis.

1.4 Empirical Results

1.4.1 Pooled OLS Regression Results

Table 1.4 presents the pooled OLS regression results. The estimates in column (1) suggest that the elasticities of instructional spending on earnings are 2.7, 3.5, and 4 percent for mean earnings six, eight, and ten years after entering college. Those are comparable in magnitude to other estimates in the literature using cross-sectional data. In column (2), I include only one additional control variable: the average SAT score of the entering class, while not exploiting the panel feature of my data. The estimated coefficients are brought down by a third. It implies that including only one educational input while not controlling for others can lead to biased estimates as many educational inputs are positively correlated. In the third column, I additionally include both the cohort fixed effects and institution fixed effects. The estimates for mean earnings eight and ten years after entry are further brought down by 30-40 percent, becoming 1.7 and 1.9 percent. The estimate for mean earning six years after entry is brought down by two-thirds and becomes insignificant. It is evident that using cross-sectional data and not controlling for the fixed effects also produces biased estimates. Lastly, in column (4), I use the full specification by including the state-specific time trends, and the estimates are not different from those in column (3).

1.4.2 Two-way Fixed Effects Regression Results

In Table 1.5, I present the results for all outcome variables using the specification described in Equation 1.4. I also summarize the estimated coefficients along with the confidence intervals in Figure 1.5 to visually present the findings. We observe positive estimated coefficients for all earning outcomes, though the coefficients for the earning outcomes six years after entry are not all statistically significant.

The first observation is that the effects are not obvious shortly after graduation (six years after entry) but become more obvious eight and ten years after entry. The magnitude of the estimated elasticities is small, being 0.5, 1.7, and 1.9 percent for mean earnings six, eight, and ten years after entry. To put the numbers into perspective, the mean value of instructional spending in my analysis sample is $8,066. The average earnings ten years after entry is $46,700. If spending increases by 10%, that is, by $806, it will lead to a 0.19% increase in earnings ten years after entry, which is $89.

The second observation is that the estimated elasticities are higher at lower percentiles in the earning distribution. This pattern holds for earnings six, eight, and ten years after entry. It suggests that higher instructional spending is overall effective in improving student earnings, and more so for students of lower earnings. However, since the estimates are elasticities, the effects in levels do not necessarily obey this pattern.

Different earning outcomes use slightly different sets of institutions due to data availability. To ensure the differences in results are not driven by the sample variation, Table 1.6 presents the results for the same specification using the same set of institutions where all earning outcome variables are available. The pattern that estimated elasticities are higher at the lower percentiles in the earning distribution remains the same. The point estimates are 1.1, 1.3, and 1.1 percent for earnings six, eight, and ten years after entry. They are not statistically different from previous results using the larger set of institutions. However, they do differ in magnitude and are less dispersed, suggesting that the differences between the estimates for earnings six, eight, and ten years after entry could be caused by differential samples. In Appendix A.2, I analyze the effect heterogeneity across the institution and the time dimensions. I find strong evidence suggesting the effects are different over time and little evidence that there is great heterogeneity across institutions.

1.4.3 Results by Types of Institution

A natural question to ask is if the above-estimated effects differ by the type of institutions. As the data cover more than 5,300 institutions, I have the statistical power to estimate those effects separately. I interact the instructional spending with an indicator variable for public institutions to allow different slopes to be estimated.

The results are presented in Table 1.7. It becomes apparent that the estimated elasticities are different for public and private institutions. The coefficients are 0.6, 2, and 2.3 percent for mean earning six, eight, and ten years after entry for private institutions. For public institutions, they are only 0.3, 0.6, and 0.35 percent and are statistically insignificant. I again visually summarize the coefficients and the confidence intervals for private institutions in Figure 1.6. The overall pattern looks similar between Figure 1.5 and Figure 1.6 that the coefficient is larger at lower earning percentiles.

It warrants further investigation of why instructional spending is effective in private institutions but ineffective in public institutions. I divide the public institutions into more established ones and less established ones. I make the distinction by whether they were observed in 1990. Only a tiny fraction of the institutions that were not observed in 1990 are recently established. As reporting was not mandatory for those not receiving Title IV funds, the majority of them chose not to report to the IPEDS in 1990 and decided to report later on. Table 1.8 presents the results for public institutions by whether they were observed in 1990. Estimated coefficients for institutions that are more established are 1.1, 1.3, and 0.8 percent. There are closer to the estimated coefficients for private institutions and are statistically significant. The fact that less established institutions face a higher marginal cost in the market of college professors may explain this difference. They usually have to offer higher compensation to attract faculty members of equal caliber than more established institutions.

I also interact the instructional spending with an indicator for whether the institution is a four-year institution. Table 1.9 presents the results. The estimated elasticities are 1.1, 3.1, and 3.5 percent for earning six, eight, and ten years after entry for four-year institutions, while they are indistinguishable from zero for two-year-or-less institutions.

1.4.4 Sensitivity for Coefficients in the Main Results

I present the results of the sensitivity analysis discussed in Section 1.3.4. Table 1.10 shows the answer to the first question: how strong a confounding factor has to be so that the estimated coefficients will be brought to zero? The coefficients and standard errors are the same as reported in Table 1.5. For results that are insignificant, they obviously can only survive a very low level of confounding. For coefficients that are significant, most of them can survive a level of confounding around $\delta = 0.5$. This is lower than the ideal threshold of $\delta = 1$ for the following reason. As I am including a rich set of control variables, the within variations explained by the controlled regression ($\tilde{R}$) are above 80%, leading to $R_{max} = 1$ when using $R_{max} = \min(1.3\tilde{R}, 1)$. As discussed in Oster (2019), $R_{max} = 1$ is a strict criteria where only 9% of the studies using nonrandomized data can survive a $\delta = 1$. To back up the above point, my coefficients for earnings ten years after entry can survive a $\delta = 1$ when the within variation explained by the controlled regression is not as high (around 70%) and $R_{max} = \min(1.3\tilde{R}, 1) < 1$.

The second question is: for a given value of δ, what will the estimate of β be after adjusting for that level of confounding? To answer it, Figure 1.7 plots the estimated coefficient against values of δ ranging from 0.1 to 1, for all outcome variables. The plot ends when the next value of δ will change the sign of the estimated β. It happens at the value of the corresponding δ shown in Table 1.10. It is worth noting that, for the elasticity of the mean earnings ten years after entry, controlling for an unobserved confounder that is half as strong as all currently controlled variables will only bring the estimated coefficients down from 1.9 percent to 0.9 percent. From Table 1.5, the standard error is 0.23 percent, indicating that the coefficient will remain positive and likely remain significant. The same holds for all the elasticities of earning outcomes ten years after entry in the distribution.

1.5 Instrumental Variable (IV) Regression

In this section, I use the variable ($Z_{i,c}$) constructed by Deming and Walters (2017) as an instrument variable for the instructional spending. Specifically,

$$Z_{i,c} = \left(\frac{Approp_{i,90}}{TotalRevenue_{i,90}}\right) \cdot \left(\frac{StateApprop_{s(i),c}}{FTEStudent_{s(i),c}}\right) \tag{1.5}$$

where for institution i, $Approp_{i,90}$ is the state appropriation in 1990 and $TotRev_{i,90}$ is the total revenue in 1990. The first factor measures institution i's historical financial dependence on state appropriations. In the second factor, $s(i)$ denotes the state of the institution i. Therefore, the second factor is the state-level average appropriation per FTE enrolled student in year c. As the state-level appropriation only affects public institutions, I run the IV regression using the set of 1,690 public institutions from my analysis sample. Not all 1,690 public institutions were observed in 1990. Nearly one-fourth of them were not.

In the first stage, I estimate the following equation, including the current and one lag of the instrumental variable, along with the entire set of control variables, institution fixed effects, and cohort fixed effects.

$$\ln(InExp_{ic}) = \alpha_1 + \beta_{11}\ln(Z_{i,c}) + \beta_{12}\ln(Z_{i,c-1}) + \boldsymbol{X_{ic}\gamma_1} + \boldsymbol{\kappa_{1i}} + \boldsymbol{\eta_{1c}} + u_{1ic} \tag{1.6}$$

I do not include state-specific time trends, which deviates from my main specification in the two-way fixed effects estimation. I do so because the instrument is largely affected by state-level time trends. The second factor $\left(\frac{StateApprop_{s(i),c}}{FTEStudent_{s(i),c}}\right)$ only varies at the state level. Including state-specific time trends will absorb almost all variations in the instrument. I include both the contemporaneous and one lag of $Z_{i,c}$ to have maximum predictive power in the first stage. The F-statistic for the joint test of the significance of the contemporaneous and one lag of the instrument is 18.6.

I recalculate the two-year or four-year average instructional spending based on the institution level with the predicted single-year instructional spending. Using that, in the second stage, I estimate the following equation

$$\ln(Earning_{ic}) = \alpha_2 + \beta_2 \ln(In\hat{E}xp_{ic}) + \boldsymbol{X_{ic}\gamma_2} + \boldsymbol{\kappa_{2i}} + \boldsymbol{\eta_{2c}} + u_{2ic} \tag{1.7}$$

As the two stages are separately estimated, I cluster-bootstrap the entire process to calculate the standard errors.

Table 1.11 presents the IV regression results. We observe that the estimated coefficients bounce around zero and are all imprecisely estimated, with the standard errors five to ten times as large as those from the two-way fixed effects regressions. This is unsurprising given that we find small to no effect for public institutions in Section 1.4.3. The IV regression results do not further support, nor contradict the two-way fixed effects results. Unfortunately, the IV regression does not provide additional insights.

1.6 Discussion

Although this paper intends to inform policymakers in making decisions on adjusting instructional spending when facing budget shocks, in no way do I suggest that a conclusive decision can be made merely based on the results presented in this paper. I investigated earning outcomes, which is only one of the outputs of the education production function. Potentially, the same set of inputs that promote the earnings margin also improves many other margins. For example, higher spending may increase the possibility of going to graduate school, the possibility of working in a job that brings less disutility,[19] and the possibility of meeting a better spouse. Those dimensions are no less important than monetary returns and await further empirical investigations.

To put my estimated coefficients into perspective, I evaluate them at my analysis sample's mean earnings and instructional spending. A standard deviation increase in instructional spending (\$5,935) from the mean (\$8,066) leads to increases in earnings eight years after entry by \$601 overall (\$687 for private institutions and \$447 for public institutions that are more established). This is comparable to the estimates in Mountjoy and Hickman (2021) where they find a standard deviation increase in a composite index of institution quality increases student earnings eight to ten years after graduation by \$753. Griffith and Rask (2016), however, find a larger elasticity coefficient. According to their estimate, such an increase will lead to an increase in earnings of \$1,333 four years after graduation, with a wide confidence interval covering zero.

[19]In Griffith and Rask (2016), they find a marginally significant positive association between higher instructional spending and the probability that one works a job matching the field of study in college.

Though I find instructional spending effective in promoting student earnings, is it still effective considering the cost? To answer this question, I plot the net present return of an investment in increasing instructional spending from the sample mean for four years in a four-year institution. I consider three different discount factors, representing different levels that a person's values a dollar one year from now. The first is based on the nominal interest rate in Oct 2021, which is 0.25% so that the discount factor is 0.9975. The second is based on the nominal interest rate in Sep 2019 to be free from the drastic changes in economic conditions due to the pandemic, and the discount factor is 0.9825.[20] The third is based on the real interest rate in Sep 2019, and the discount factor is 1.0006.[21] The first and the third choices of the discount factors reflects a highly forward-looking perspective due to the recent economic conditions in the United States. The third choice corresponds to a view that values a dollar in the future more than a dollar now. The cumulative return assumes the estimated elasticities (in Table 1.5, Table 1.7, and Table 1.8) for earnings eight years after first attendance are persistent throughout students lifetime.[22]

Figure 1.9 plots the net present return as the cumulative net present value divided by total investment. The total investment is an increase in instructional spending every year for four years.[23] Using the discount factor 0.9975, a student has to work for 42 years (37 years for private institution and 58 for public institution attendees) until the net present return becomes non-negative. Using the less forward-looking discount factor 0.9825, a private institution attendee has to work for 58 years (or to the age of 80, assuming

[20]The monthly average of the 1-year treasury bill secondary market rate is 1.75%, retrieved from the Federal Reserve Bank.

[21]Annual inflation rate is 1.81% in 2019, retrieved from the Bureau of Labor Statistics.

[22]The combined evidence in the literature supports the assumption that the elasticities can be persistent.

[23]I consider four-year institutions only because in Section 1.4.3, only four-year institutions were found to have positive effects.

graduating at 22) until the net present return becomes positive. Using the third discount factor 1.0006, where the student values the future more than the current, she has to work for 39 years (35 years for private and 53 years for public institution attendees) until the investment in instructional spending start to pay off.

Alternatively, I evaluate for what value the discount factor has to take so that the net present return is non-negative after working for a fixed number of years. Figure 1.10 plots such relationships for overall, private, and public institutions assuming a student works for 35 to 55 years. First, public institutions require an extremely forward-looking perspective, given their low estimated return. In other words, unless one view future dollars more than current dollars, investment in instructional spending cannot pay off. For example, if one works for 35 years, a discount factor of 1.02 is required, which is unrealistic. For private institutions, up to a small degree of discounting, investment in instructional spending can still pay off. If one works for 45 years, a discount factor of 0.9889 will break even.

The results above may justify that institutions have been shifting away from allocating additional funds to instructional spending. However, the budget decision process surely does not solely rely on the perceived effectiveness of instructional spending but is oriented by the values and goals of the different institution. Future research is warranted to evaluate how institutions make their budget decisions.

Table 1.1: Summary Statistics - Outcome Variables

	(1) Full Sample			(2) Analysis Sample		
	Mean	SD	N	Mean	SD	N
Earning, six years after entry						
Mean	36204	9393	40122	36626	9086	32741
25th percentile	19494	6642	35834	19880	6555	29545
Median	32922	8707	40122	33373	8549	32741
75th percentile	47769	11254	35834	48246	10885	29545
Earning, eight years after entry						
Mean	41459	11562	33803	41980	11175	27535
25th percentile	22385	7788	30240	22859	7701	24866
Median	37248	9997	33803	37782	9830	27535
75th percentile	54199	13559	30240	54771	13094	24866
Earning, ten years after entry						
Mean	46090	13691	27872	46700	13251	22669
25th percentile	24517	8675	24957	25046	8580	20527
Median	40799	11232	27872	41401	11062	22669
75th percentile	60032	16027	24957	60706	15490	20527

Notes: Data from College Scorecard project. The full sample consists of observations where instructional spending and at least one of the earning outcome variables are nonmissing. Analysis sample removes from the full sample (1) institutions that have changed their levels (two-year vs. four-year), and (2) institutions that had higher than $100,000 or lower than $50 instructional spending per student. Mean and standard deviation are weighted by the number of undergraduate enrollments. All monetary terms are in 2014 dollars.

Table 1.2: Summary Statistics - Treatment and Control Variables

	(1) Full Sample			(2) Analysis Sample		
	Mean	SD	N	Mean	SD	N
Instructional spending per student	9722	129516	50263	8066	5935	40962
Average SAT	1077	119	11311	1076	115	10668
Family income	38938	19908	50231	40244	19957	40939
Median household income	82071	17743	24716	82667	17681	20102
Age at entry	24.0	2.9	50231	23.9	2.9	40939
White	74.4	15.1	24716	74.6	15.1	20102
Black	12.8	11.8	24716	12.7	11.8	20102
Asian	3.9	4.8	24716	3.9	4.8	20102
Hispanic	13.9	17.4	24716	13.4	16.6	20102
Bachelor	15.6	4.0	24613	15.7	4.0	20028
Graduate or professional	8.7	2.9	24613	8.8	2.9	20028
Born in U.S.	87.4	10.5	24613	87.7	10.2	20028
Unemployment rate	3.77	1.08	24716	3.74	1.05	20102
Poverty rate	10.02	6.00	24716	9.85	5.77	20102
Female	0.60	0.11	45688	0.60	0.10	37290
Married	0.15	0.10	47865	0.15	0.10	38888
Dependent	0.60	0.22	49347	0.62	0.22	40247
Veteran	0.03	0.03	33178	0.03	0.03	27221
First generation	0.43	0.12	47374	0.42	0.12	38703

Notes: Data from College Scorecard project. See notes in Table 1.1 for sample restrictions. Mean and standard deviation are weighted by the number of undergraduate enrollments. White refers to the percent of the population from students' zip codes who are White, via Census data. The same holds for Black, Asian, Hispanic, Bachelor, Graduate or professional, Born in the U.S., Unemployment rate, and Poverty rate. All monetary terms are in 2014 dollars.

Table 1.3: Correlation of Different Educational Inputs

	SAT	InExp	Retention
SAT	1		
InExp	0.611	1	
Retention	0.707	0.448	1

	Res_InExp	Res_Retention
Res_InExp	1	
Res_Retention	0.0287	1

	Dm_InExp	Dm_SAT	Dm_Retention
Dm_InExp	1		
Dm_SAT	0.185	1	
Dm_Retention	0.0453	0.112	1

Notes: Data from College Scorecard project. The top panel shows the correlation between the raw values of the average SAT score, instructional spending per FTE student, and first-year retention rate. In the mid panel, I correlate residuals of instructional spending and retention rate from two separate regressions on the SAT score, respectively. In the bottom panel, I correlate the demeaned value of the three variables after subtracting from them the institution average.

Table 1.4: Effect of Instructional Spending on Earnings, Pooled OLS Regression

	(1)	(2)	(3)	(4)
Log mean earning six years after entry				
ln(InExp/FTE)	0.0270	0.0186	0.00545	0.00535
	(0.0053)	(0.0035)	(0.0030)	(0.0027)
Number of Institutions	4721	4721	4721	4721
Log mean earning eight years after entry				
ln(InExp/FTE)	0.0351	0.0271	0.0173	0.0175
	(0.0049)	(0.0038)	(0.0030)	(0.0028)
Number of Institutions	4323	4323	4323	4323
Log mean earning ten years after entry				
ln(InExp/FTE)	0.0403	0.0276	0.0192	0.0194
	(0.0052)	(0.0037)	(0.0023)	(0.0023)
Number of Institutions	3998	3998	3998	3998
Other Controls	Yes	Yes	Yes	Yes
Average SAT	No	Yes	Yes	Yes
Fixed Effects	No	No	Yes	Yes
State-Specific Time Trends	No	No	No	Yes

Notes: Data from College Scorecard project for cohorts between 1996 and 2007. See notes in Table 1.1 for sample restrictions. Control variables include average SAT score, average family income, median household income, the average age at entry for the entering cohort, racial composition, educational level, unemployment rate and the poverty rate of the place of origin of the entering students, the fraction of the enrolled students that are female, married, dependent, veteran, and first-generation college student, and composition of degrees conferred by Classification of Instructional Programs (CIP) codes. Regressions are weighted by the average cohort size in each institution. Standard errors in parentheses are calculated using cluster-robust standard errors at the institution level.

Table 1.5: Effect of Instructional Spending on Earnings

	(1) Mean	(2) Pct25	(3) Pct50	(4) Pct75
Log earnings six years after entry				
ln(InExp/FTE)	0.0053	0.014	0.0057	0.0030
	(0.0027)	(0.0044)	(0.0029)	(0.0026)
Number of Institutions	4721	4721	4721	4721
Log earnings eight years after entry				
ln(InExp/FTE)	0.017	0.025	0.019	0.015
	(0.0028)	(0.0039)	(0.0029)	(0.0027)
Number of Institutions	4323	4323	4323	4323
Log earnings ten years after entry				
ln(InExp/FTE)	0.019	0.031	0.023	0.017
	(0.0023)	(0.0034)	(0.0023)	(0.0021)
Number of Institutions	3998	3998	3998	3998

Notes: Data from College Scorecard project for cohorts between 1996 and 2007. See notes in Table 1.1 for sample restrictions. All regressions include cohort fixed effects, institution fixed effects, state-specific time trends, and a set of control variables. Control variables include average SAT score, average family income, median household income, the average age at entry for the entering cohort, racial composition, educational level, unemployment rate and the poverty rate of the place of origin of the entering students, the fraction of the enrolled students that are female, married, dependent, veteran, and is first-generation college student, and composition of degrees conferred by Classification of Instructional Programs (CIP) codes. Regressions are weighted by the average cohort size in each institution. Standard errors in parentheses are calculated using cluster-robust standard errors at the institution level.

Table 1.6: Effect of Instructional Spending on Earnings, Same Institutions

	(1) Mean	(2) Pct25	(3) Pct50	(4) Pct75
Log earnings six years after entry				
ln(InExp/FTE)	0.011	0.017	0.011	0.0087
	(0.0024)	(0.0036)	(0.0024)	(0.0023)
Log earnings eight years after entry				
ln(InExp/FTE)	0.013	0.022	0.014	0.010
	(0.0023)	(0.0030)	(0.0023)	(0.0023)
Log earnings ten years after entry				
ln(InExp/FTE)	0.011	0.019	0.014	0.0092
	(0.0020)	(0.0029)	(0.0020)	(0.0019)
Number of Institutions	3826	3826	3826	3826

Notes: Data from College Scorecard project for cohorts between 1996 and 2007. See notes in Table 1.1 for sample restrictions. All regressions include cohort fixed effects, institution fixed effects, state-specific time trends, and a set of control variables. Control variables include average SAT score, average family income, median household income, the average age at entry for the entering cohort, racial composition, educational level, unemployment rate and the poverty rate of the place of origin of the entering students, the fraction of the enrolled students that are female, married, dependent, veteran, and is first-generation college student, and composition of degrees conferred by Classification of Instructional Programs (CIP) codes. Regressions are weighted by the average cohort size in each institution. Standard errors in parentheses are calculated using cluster-robust standard errors at the institution level.

Table 1.7: Effect of Instructional Spending on Earnings, by Public and Private Institutions

Treatment: Ln(InExp/FTE)	(1) Mean	(2) Pct25	(3) Pct50	(4) Pct75
Log earnings six years after entry				
Private	0.0059	0.016	0.0056	0.0030
	(0.0032)	(0.0051)	(0.0033)	(0.0030)
Public	0.0030	0.0068	0.0065	0.0034
	(0.0044)	(0.0066)	(0.0048)	(0.0045)
Number of Institutions	4721	4721	4721	4721
Log earnings eight years after entry				
Private	0.020	0.029	0.021	0.017
	(0.0032)	(0.0045)	(0.0033)	(0.0030)
Public	0.0060	0.0068	0.0079	0.0047
	(0.0053)	(0.0065)	(0.0055)	(0.0051)
Number of Institutions	4323	4323	4323	4323
Log earnings ten years after entry				
Private	0.023	0.037	0.027	0.020
	(0.0025)	(0.0039)	(0.0026)	(0.0024)
Public	0.0035	0.0048	0.0036	0.0012
	(0.0040)	(0.0050)	(0.0036)	(0.0037)
Number of Institutions	3998	3998	3998	3998

Notes: Data from College Scorecard project for cohorts between 1996 and 2007. See notes in Table 1.1 for sample restrictions. All regressions include cohort fixed effects, institution fixed effects, state-specific time trends, and a set of control variables. See notes in Table 1.4 for descriptions of the control variables. The treatment variable is instructional spending per FTE student. Instructional spending is interacted with the type of institution. Outcome variables are mean, 25th percentile, median, and 75th percentile of earnings six, eight, and ten years after entry. Regressions are weighted by the average cohort size in each institution. Standard errors in parentheses are calculated using cluster-robust standard errors at the institution level.

Table 1.8: Effect of Instructional Spending on Earnings, Public Institutions

Treatment: Ln(InExp/FTE)	(1) Mean	(2) Pct25	(3) Pct50	(4) Pct75
Log earning six years after entry (No. of Institutions: 1686)				
Not Observed in 1990	-0.0063	-0.0016	-0.0059	-0.0087
	(0.0050)	(0.0064)	(0.0054)	(0.0051)
Observed in 1990	0.011	0.012	0.014	0.012
	(0.0037)	(0.0053)	(0.0040)	(0.0037)
Log earning eight years after entry (No. of Institutions: 1648)				
Not Observed in 1990	-0.0012	0.0037	0.00077	-0.0023
	(0.0055)	(0.0075)	(0.0059)	(0.0045)
Observed in 1990	0.013	0.012	0.014	0.012
	(0.0050)	(0.0066)	(0.0055)	(0.0051)
Log earning ten years after entry (No. of Institutions: 1609)				
Not Observed in 1990	-0.0021	0.011	0.0042	-0.0043
	(0.0064)	(0.0084)	(0.0061)	(0.0058)
Observed in 1990	0.0079	0.0094	0.0058	0.0048
	(0.0034)	(0.0044)	(0.0031)	(0.0032)

Notes: Data from College Scorecard project for cohorts between 1996 and 2007. See notes in Table 1.1 for sample restrictions. All regressions include cohort fixed effects, institution fixed effects, state-specific time trends, and a set of control variables. See notes in Table 1.4 for descriptions of the control variables. The treatment variable is instructional spending per FTE student. Instructional spending is interacted with whether the institution was observed in 1990. Outcome variables are mean, 25th percentile, median, and 75th percentile of earnings six, eight, and ten years after entry. Regressions are weighted by the average cohort size in each institution. Standard errors in parentheses are calculated using cluster-robust standard errors at the institution level.

Table 1.9: Effect of Instructional Spending on Earnings, by Institution Levels

Treatment: Ln(InExp/FTE)	(1) Mean	(2) Pct25	(3) Pct50	(4) Pct75
Log earning six years after entry (No. of Institutions: 4721)				
Two-year or less	-0.0029	-0.0045	-0.0024	-0.00081
	(0.0027)	(0.0044)	(0.0029)	(0.0025)
Four-year	0.011	0.028	0.012	0.0059
	(0.0041)	(0.0068)	(0.0044)	(0.0040)
Log earning eight years after entry (No. of Institutions: 4323)				
Two-year or less	0.0024	0.0024	0.0034	0.0036
	(0.0027)	(0.0038)	(0.0028)	(0.0026)
Four-year	0.031	0.047	0.034	0.025
	(0.0046)	(0.0067)	(0.0048)	(0.0044)
Log earning ten years after entry (No. of Institutions: 3998)				
Two-year or less	0.0052	0.0048	0.0065	0.0054
	(0.0022)	(0.0034)	(0.0023)	(0.0021)
Four-year	0.035	0.059	0.040	0.029
	(0.0036)	(0.0049)	(0.0037)	(0.0034)

Notes: Data from College Scorecard project for cohorts between 1996 and 2007. See notes in Table 1.1 for sample restrictions. All regressions include cohort fixed effects, institution fixed effects, state-specific time trends, and a set of control variables. See notes in Table 1.4 for descriptions of the control variables. The treatment variable is instructional spending per FTE student. Instructional spending is interacted with whether the institution is a four-year institution. Outcome variables are mean, 25th percentile, median, and 75th percentile of earnings six, eight, and ten years after entry. Regressions are weighted by the average cohort size in each institution. Standard errors in parentheses are calculated using cluster-robust standard errors at the institution level.

Table 1.10: Sensitivity: What value of δ will drive β to 0?

	(1) Mean	(2) Pct25	(3) Pct50	(4) Pct75
Log earnings six years after entry				
ln(InExp/FTE)	0.00535	0.0141	0.00573	0.00303
	(0.0027)	(0.0044)	(0.0029)	(0.0026)
$\tilde{R}$	0.900	0.903	0.917	0.897
$R_{max} = \min(1.3 \cdot \tilde{R}, 1)$	1	1	1	1
δ	0.302	0.508	0.343	0.183
Log earnings eight years after entry				
ln(InExp/FTE)	0.0175	0.0253	0.0191	0.0149
	(0.0028)	(0.0039)	(0.0029)	(0.0027)
$\tilde{R}$	0.806	0.808	0.838	0.801
$R_{max} = \min(1.3 \cdot \tilde{R}, 1)$	1	1	1	1
δ	0.486	0.483	0.561	0.431
Log earnings ten years after entry				
ln(InExp/FTE)	0.0194	0.0310	0.0225	0.0168
	(0.0023)	(0.0034)	(0.0023)	(0.0021)
$\tilde{R}$	0.686	0.662	0.753	0.687
$R_{max} = \min(1.3 \cdot \tilde{R}, 1)$	0.892	0.860	0.980	0.893
δ	0.974	1.139	0.893	0.822

Notes: This table applies the sensitivity analysis described in Oster (2019) to the main estimation equation. The coefficients and the standard errors are the same as in Table 1.5. $\tilde{R}$ is the R-squared for the controlled regression. R_{max} is conceptually the maximum variation in the outcome variable that could be explained by including all possible confounding variables. Using $R_{max} = \min(1.3 \cdot \tilde{R}, 1)$ was suggested by Oster (2019) based on a review of existing published articles. δ is the degree of selection on potential confounding variables relative to the degree of selection on observed variables that are already included in the equation. Larger δ indicates more robust estimated coefficient.

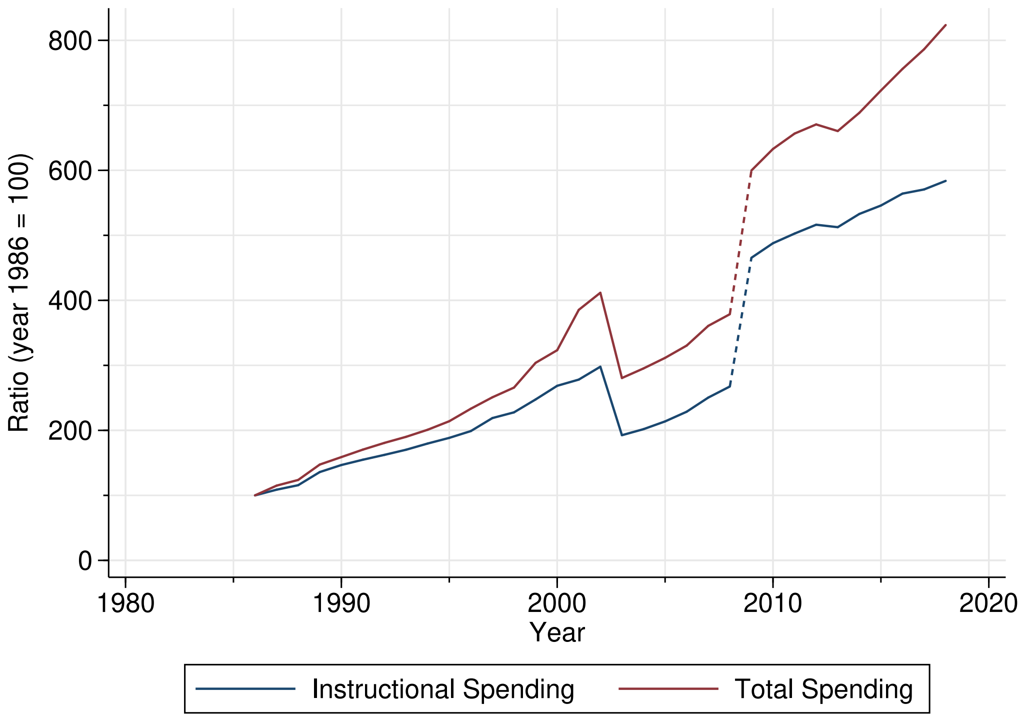
.

Table 1.11: IV Regression: Effect of Instructional Spending on Earnings, Public Institutions

	(1) Mean	(2) Pct25	(3) Pct50	(4) Pct75
Log earnings six years after entry				
Ln(InExp/FTE)	0.016	-0.041	0.006	0.007
	[.026]	[.044]	[.029]	[.026]
Number of Institutions	1266	1266	1266	1266
Log earnings eight years after entry				
Ln(InExp/FTE)	0.007	-0.066	-0.021	0.008
	[.02]	[.038]	[.025]	[.019]
Number of Institutions	1260	1260	1260	1260
Log earnings ten years after entry				
Ln(InExp/FTE)	-0.009	-0.047	-0.010	0.002
	[.021]	[.03]	[.021]	[.024]
Number of Institutions	1252	1252	1252	1252

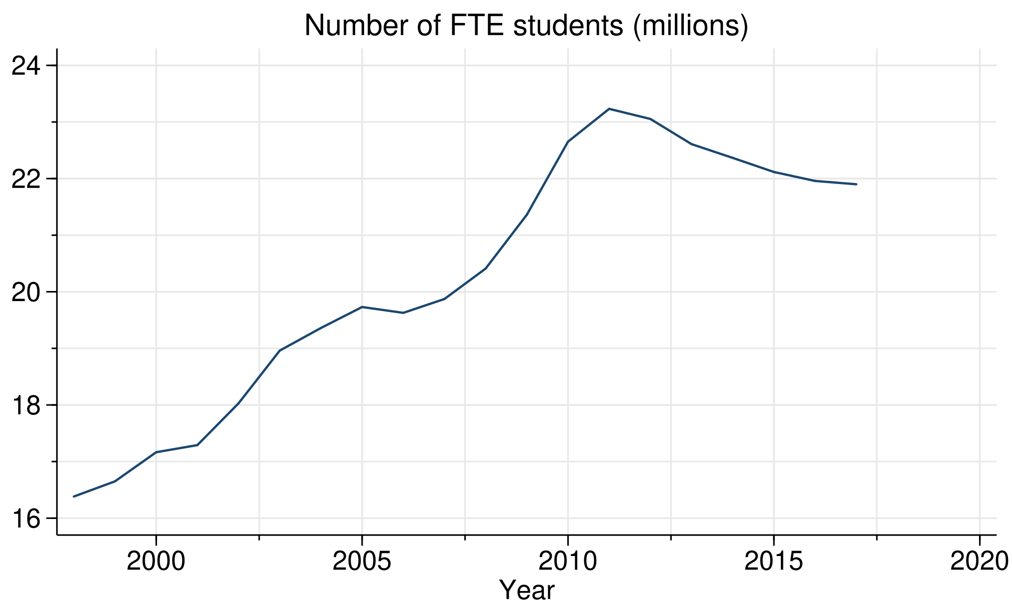
Notes: This table presents the IV regression results. The instrument is the log of the institution's financial dependence on state appropriations in 1990 multiplied by average state appropriations per student, and its one-lag value. See Section 1.5 for more detailed discussion. Data from the College Scorecard project and State Higher Education Finance project for cohorts between 1996 and 2007. The analysis sample consists of institutions that have remained consistent in their level and have reported instructional spending per student between $50 to $100,000. All regressions include a set of control variables, institution fixed effects, and cohort fixed effects. See notes in Table 1.4 for descriptions of the control variables. The first-stage F-statistic in 18.6. Standard errors in brackets are calculated based on 100 replications of cluster-bootstrap of the entire process at the institution level.

Figure 1.1: Growth of Instructional Spending and Total Spending



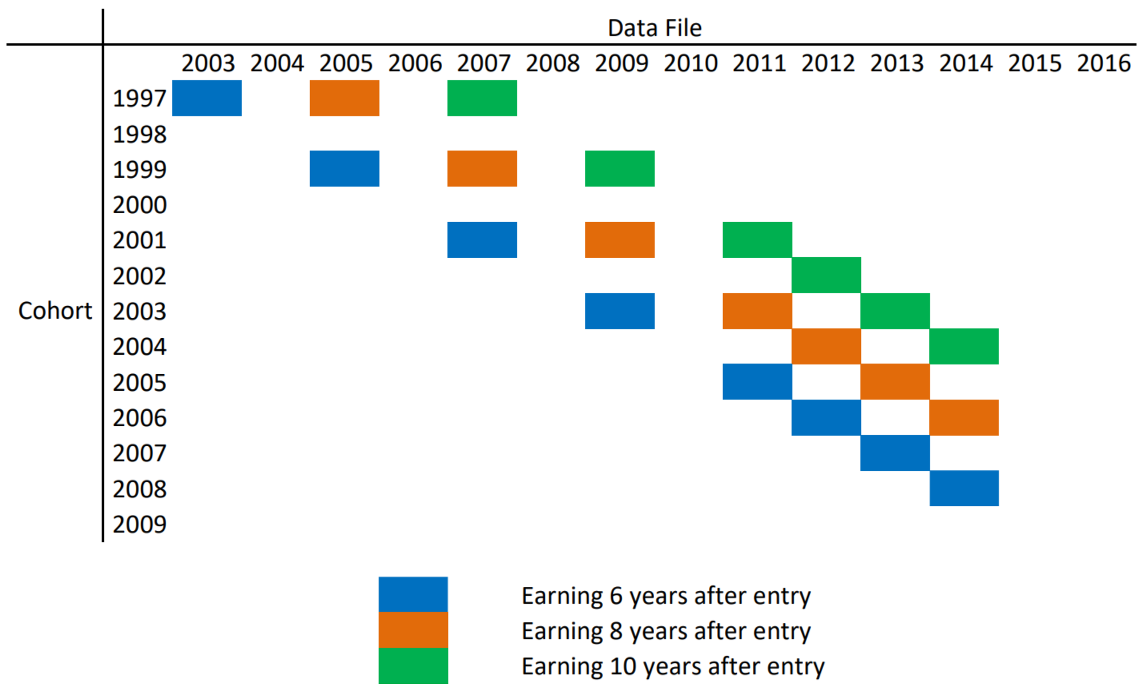
Notes: Author's calculation based on data from the Integrated Postsecondary Education Data System (IPEDS). The dashed line represents an institutional change in the data collection process. Comparisons should not be made before and after the change.

Figure 1.2: Change in Instructional Spending and Change in Wage Earnings of Young Workers with College Degree



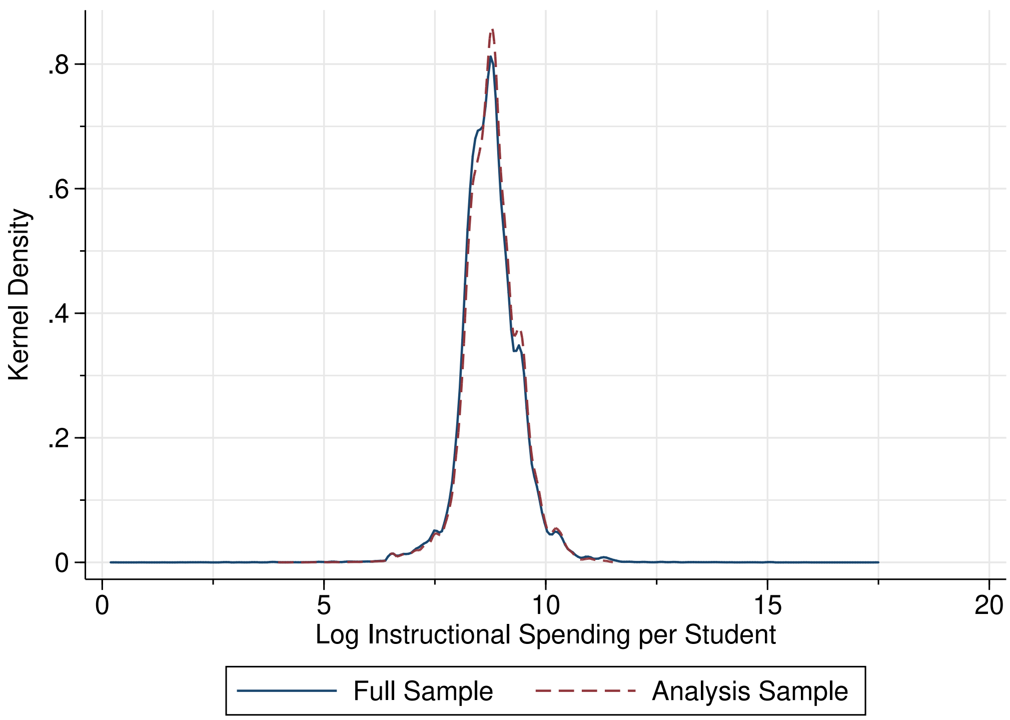
Notes: Each point represent a state. States are grouped into four regions and are indicated by different colors. The horizontal axis shows the percentage change in spending per FTE student between 2009 and 2016. The vertical axis shows the percentage change in wage earnings for young workers (age 22-30) with at least an associate degree between 2012 and 2019. The dashed line is a linear fit weighted by the population of young workers with at least an associate degree in 2019. Calculations are based on data from Integrated Postsecondary Education Data System (IPEDS) and American Community Survey (ACS). All monetary terms are deflated using Consumer Price Index from the Bureau of Labor Statistics.

Figure 1.3: Data Structure and Availability for Outcome Variables



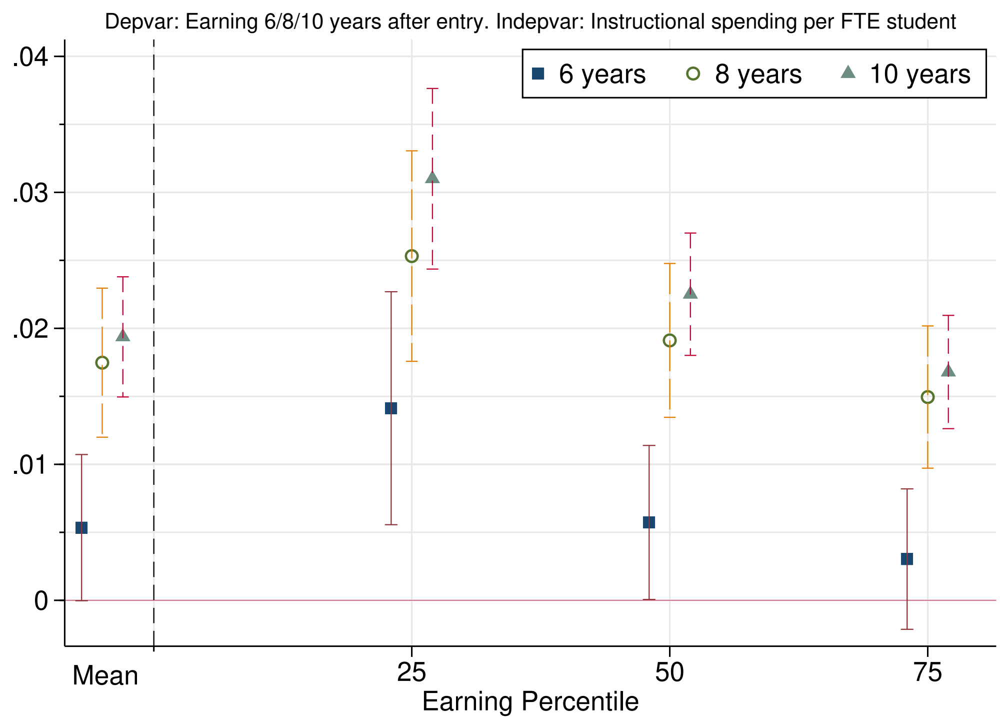
Notes: Colored area indicates that outcome variable is available in that year. Cohorts are pooled at the two-year level. Cohort 1997 includes all students entering the institution in Fall 1996 and Fall 1997.

Figure 1.4: Distribution of the Instructional Spending



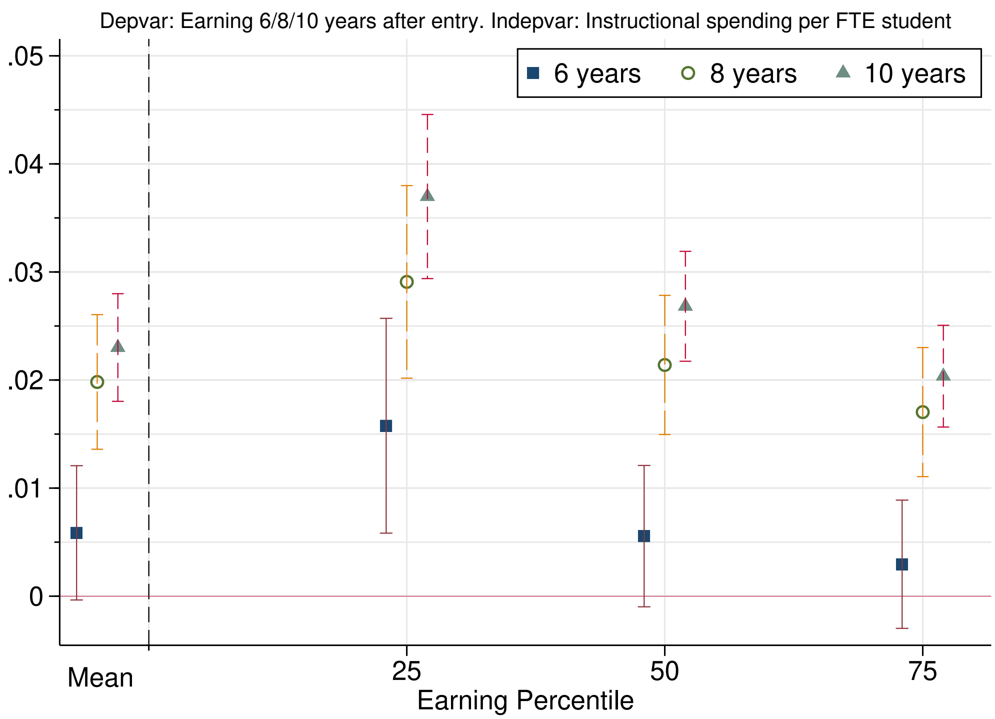
Notes: The full sample consists of 51,779 institution-cohort observations. The analysis sample consists of 42,244 observations, which are institutions that have remained consistent in their levels, and have not reported to have instructional spending per student falling outside the range of $50 to $100,000.

Figure 1.5: Effect of Instructional Spending per FTE Student on Earnings (All Institutions)



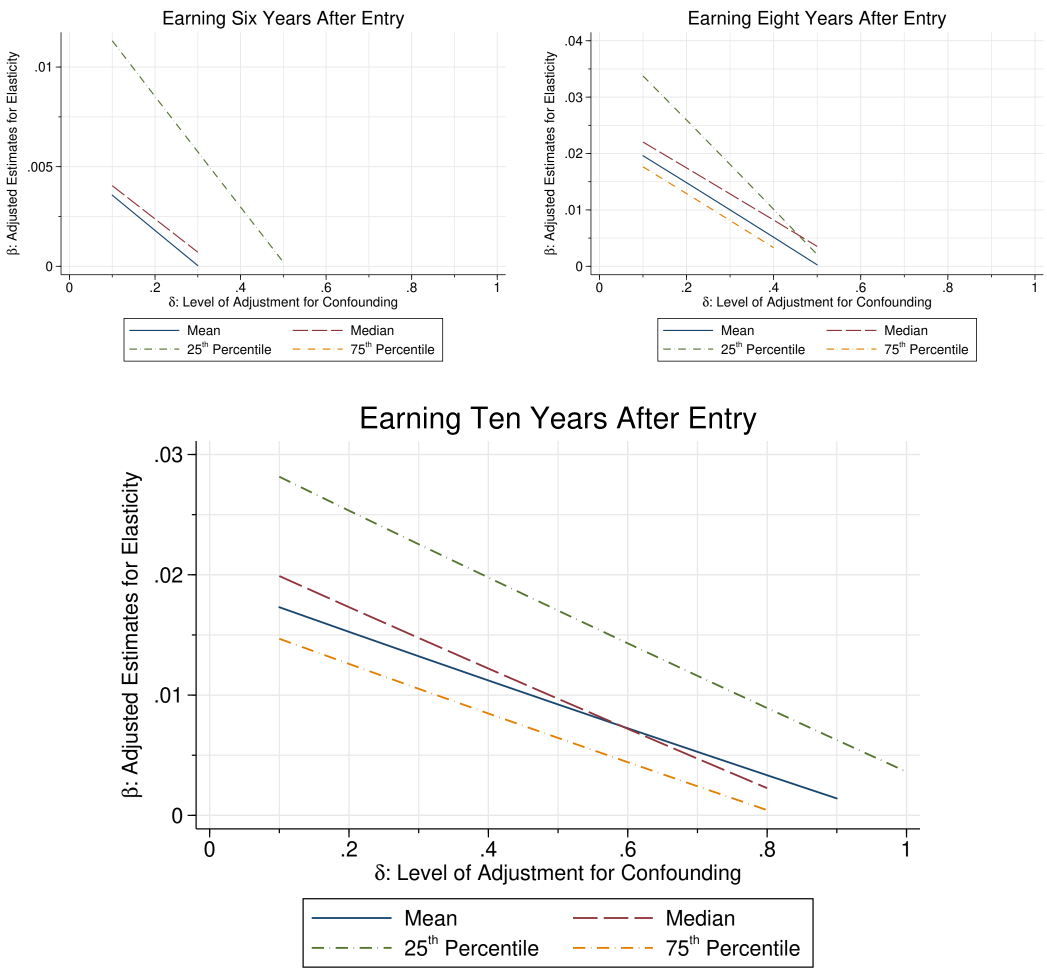
Notes: This figure visually summarizes estimation results of the effect of instructional spending per student on average student earnings. 95% confidence intervals are attached to the point estimates. See Table 1.5 for the exact values.

Figure 1.6: Effect of Instructional Spending per FTE Student on Earnings (Private Institutions)



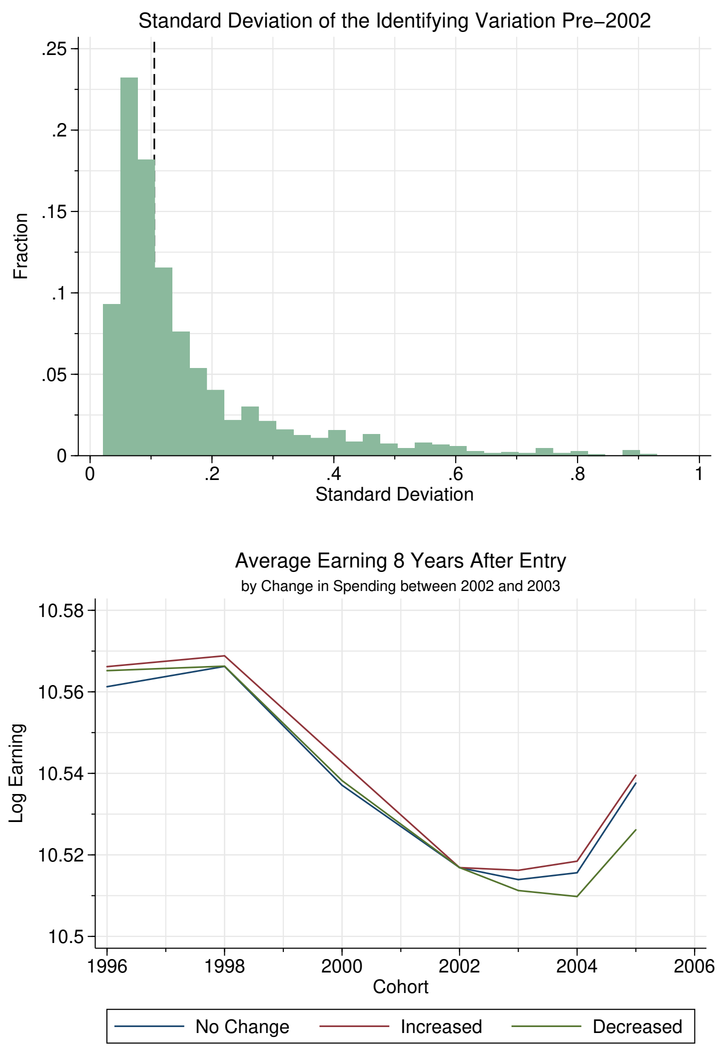
Notes: This figure visually summarizes estimation results of the effect of instructional spending per student on average student earnings for private institutions. 95% confidence intervals are attached to the point estimates. See Table 1.7 for the exact values.

Figure 1.7: Confounding Adjusted Estimated Elasticities (All Institutions)



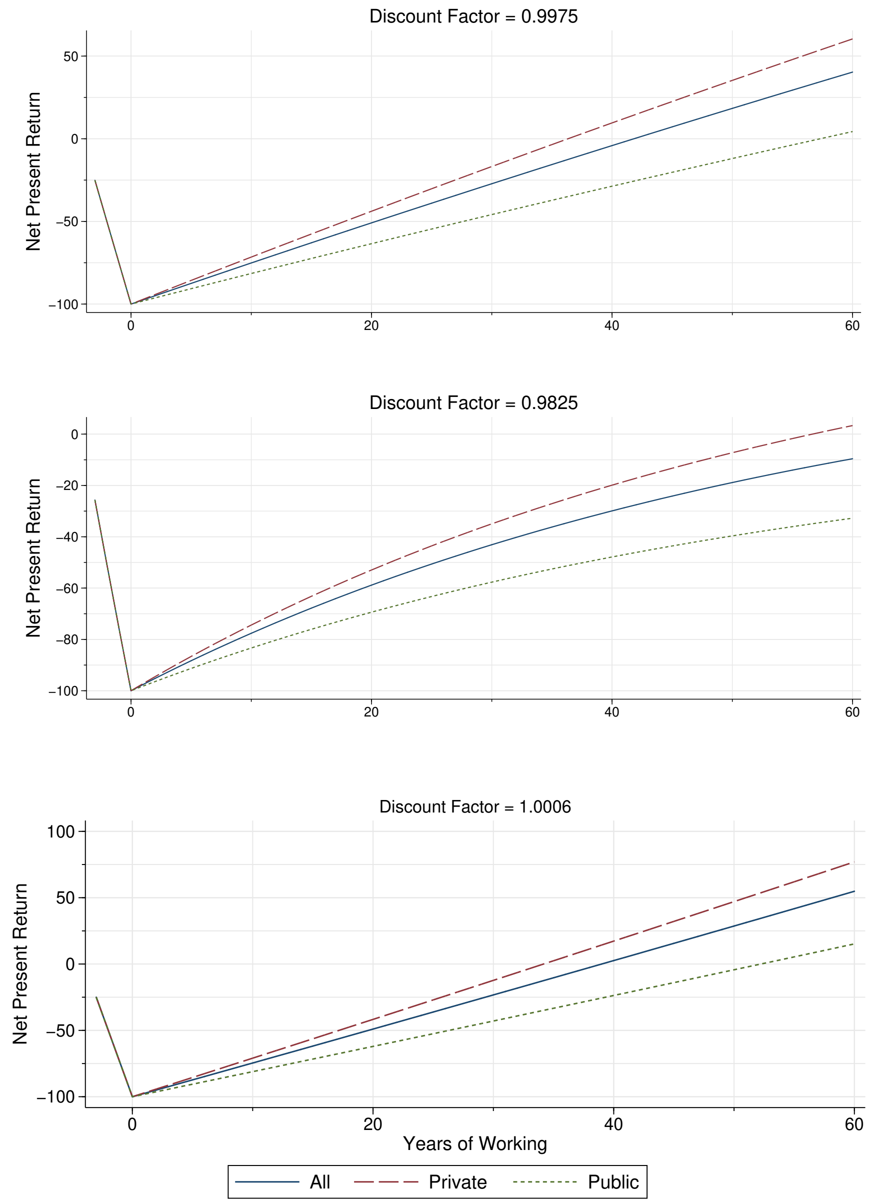
Notes: This figure plots the estimated coefficients in Table 1.5 against levels of adjustment for confounding factors with the method discussed in Oster (2019). The parameter $R_{\max}$ is chosen as $\min(1.3\tilde{R}, 1)$ where $\tilde{R}$ is the R-squared in the controlled regression, as suggested in the original paper. Remaining positive for a high level of confounding adjustment indicates a more robust estimated coefficient.

Figure 1.8: Parallel Trends Assumption



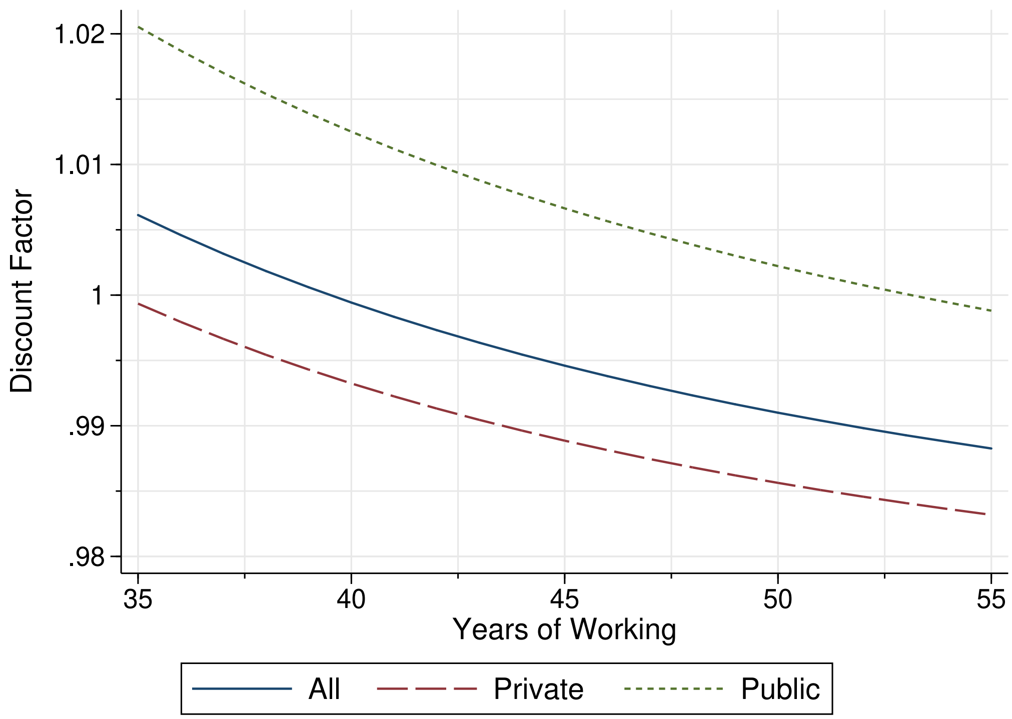
Notes: The figure in the top panel plots the histogram of the standard deviations of the identifying variation in the artificial "pre-period" (before 2002). The identifying variation is the predicted residuals from a regression of instructional spending on all the control variables, cohort fixed effects, institution fixed effects, and state-specific time trends. The figure in the bottom panel plots the average earnings eight years after first attendance for the above three groups over time.

Figure 1.9: Net Present Return of Instructional Spending



Notes: This figure plots the cumulative return of instructional spending as a percentage of total investment in instructional spending at a four-year college.

Figure 1.10: Discount Factor for the Investment in Instructional Spending to Break Even



Notes: This figure plots the value of the discount factor so that the net present return of the investment in instructional spending after working for a fixed number of years will break even. The total investment is an increase in instructional spending each year for four years. The cumulative return assumes the estimated returns (in Table 1.5, Table 1.7, and Table 1.8) for earnings eight years after first attendance are persistent throughout students' lifetime.

Chapter 2

Competition for Better Students and College Budget Decisions

2.1 Introduction

Going to college is a mixture of many experiences, including investing in human capital and consuming amenities. The budget allocation decisions in college determine the provision of educational inputs, consumption amenities, and many other services that directly affect the college experience and, in turn, impact student outcomes. Over the past decades, the composition of college budget spending underwent notable changes. Nationally, instructional spending as a share of total spending decreased from 41% in 1986 to 29% in 2018. The percentage of spending on research has also declined, even for four-year institutions. Over the same period, the three spending categories with the most notable increases are hospitals, student services, and others. Moreover, there exist considerable variation in budget spending decisions across institutions, even among institutions of the same type. Therefore, a key question is to understand how different factors contribute to the decision-making process regarding budget allocations.

This paper focuses on evaluating whether and to what extent colleges make their budget decisions to compete for good students. We construct a model for the college admission and human capital production process. Based on the idea that improving reputation is the most common goal for colleges, and the ability to promote good student outcomes is the primary reason for a college to be perceived as outstanding, we assume colleges aim to maximize the average earnings of their students. This objective differs from other papers in the literature. Many assumed that the colleges' objective was to maximize enrollment, peer quality at admission, total revenue (or profit), or a function of those. We argue that while those measures are commonly used as proxies for a college's reputation, they are the results rather than the cause. The outcomes of the graduating cohorts contribute most to the movements in the way the general public perceives a college, or say its reputation.[1]

Using student earnings as the objective function creates an interesting trade-off for colleges. On the one hand, by increasing the share of spending on educational inputs, a college can better create value-added to the students through the human capital production process. On the other hand, a higher level of consumption for amenities makes a college generally more attractive, and it can more likely attract students with higher baseline ability levels. Jacob et al. (2018) finds that most students value the consumption of amenities, while the taste for academic quality is confined to students of higher ability. It is, therefore,

[1] MacLeod et al. (2017) finds that employers are willing to pay a wage premium for a higher college reputation. Higher reputation is also correlated with the earning growth path, suggesting that reputation matters beyond the signaling of individual skills.

essential to model the different student types with heterogeneous tastes for educational inputs and consumption of amenities.[2]

While we do not estimate the structural parameters of our model, we specify them based on existing knowledge in the literature and match moments in the data. We simulate the case with two competing colleges of similar quality to illustrate whether and how a college responds to budget changes of the competing college. We find that when the competing college is spending a relatively high share of its budget on educational inputs, the best choice of the college is to undercut its competitor. This reflects the idea that by slightly decreasing the spending on educational inputs and increasing the spending on consumption amenities, a college can become better off by attracting a student body of higher quality. Over this range, our model predicts a positive relationship between the budget decisions among two competing colleges.

When the competing college spends a relatively low share of its budget on educational inputs, and as it further decreases its share, the best response of the college is to spend increasingly more on educational inputs. The reason is that when the competitor is spending a small share on educational inputs, they are particularly unattractive to students with high learning abilities. They are also generally unattractive to an average student as students value their earnings after graduation, and a low level of spending on educational inputs does not promote that margin as much. Hence the competitor imposes a meager amount of competitive pressure. In such cases, the college does not need to spend as much on consumption amenities to attract good students. They will spend increasingly more on

[2] We acknowledge that the spending on educational inputs and consumption amenities is not necessarily mutually exclusive. For example, events that promote social networks are both investment and consumption. Practically, there is a continuum of trade-offs. To fix the idea and illustrate our main point, we take the dichotomous approach in categorizing spending.

educational inputs to better promote the outcomes of the students they admit. Our model thus predicts a negative relationship over this range. Overall, the best response function is U-shaped between two colleges of similar characteristics.

We simulated a second case of the competition between two colleges with lower competitive pressure. Intuitively, the responsiveness declines as the competitive pressure become lower. Our simulated result confirms that intuition. As two colleges become more heterogeneous, the region where one institution is unresponsive to its competitor expands until a certain point where the best response functions become flat lines.

We verify the above findings with our empirical analysis, using data from the Integrated Postsecondary Education Data System (IPEDS). We find that most colleges respond positively to increases in spending on educational inputs by their competitors, which is robust across several ways in defining the competing institution. Overall, a 10 percentage points increase in the average share of spending on educational inputs by competitors is associated with a 4 to 7 percentage points response. Associate's colleges, doctoral universities, master's colleges, and universities, special focus four-year colleges, and others are more responsive than special focus two-year colleges and baccalaureate colleges.

Our empirical evidence confirms the upward sloping part of the U-shaped best response function. We estimate a different slope parameter over five intervals for each type of institution. Only special focus two-year colleges have a negative slope at the bottom end of the unit interval and positive slopes over the remaining unit interval, consistent with the U-shaped predictions. All other types of colleges have positive slopes throughout the unit interval.

We also estimate the responsiveness in the log level of spending on educational inputs. As is expected, we find positive associations between a college's choice and its competitors, with elasticity parameters estimated to be between 5 and 18 percent. We consider another source of the decline of competitive pressure - increasing distances between colleges. We find that such association fades away as we expand the definition of competing colleges by including colleges further away in our calculation of average spending, except for doctoral universities. As we calculate the average level of spending on educational inputs among colleges within a radius of 1,000 miles, the association found above completely disappears.

This paper contributes to the growing literature analyzing education institutions as markets. While extensive studies have analyzed the impact of educational spending decisions on student outcomes (Hanushek, 1981; Grogger, 1996; Papke, 2005; Webber and Ehrenberg, 2010; Jackson et al., 2016; Hyman, 2017; Lafortune et al., 2018), less is known about the determinants of those decisions, particularly for colleges. Robins (1973) provide some institutional knowledge. Rothschild and White (1993, 1995) provide insights into the determinants of pricing strategies of higher education institutions. This paper takes the angle of competition for good students between colleges and studies how that can affect budget allocation decisions.[3]

Jacob et al. (2018) is the closest study to this paper that estimates the demand elasticities for academic quality and consumption amenities using net revenue as the objective function of colleges. The types of students that colleges compete for in their setting

[3]We abstract away from the joint decision-making process on the enrollment and revenue maximization margins.

are wealthy students willing to pay a higher tuition price. In our setup, colleges compete for students with higher baseline levels of human capital or higher learning capabilities.

This paper is also related to another strand of the literature that studies the competition between educational institutions. Hoxby (2000) finds that competition improves production efficiency in public schools and improves student outcomes. Belfield and Levin (2002) make a good review of the early literature. We approach our question by considering the competition by providing better consumption amenities. Epple et al. (2006) provides a general equilibrium model without considering the consumption of amenities. They focus less on the budget decision-making process and more on the hierarchy of colleges with different educational qualities.

The paper proceeds as follows. In Section 2.2, we describe the model setup and the predictions from the simulated model. In Section 2.3, we introduce the data we use, the empirical specification, and the results. Section 2.4 concludes.

2.2 Model

2.2.1 Students

An individual student values her college experience. Her utility depends on the consumption of amenities when she attends a certain college and her wage earnings after graduation. For a student i, her utility function is given by

$$u_i = \log(c_{i1}) + \beta_i \log(c_{i2}) \tag{2.1}$$

where c_{i1} is her consumption when attending college, c_{i2} is her consumption after graduation, and β_i is an aggregate discount factor.[4]

By attending college j, c_{i1} equals to A_j the spending on consumption amenities by college j. c_{i2} equals to her wage income, which is assumed to equal her human capital at graduation. That is, $c_{i2} = H_{i2}$.

We model the human capital production function based on the classic literature and the empirical evidence found in Jacob et al. (2018), with a slight modification. They find that most students value the consumption of amenities, while the taste for academic quality is confined to students of higher ability. They measure the ability level by test scores before entry, which we argue is highly correlated with a student's learning ability. We, therefore, model that students with higher learning abilities can better utilize the spending on educational inputs by a college, that scales their initial level of ability, or human capital through the production process.

Specifically, through attending college j, a student's human capital evolves according to the following process

$$H_{i2} = \zeta_j \cdot E_j^{\pi_i} \cdot H_{i1} \tag{2.2}$$

[4] The utility function is an aggregation of period utility function where the period utility is $u_{it} = \log(c'_{it})$ for $t = 0, \ldots, T$. Suppose student i has a (conventional) discount factor β_{i0}. Her utility is then $u_i = \sum_{t=0}^{T}(\beta_{i0}^t \log(c'_{it}))$. We aggregate her life into two periods, the first in college (assuming four years) and the second in her work life. We assume her consumption in college is constant, that is $c'_{it} = c_{i1}$ for $t = 0, \ldots, 3$, and her consumption trajectory in work life is an exponential function of her initial earning after graduation, that is $c'_{it} = c_{i2}^{v(i,t)}$ for $t = 4, \ldots, T$ where $v(i,t)$ can be a simple linear function in t to represent wage growth with experience, or take a more complex form. It may vary across individual i to reflect individual heterogeneity in the wage trajectory. Her utility can thus be written as $u_i = (\sum_{t=0}^{3} \beta_{i0}^t) \log(c_{i1}) + (\sum_{t=4}^{T} \beta_{i0}^t v(i,t)) \log(c_{i2})$, or $u_i = \log(c_{i1}) + \beta_i \log(c_{i2})$ where $\beta_i = [\sum_{t=4}^{T} \beta_{i0}^t v(i,t)]/[\sum_{t=0}^{3} \beta_{i0}^t]$.

where ζ_j is the human capital production technology for institution j, E_j is the spending on educational inputs at institution j, π_i is student i's ability in learning, and H_{i1} is the initial level of human capital she is endowed with.

2.2.2 Colleges

Colleges have many objectives. To name a few, most colleges aim to promote leadership, innovation, equity, and social responsibility. As efforts to achieve those objectives are usually complementary rather than contradictory, the literature has so far used either a single measure as the objective or a combination of a few, while there is not a consensus on the objective function of colleges. An arguably most common objective for colleges is to promote their reputation. While different indicators have been used as proxies for reputation, such as admission rates and peer quality at admission, they are the results rather than the causes of changes in college reputation. Using Ph.D. programs as an example, if placements have been undesirable for consecutive years, the program will likely become less attractive to potential incoming students. Similarly, if earnings for graduating cohorts at a college have been lower than the expectation, after taking into consideration changes in general labor market conditions, it will likely constitute a decline in the reputation of the college. Moreover, although there has been some adoption of the value-added measure in postsecondary institutions in evaluating the effectiveness of education (e.g., Mountjoy and Hickman (2021)), it is not as accessible to the general public. We, therefore, argue that it is the earning outcomes of the graduating cohorts that contribute most to the movements in the perceived reputation of colleges by the public, and is also what colleges aim to maximize.

We assume that colleges aim to maximize the average earnings of their graduating students, that is

$$U_j = \frac{1}{N_j} \sum_{i \in I_j} w_{i2} \tag{2.3}$$

where N_j is the number of students admitted, I_j is the set of the index of students i that attend college j, and w_{i2} is the wage earning for student i at graduation, which we assume to be equal to her human capital H_{i2} after the knowledge learning (production) process at college j.

We further assume that there is more demand for colleges than capacities available, and the maximum capacities are exogenously determined. Colleges cannot turn students away until they are at full capacity. This assumption ensures in our model that no college can maximize the average student earnings by choosing to admit a single high-achieving student and reject all other applications. Therefore, the preference of college over every student is merely based on their w_{i2}, or equivalently H_{i2}, which is deterministic on the college's characteristics and choice. Practically, colleges make many decisions in companion with their budget decisions, such as the number of students they admit (subject to a physical capacity), the price they charge, and the programs they offer, and in the meantime, they also face a budget constraint. As our intention with this model is to illustrate how the competition between colleges affects budget allocation decisions rather than to fully describe the complex joint decision-making process, we simplify those aspects and take them as given.[5]

[5]Epple et al. (2006) study a general equilibrium model that predicts a hierarchy of colleges that differ by the educational quality provided, without the component of consumption amenities.

For a given budget (total spending per student) B_j, a college chooses s_j, the share of spending on educational inputs. The remaining is then the spending on consumption amenities. Effectively, $E_j = s_j B_j$ and $A_j = (1 - s_j) B_j$.

2.2.3 Admission

The student utility conditional on attending college j can be written as

$$u_i = \underbrace{\log(1 - s_j) + \beta_i \pi_i \log(s_j)}_{\text{Budget decision of college } j} + \underbrace{\log(B_j)(1 + \beta_i \pi_i) + \beta_i \log(\zeta_j)}_{\text{Characteristics of college } j} + \beta_i \log(H_{i1}) \tag{2.4}$$

The terms in the first brace depend on the budget decision of college j, and student characteristics. The terms in the second brace depend on college characteristics that are assumed to be exogenously determined in this model, as well as student attributes. The last term does not depend on college choices or characteristics. It only reflects a shift in the level of utility a student can achieve given her initial level of human capital.

For a specific type of student (β_i, π_i, H_{i1}), there is an optimal level of s_j that college j can choose that will maximize the terms in the first brace. However, it does not necessarily make the college j most attractive to that type of student if there is another college j' with better human capital production technology $\zeta_{j'}$ and/or higher overall budget spending per student $B_{j'}$, and an adequate choice of $s_{j'}$ that will yield a higher overall utility.

In fact, considering all postsecondary institutions and all pairwise comparisons, many of those pairs are of vastly different in their ζ_j and B_j such that the student body they each try to attract do not overlap. In such cases, the decision made by one college will not impact the other from the perspective of competition for good students.[6] In our model, if the difference between two colleges in the terms in the second brace is too large, such that regardless of the choice of $s_{j'}$, there is no s_j that will make $u_{i|j} > u_{i|j'}$ for any student type, then the choice of $s_{j'}$ should not have a direct impact on s_j.[7]

We assume that the matching of colleges and students follows a student-proposing deferred acceptance mechanism proposed in Gale and Shapley (1962). As the preferences for both sides are deterministic based on the student characteristics (β_i, π_i, H_{i1}), college characteristics (ζ_j, B_j) and college budget decision (s_j), there is a deterministic matching outcome

$$\mathcal{I}(s_1, \ldots, s_J) = \{I_1(s_1, \ldots, s_J), \ldots, I_J(s_1, \ldots, s_J)\} \tag{2.5}$$

where I_j index the set of students i that are admitted to college j. We emphasis here that for all $j = 1, \ldots, J$, I_j depends on the vector of choices $(s_1, \ldots, s_J)$ that is made by all colleges. Moving beyond, we write $\mathcal{I} = \{I_1, \ldots, I_J\}$ to make the expression less cumbersome.

The college utility is thus

$$U_j = \frac{1}{N_j} \sum_{i \in I_j} \zeta_j \cdot (s_j B_j)^{\pi_i} \cdot H_{i1} \tag{2.6}$$

College j makes its choice of s_j to maximize U_j, conditional on the vector of choices s_{-j} made by other colleges. It is analytically infeasible to find the global optimum using the

[6]Practically, a community college will likely not consider the competition for good students with universities in the Ivy League.

[7]The choice of $s_{j'}$ may have an impact on s_j indirectly, through the competition for students with college(s) j'' that are of technology $\zeta_{j''}$ and budget $B_{j''}$ in between the values of j and j'.

first order condition $\frac{\partial U_j(s_j, s_{-j})}{\partial s_j}$ as the set of individuals over which U_j is calculated, I_j, is also changing as s_j changes. Let $s_j^*(s_{-j}) = \arg\max_{s_j} U_j$ denote the optimal choice made by college j, conventionally known as the best response function of college j to other college's choices.

The equilibrium is at the intersection of the best response functions of all colleges. It is defined as the set of choices of $(s_1, \ldots, s_J)$ such that $s_j^* = s_j^*(s_{-j}^*)$ for all $j = 1, \ldots, J$, or to put in words, each college's choice is maximizing their utility in response to the choices made by all other colleges. The set of students admitted to each college in this equilibrium is a deterministic function of $(s_1, \ldots, s_J)$, and the fact that the matching is made through a centralized mechanism ensures that the capacity constraints for colleges are satisfied.

While the determination of the equilibrium and its uniqueness are important in estimating the model parameters, the computational burden grows exponentially as the number of institutions that we consider increases. It soon passes what can be calculated within a feasible amount of time. We hence turn to simulated examples to illustrate the idea and obtain insights. We are particularly interested in the sign of the slopes of the best response functions, that is, how s_j changes as s_{-j} changes. We test empirically whether and how colleges respond to their competitors' choices of s_{-j}.

2.2.4 Simulated Example

Students are heterogeneous in their aggregate discount factor β_i for future consumption, learning abilities π_i, and initial human capital H_{i1}. We assume that they follow a joint normal distribution

$$
\begin{bmatrix} \beta_i \\ \pi_i \\ H_{i1} \end{bmatrix} \sim N \left[\begin{pmatrix} \mu_\beta \\ \mu_\pi \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_\beta^2 & \rho_{12}\sigma_\beta\sigma_\pi & \rho_{13}\sigma_\beta\sigma_H \\ \rho_{12}\sigma_\beta\sigma_\pi & \sigma_\pi^2 & \rho_{23}\sigma_\pi\sigma_H \\ \rho_{13}\sigma_\beta\sigma_H & \rho_{23}\sigma_\pi\sigma_H & \sigma_H^2 \end{pmatrix} \right] \tag{2.7}
$$

If ρ_{12} is positive, that is if students who value future consumption more are in the meantime more likely to be of higher learning abilities, the extent that colleges can become generally more attractive to students by increasing their spending on consumption amenities will be weakened. To the extent that they are perfectly positively correlated, the above strategy remains valid. If ρ_{12} is negative, the college will be more able to strategically attract better students by increasing their spending on consumption amenities.

By similar arguments, the sign of ρ_{13} and ρ_{23} only affect the extent to which colleges can strategically make their choices, but do not eliminate the possibility. To simplify the model, we restrict that $\rho_{12} = \rho_{13} = \rho_{23} = 0$.[8] The model predictions are qualitatively unaffected when we choose positive (or negative) values of ρ.[9]

Instead of estimating the parameters $(\mu_\beta, \mu_\pi, \mu_H)$ and $(\sigma_\beta, \sigma_\pi, \sigma_H)$, we specify them based on prior knowledge. β_i is an aggregate discount factor as discussed in section 2.2.2. Based on a yearly discount factor of 0.96, an annual wage growth rate of 2%, and assuming a 40-years work life, we set $\mu_\beta = 6.5$. The mean value of π_i is chosen to match the moment observed in our data. In Equation 2.4, the optimal level of s_j is $\frac{\beta_i\pi_i}{1+\beta_i\pi_i}$. In the data, we observe that the average share of spending on educational inputs is around

[8]Intuition may suggest a positive ρ_{23}. Students with a higher initial endowment of human capital, which in our setting is the amount of human capital when they graduate from high school and before entering college, should also be more likely to have higher learning abilities, as students who are better at learning will have learned more before graduating from high school. It is less intuitive in terms of the signs of ρ_{12} and ρ_{13}.

[9]See Figure B.1 and Figure B.2 in Appendix B.2.

30%. Therefore, we set $\mu_\pi = 0.07$ so that $E\left(\frac{\beta_i \pi_i}{1+\beta_i \pi_i}\right) \approx 0.3$. The distribution of the initial amount of human capital (H_{i1}) cannot be separately identified from the human capital production technology in our model. Fortunately, it only provides variation in the students' baseline quality and does not interact with the school's choice or the school's characteristics. Therefore, we set it arbitrarily with $\mu_H = 1$. The standard deviations $(\sigma_\beta, \sigma_\pi, \sigma_H)$ are chosen as $(1.4, 0.017, 0.2)$ to generate the necessary variation in the student attributes.

We further simplify the case to two colleges in our simulated example. There are several reasons for making this simplification. First and foremost, as we are interested in whether a college strategically responds to budget changes of its competitors, but not whether and how such responsiveness differs when budget decisions are made deferentially among competing institutions, it is not necessary to have more than one competing institution. Second, we may conceptually view the combination of all competitors as a single institution. Third, it is computationally costly and visually difficult to calculate and represent the best response function of choices made by the competitors when there is more than one competitor. It has been a common practice in the literature when joint admission decisions are made for many colleges to aggregate multiple institutions of reasonably similar characteristics into one (Jacob et al. (2018); He et al. (2021)). We acknowledge that the results obtained under such simplification can not be generalized. We leave it to future work to consider the more complex strategic responds among multiple institutions.

In Figure 2.1, we plot the simulated best response functions of two colleges. The parameters we set for the two institutions are very close so that the pressure from its competitor's budget choice remains relevant throughout the domain. We see that when the competitor is setting a high level of s, that is, when the competitor is spending a high share of its budget on educational inputs, it is best to undercut the competitor. This reflects the idea that by slightly lowering the spending on educational inputs and increasing that on consumption amenities, a college will be able to become slightly more attractive overall, and hence admit a student body with a higher baseline level of human capital. We expect a positive relationship between a college's spending and its competing institution over this range. The point where the two functions intersect is at around 35%. This is a construct of the way we set μ_π. The best response functions below 35% are negatively sloped. Over this range, the competitor is setting a very low share of spending on educational inputs, and it is also below the optimal level for most types of students, as we can see from Equation 2.4. Over this range, the competitor is becoming generally less attractive as it lowers its s. The pressure from the competition for good students, therefore, becomes less relevant. The college can increase its utility by setting a higher level of s, as a higher level of s directly improves the human capital production outcomes of its students.

We simulate the second case of two colleges that are more heterogeneous. In Figure 2.2, the difference in human capital production technology ζ_j and total spending per student B_j is more significant between the two colleges than what we set in Figure 2.1. When the two colleges are less homogeneous, college 2 does not respond to budget choices made by college 1 when s_1 is between 0.29 and 0.43. Within this range, college 1 is more attractive to students than college 2, regardless of the choice of college 2. College 2 is therefore unable to compete for a better student body through its choice of s_2, hence setting s_2 to the maximum possible value as doing so will best promote the student it can admit. As stated above, the competitor (college 1 in this case) is the combination of colleges that are relevant in college 2's decision-making process. As we include more colleges that are more heterogeneous from college 2 in that set, we expect college 2's responsiveness to decline.

As the model parameters are specified rather than estimated, we refrain from drawing quantitative conclusions based on our model but only predict the signs of the slope. We turn to quantitative analysis in our empirical section.

2.3 Data and Empirical Results

2.3.1 Data

In this paper, we use the college characteristics from the Integrated Postsecondary Education Data System (IPEDS). We specifically use the variables of spending on different expenditure categories (instruction, research, student service, etc) to calculate s_j in our model. The data contain information on the latitude and longitude of institutions, allowing us to calculate the distance between two institutions.[10] We also utilize the Carnegie Classifications of colleges which are based on the purpose of the institution. We aggregate over twenty detailed categories into seven broad categories.[11] We assume that colleges within each of those categories are more likely to target the same student body when making strategic decisions. In robustness checks, we relax this assumption.

[10] We use the formula $d(j,j') = 3963 \cdot \arccos(\sin(lat_j) \cdot \sin(lat_{j'}) + \cos(lat_j) \cdot \cos(lat_{j'}) \cdot (\cos(long_j - long_{j'}))))$ to calculate the earth surface distance, where $(lat_j, long_j)$ and $(lat_{j'}, long_{j'})$ are the coordinates of institution j and j'.

[11] The seven categories are associate's colleges, special focus two-year colleges, doctoral universities, master's colleges and universities, baccalaureate colleges, special focus four-year, and others.

2.3.2 Empirical Specification

We consider the following regression equation

$$s_{j,t} = \beta \cdot \bar{s}_{-j,t} + \eta_j + \theta_t + \mathbf{C}_{j,t} + \varepsilon_{j,t} \tag{2.8}$$

where $s_{j,t}$ is a measure of spending on educational inputs at college j in year t, $\bar{s}_{-j,t} = \frac{\sum_{j' \in I_{-j,t}} s_{j',t}}{|I_{-j,t}|}$ is an average of $s_{-j,t}$ over the set of relevant competing institutions, denoted as $I_{-j,t}$ which we will discuss in more detail below about how we construct it empirically, η_j is the institution fixed effects, θ_t is the time fixed effects, $\mathbf{C}_{j,t}$ is a vector of institution characteristics, and $\varepsilon_{j,t}$ are assumed to be idiosyncratic errors.

Identification issues will arise if $\varepsilon_{j,t}$ is correlated with $\bar{s}_{-j,t}$, conditional on all other variables included. As $\bar{s}_{-j,t}$ is a measure of group averages, the issue caused by the association between any individual college's budget choice $s_{j',t}$ for $j' \in I_{-j,t}$ and $\varepsilon_{j,t}$ is ameliorated after the aggregation. The group level association between $\bar{s}_{-j,t}$ and $\varepsilon_{j,t}$ can be more serious for cases where the number of institutions in $I_{-j,t}$ is small. For example, if two institutions are cooperating and make their budget decisions jointly, we will underestimate the competitive pressure. We argue that in such cases, the cooperative relationship is usually long-lasting. By including institution fixed effects, our estimates are less susceptible to biases caused by such issues. Another argument for this concern is that for the vast majority of the institution-year observations in our sample, the $\bar{s}_{-j,t}$ is averaged over 10 colleges. It is much less likely for the average of a large group of colleges to be associated with the idiosyncratic error term. To further account for issues such as changes in general labor market conditions, we also include the year fixed effects in our regression.

2.3.3 Share of Spending on Educational Inputs

We estimate the following regression equation

$$s_{j,t} = \sum_{p=1}^{7} (\beta_p \cdot 1(CC_j = p) \cdot \bar{s}_{-j,t}) + \eta_j + \theta_t + \mathbf{C}_{j,t} + \varepsilon_{j,t} \tag{2.9}$$

where $s_{j,t}$ is the share of spending on educational inputs at college j in year t, $\bar{s}_{-j,t} = \frac{\sum_{j' \in I_{-j,t}} s_{j',t}}{|I_{-j,t}|}$ is an average of $s_{-j,t}$ over the set of relevant competing institutions, denoted as $I_{-j,t}$, and p from 1 to 7 indicates the seven Carnegie Classifications. We interact the average share of spending at competing institutions with the type of college to allow for a different slope coefficient. In this section, we construct $I_{-j,t,x} = \{j' : (1-x) \cdot B_{j,t} \leq B_{j',t} \leq (1+x) \cdot B_{j,t}\}$ where $x = 0.05$, 0.1, 0.25, and 0.5 so that the results are robust to the choice of x, and B_j is the overall budget of institution j in year t. Such construction is based on the simulation results where we find only colleges of close B_j and ζ_j compete against each other. As we do not observe and cannot quantify ζ_j empirically, we choose the set of relevant competing institutions as those with similar levels of the overall budget.

Table 2.1 presents the estimation results. Across the board, colleges of all types respond to spending changes in competing institutions. A 10 percentage points increase in the average share of spending on educational inputs at competing institutions is associated with a 4 to 7 percentage points increase in the share of spending on educational inputs. Associate's colleges, doctoral universities, master's colleges and universities, special focus four-year colleges, and others are more responsive at between 6 to 7 percentage points. Special focus two-year colleges and baccalaureate colleges are less responsive at between 4 and 5 percentage points. The results are qualitatively similar for different values of x.

As is shown in the simulated example, the best response function is U-shaped. We further investigate if such a pattern is observed empirically. To do so, we estimate the following regression equation

$$s_{j,t} = \sum_{k=1}^{5}\sum_{p=1}^{7}\left(\beta_{p,k} \cdot 1(CC_j = p) \cdot 1\left(\frac{k-1}{5} \leq \bar{s}_{-j,t} \leq \frac{k}{5}\right) \cdot \bar{s}_{-j,t}\right) + \eta_j + \theta_t + \mathbf{C}_{j,t} + \varepsilon_{j,t} \quad (2.10)$$

where the terms are defined the same as above and we further allow five different slopes for different values of $\bar{s}_{-j,t}$ for each type of college p.[12] A U-shaped best response function predicted by our model will correspond to negative values of $\beta_{p,k}$ when $k = 1$ or $k = 2$, which increase and become positive as k increases.

Table 2.2 shows the results for the above exercise. We find mixed evidence for the above hypothesis. For associate's colleges, we find small and positive coefficients when $k =$, which increases as k increases. Special focus two-year colleges are closest to our prediction of a U-shaped function. The slope is negative when $k = 1$ and is positive when $k \geq 2$, though not increasing with k. Doctoral universities, master's colleges and universities, and others all have large and positive slopes when $k = 1$, and the slopes decrease as k increases. For baccalaureate colleges and special focus four-year colleges, the pattern is unclear. The empirical evidence does not confirm the U-shaped best response function, possibly because that our model did not completely consider the joint decision-making process.

[12] Dividing the unit interval into five is a compromise. Ideally, we want to divide it as fine-grained as possible to allow for more flexible slope coefficients. In the data, very few institution-year observations are associated with an average share of spending on educational inputs by their competing institutions that are below 0.3 or above 0.6. Our attempt to empirically shed light on the best response function away from the equilibrium value is impeded by the insufficient amount of observations in those ranges.

2.3.4 Log of Spending on Educational Inputs

We then test the prediction of whether a lower level of competitive pressure is associated with a lower level of responsiveness. We consider the case of using the log level of spending on educational inputs as our variable of interest.[13] For values of d ranging from 10 to 500 and for each one of the seven types of institutions, we estimate the following regression equation

$$\ln(E_{j,t,d}) = \beta_1 \cdot \ln(\bar{E}_{-j,t,d}) + \eta_j + \theta_t + \varepsilon_{j,t,d} \tag{2.11}$$

where $E_{j,t,d}$ is the spending per student on educational inputs at college j in year t. Here, we consider the relevant competing colleges as those of the same type of colleges within a radius d. Let $d(j, j')$ denote the distance between two colleges j and j'. The set of relevant colleges for college j in year t is defined as $I_{-j,t,d} = \{j' : CC_j = CC_{j'}, d(j, j') \leq d\}$ where CC_j is the Carnegie Classification of college j. Therefore, $\bar{E}_{-j,t,d} = \frac{\sum_{j' \in I_{-j,t,d}} E_{j',t}}{|I_{-j,t,d}|}$ is the average spending per student on educational inputs over the set of relevant competing institutions.

Figure 2.3 plots the estimated coefficients and the corresponding 95% confidence interval for the above regression equations. For doctoral universities, baccalaureate colleges, and special focus four-year colleges, we cannot reject the null hypothesis that they do not respond to changes in spending on educational inputs by other colleges of the same type due to imprecisely estimated standard errors. Associate's colleges, special focus two-year colleges, and master's colleges and universities respond to a 10 percent increase from their

[13] We are slightly departing from the model. Practically, spending on educational inputs and consumption amenities does not exhaust all spending categories of a college.

competing institutions by increasing their spending on educational inputs by approximately 1 percent.

A common concern is whether the results are driven by the definition of competing institutions. Figure B.3 in Appendix B.2 shows the results when the competing institutions are not restricted to colleges of the same type, that is, we set $I_{-j,t,d} = \{j' : d(j,j') \leq d\}$. Another concern is whether the results are driven by changes in the composition of the analysis sample. Figure B.4 in Appendix B.2 presents the same results for colleges with at least one competing institution within 30 miles radius. We observe generally similar patterns and are thus assured that the above findings are robust to alternative constructions of the sample.

We further observe that colleges are most responsive to budget changes to competing colleges of close distance. While not modeled in this paper, it is well documented in the literature that students generate dislike for colleges far away from their homes, which can serve as another source of decreasing levels of competitive pressure. By relaxing the definition of relevant competing colleges to include colleges that are further away, the competitive pressure becomes lower. To test whether the responsiveness decreases as competitive pressure decreases, we estimate the following regression equation

$$\ln(E_{j,t,d}) = \beta_1 \cdot \ln(\bar{E}_{-j,t,d}) + \beta_2 \cdot \ln(\bar{E}_{-j,t,d}) \cdot d + \eta_{j,d} + \theta_t + \varepsilon_{j,t,d} \qquad (2.12)$$

where $E_{j,t,d}$ is the spending per student on educational inputs at college j in year t, and it does not vary as d varies. There are approximately 50 times more institution-year observations as a result of this construction. We thus include the institution-distance fixed effects $\eta_{j,d}$ and the year fixed effects θ_t to reduce potential bias. We also use the cluster-robust standard error at the institution-distance level to obtain reliable inference results.

Table 2.3 presents the regression results for six different types of colleges. We first observe that in general, colleges are responding positively to increases in spending on educational inputs by their competing colleges. This was not clear from the previous figures, possibly due to imprecisely estimated standard errors. Associate's college, special focus two-year colleges, and baccalaureate colleges are most responsive to such changes. If competing institutions increase spending by 10 percent, those three types of colleges will respond by increasing spending by 1.6 to 1.8 percent. Doctoral universities, master's colleges and universities, and special focus four-year colleges also respond positively, but to a lesser extent. A 10 percent increase in spending on educational inputs by their competing institutions is only associated with a 0.5 to 0.9 percent increase in response. The estimates are consistent with what we see in Figure 2.3.

The estimated coefficients on the product term of the average spending and the distance in the second row shed light on whether a higher distance between institutions leads to lower levels of responsiveness to budget changes. Except for doctoral universities, the responsiveness declines for all other types of colleges and will reach zero when we expand the relevant competing institutions to those that are around 750 to 1,000 miles far. It confirms our conjecture that the competitive pressure is higher among institutions that have a more common target student body.

2.4 Conclusion

In this paper, we construct a simple model to illustrate the idea that colleges make budget decisions as an attempt to compete for good students. A simulated example shows that the best response function of a college to its competitor's budget decisions can be U-shaped. We also show that such responsiveness declines as the competitive pressure weaken when the colleges become more heterogeneous. While our theoretical model has simplified the joint decision-making process, we view it as a base model that is informative and also can be extended for more comprehensive analysis.

Our empirical analysis finds evidence that generally supports the predictions from the simulated example. Overall, colleges respond positively to budget increases in educational inputs by their competitors. In terms of share of spending on educational inputs, the response is between 4 to 7 percentage points to an increase of 10 percentage points. In terms of response to log levels of spending, the estimated elasticities are between 5 to 18 percent. Those results are supportive of the upward sloping part of the U-shaped best response function. We attempt to search for evidence for the downward sloping part, and we find mixed evidence. Of the seven Carnegie classifications of colleges that we consider, only the special focus two-year colleges exhibit a negative slope when competitors are spending a low share on educational inputs, as the model would predict. Other types of colleges always have positive slopes, unless when imprecisely estimated.

The findings in this paper can be extended in several ways. First, our model does not reflect the dynamics in the accumulation of human capital production technology at colleges. The competition for good students can be even more important in a model where the human capital production technology (ζ_j) evolves based on the quality of the graduating students. Unlike in firms where production technology benefits from investment in R&D, the human capital production technology can benefit from the interactions between good students and teachers, which does not require monetary input but requires high-quality students as input. Second, we assumed an exogenously determined capacity of colleges so that nothing happens on the enrollment margin as college budget spending changes. While such simplification allows us to better focus on the idea we propose, it ignores a potentially interesting interaction between the number of students enrolled and their quality. To jointly consider those issues, a dynamic structural model that extends the work by Epple et al. (2006) with its structural parameters estimated is called for.

Table 2.1: Response in Spending to Changes in Competing Institutions, by Type

	(1) $I_{5\%}$	(2) $I_{10\%}$	(3) $I_{25\%}$	(4) $I_{50\%}$
Associate's Colleges$\times\bar{s}_{-j}$	0.59***	0.67***	0.74***	0.79***
	(0.03)	(0.04)	(0.05)	(0.05)
Special Focus Two-Year Colleges$\times\bar{s}_{-j}$	0.41***	0.36**	0.40**	0.47***
	(0.09)	(0.11)	(0.13)	(0.13)
Doctoral Universities$\times\bar{s}_{-j}$	0.57***	0.62***	0.67***	0.71***
	(0.04)	(0.04)	(0.05)	(0.05)
Master's Colleges and Universities$\times\bar{s}_{-j}$	0.58***	0.66***	0.72***	0.78***
	(0.04)	(0.04)	(0.05)	(0.05)
Baccalaureate Colleges$\times\bar{s}_{-j}$	0.40***	0.43***	0.46***	0.52***
	(0.04)	(0.05)	(0.05)	(0.06)
Special Focus Four-Year Colleges$\times\bar{s}_{-j}$	0.58***	0.59***	0.59***	0.62***
	(0.06)	(0.07)	(0.08)	(0.08)
Others$\times\bar{s}_{-j}$	0.68***	0.70***	0.74***	0.79***
	(0.02)	(0.03)	(0.03)	(0.03)
Observations	162282	162282	162282	162282

Notes: The dependent variable is the share of spending on educational inputs. The independent variable is the same measure averaged over colleges of similar overall budget. Column (1) averages over colleges with overall budget within 5% difference. Column (2), (3), and (4) average over colleges with overall budget within 10%, 25%, and 50%, respectively. A different slope is estimated for $\bar{s}_{-j}$ for each type of the college. All regressions include institution fixed effects and year fixed effects. Standard errors are clustered at the institution level.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.2: Response in Spending to Changes in Competing Institutions, by Institution Type and Interval

	(1) $I_{5\%}$	(2) $I_{10\%}$	(3) $I_{25\%}$	(4) $I_{50\%}$
Associate's Colleges:				
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0, 0.2])$	0.39	0.32	-0.47***	-
	(0.51)	(0.35)	(0.11)	-
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.2, 0.4])$	0.43***	0.46***	0.51***	0.57***
	(0.04)	(0.05)	(0.06)	(0.06)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.4, 0.6])$	0.47***	0.51***	0.56***	0.61***
	(0.03)	(0.04)	(0.05)	(0.06)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.6, 0.8])$	0.66***	0.64***	1.05***	1.02***
	(0.03)	(0.09)	(0.03)	(0.07)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.8, 1])$	-	-	-	-
Special Focus Two-Year Colleges:				
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0, 0.2])$	-1.39*	-1.48*	-0.94*	-0.85
	(0.67)	(0.69)	(0.40)	(0.70)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.2, 0.4])$	0.37**	0.32*	0.25	0.44**
	(0.12)	(0.14)	(0.15)	(0.16)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.4, 0.6])$	0.39***	0.33**	0.30*	0.44**
	(0.11)	(0.13)	(0.14)	(0.15)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.6, 0.8])$	0.32**	0.33*	0.31*	0.45**
	(0.12)	(0.14)	(0.15)	(0.16)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.8, 1])$	0.28***	0.18	-	-
	(0.07)	(0.11)	-	-
Doctoral Universities:				
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0, 0.2])$	0.68***	0.81***	0.83***	0.85***
	(0.07)	(0.08)	(0.10)	(0.11)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.2, 0.4])$	0.65***	0.73***	0.78***	0.82***
	(0.04)	(0.05)	(0.06)	(0.06)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.4, 0.6])$	0.62***	0.69***	0.73***	0.77***
	(0.04)	(0.05)	(0.05)	(0.06)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.6, 0.8])$	0.50***	0.46***	0.45***	0.48***
	(0.07)	(0.03)	(0.04)	(0.05)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.8, 1])$	-	-	-	-
Master's Colleges and Universities:				
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0, 0.2])$	-	-	-	-
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.2, 0.4])$	0.74***	0.88***	1.00***	1.12***
	(0.04)	(0.05)	(0.06)	(0.07)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.4, 0.6])$	0.71***	0.84***	0.94***	1.05***
	(0.04)	(0.05)	(0.05)	(0.06)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.6, 0.8])$	0.35***	0.42***	0.56***	-
	(0.02)	(0.02)	(0.03)	-
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.8, 1])$	-	-	-	-

Table 2.2 (Continued)

	(1)	(2)	(3)	(4)
	$I_{5\%}$	$I_{10\%}$	$I_{25\%}$	$I_{50\%}$
Baccalaureate Colleges:				
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0, 0.2])$	-	-	-	-
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.2, 0.4])$	0.45***	0.47***	0.57***	0.70***
	(0.05)	(0.05)	(0.07)	(0.07)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.4, 0.6])$	0.44***	0.46***	0.54***	0.65***
	(0.04)	(0.05)	(0.06)	(0.07)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.6, 0.8])$	-	-	-	-
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.8, 1])$	-	-	-	-
Special Focus Four-Year Colleges:				
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0, 0.2])$	0.39*	0.50*	0.58	0.46
	(0.18)	(0.20)	(0.35)	(0.30)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.2, 0.4])$	0.66***	0.69***	0.72***	0.78***
	(0.07)	(0.08)	(0.09)	(0.09)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.4, 0.6])$	0.62***	0.65***	0.67***	0.72***
	(0.06)	(0.07)	(0.08)	(0.09)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.6, 0.8])$	0.49***	0.54***	0.46***	0.56***
	(0.06)	(0.07)	(0.08)	(0.08)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.8, 1])$	-	-	-	-
Others:				
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0, 0.2])$	0.72***	0.51**	0.49*	0.87
	(0.15)	(0.19)	(0.21)	(0.47)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.2, 0.4])$	0.66***	0.64***	0.67***	0.70***
	(0.03)	(0.04)	(0.04)	(0.04)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.4, 0.6])$	0.68***	0.68***	0.72***	0.74***
	(0.03)	(0.03)	(0.04)	(0.04)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.6, 0.8])$	0.64***	0.62***	0.63***	0.69***
	(0.02)	(0.03)	(0.04)	(0.04)
$\times\bar{s}_{-j} \times 1(\bar{s}_{-j} \in [0.8, 1])$	0.65***	0.65***	0.68***	0.70***
	(0.03)	(0.04)	(0.06)	(0.07)
Observations	162282	162282	162282	162282

Notes: The dependent variable is the share of spending on educational inputs. The independent variable is the same measure averaged over competing colleges. Column (1), (2), (3), and (4) averages over colleges with overall budget within 5%, 10%, 25%, and 50% difference, respectively. A different slope is estimated for $\bar{s}_{-j}$ in one of the five intervals, interacted with the type of the college. All regressions include institution fixed effects and year fixed effects. Standard errors are clustered at the institution level.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

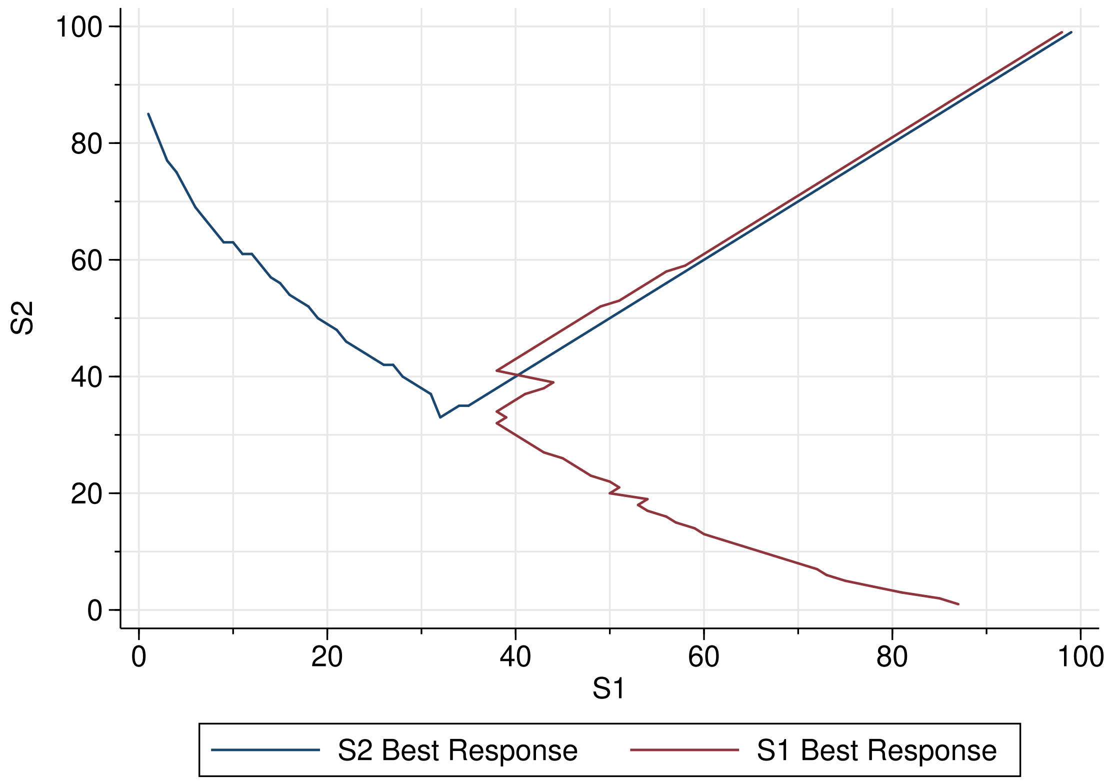
Table 2.3: Response in Spending to Changes in Competing Institutions

	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Avg_InExp)	0.16***	0.18***	0.05***	0.09***	0.16***	0.06***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)
Ln(Avg_InExp)	-0.21***	-0.19***	0.01	-0.11***	-0.25***	-0.10**
×Distance (1,000 miles)	(0.02)	(0.04)	(0.02)	(0.03)	(0.07)	(0.03)
Observations	1075744	248611	422652	670528	642949	716366

Notes: The dependent variable is spending per student on educational inputs. The independent variable is the same measure averaged over colleges of the same type within a radius d that ranges from 10 to 500 miles. The seven types of institutions are categorized based on the Carnegie Classifications. Each column represent the regression results for one type of the institution. All regressions include institution-distance fixed effects and year fixed effects. Standard errors are clustered at the institution-distance level. Column (1): Associate's Colleges; Column (2): Special Focus Two-Year; Column (3): Doctoral Universities; Column (4): Master's Colleges and Universities; Column (5): Baccalaureate Colleges; Column (6): Special Focus Four-Year.

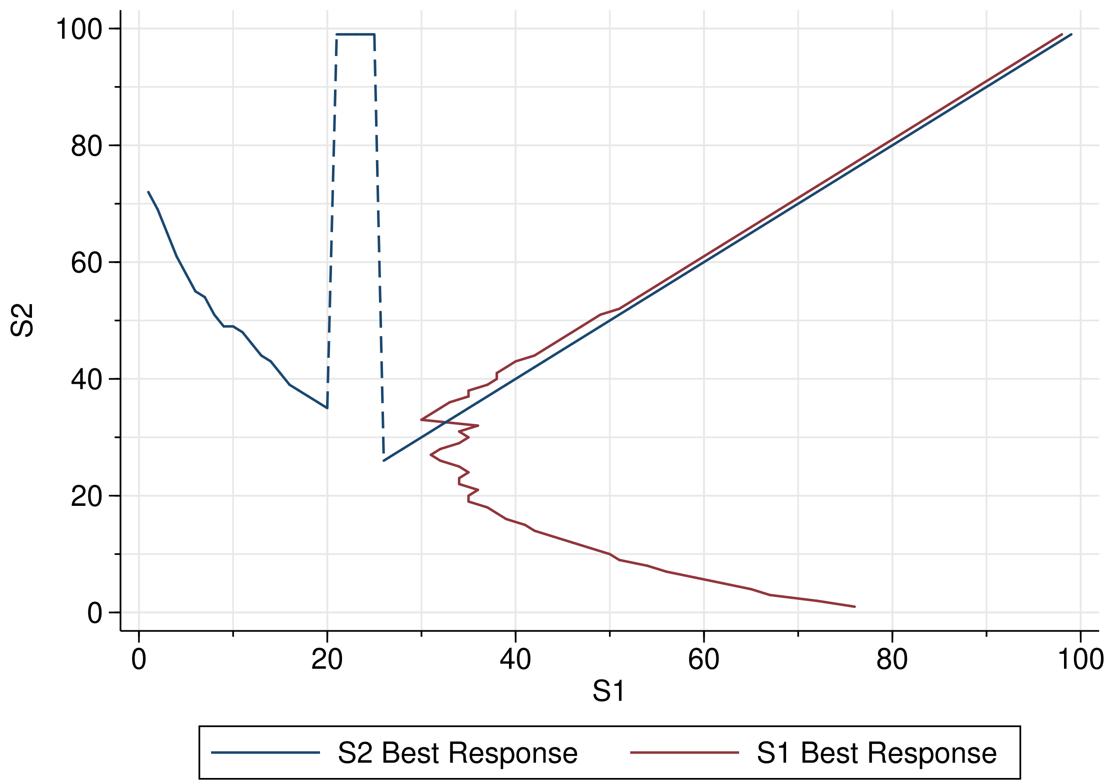
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 2.1: Best response functions between two colleges, higher competitive pressure



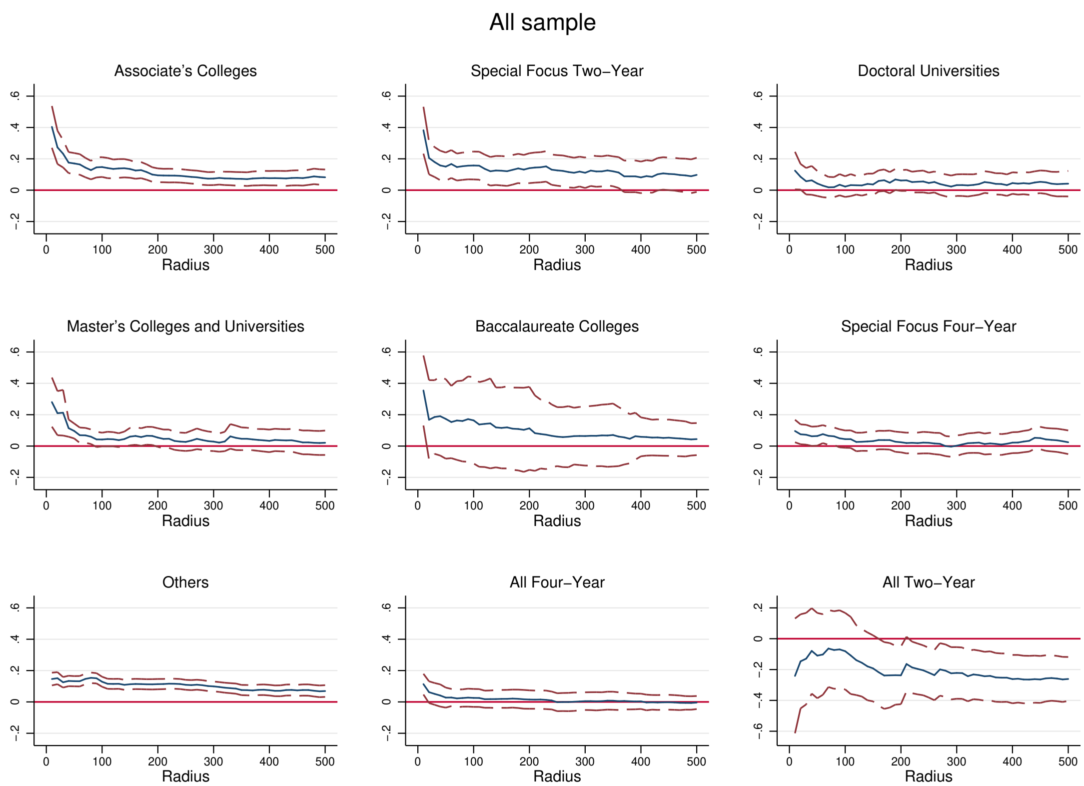
Notes: This figure plots the best response functions of two colleges in setting their share of spending on educational inputs s_j to each other's decisions. Characteristics for college 1 is technology in human capital production $\zeta_1 = 1.1$, total spending per student $B_1 = 8010$, and capacity $N_1 = 800$. Characteristics for college 2 is $\zeta_2 = 1.101$, $B_2 = 8000$, and $N_2 = 1,500$. Total number of students is $N = 10,000$ and is drawn from the distribution in Equation 2.7.

Figure 2.2: Best response functions between two colleges, lower competitive pressure



Notes: This figure plots the best response functions of two colleges when the competitive pressure is lower. Characteristics for college 1 is technology in human capital production $\zeta_1 = 1.11$, total spending per student $B_1 = 8000$, and capacity $N_1 = 1,500$. Characteristics for college 2 is $\zeta_2 = 1.1$, $B_2 = 8100$, and $N_2 = 800$. Total number of students is $N = 10,000$ and is drawn from the distribution in Equation 2.7.

Figure 2.3: Response to change in spending on educational inputs by competing institutions



Notes: This figure plots the estimated coefficients and the 95% confidence intervals as responses of a college to changes in its competing institutions' spending on educational inputs. Competing institutions are defined as institutions located within the specified miles radius and are of the same Carnegie classification.

Chapter 3

Bounding Average Treatment Effects Using Observed or Unobserved Variable

3.1 Introduction

The average causal effect of one treatment variable on an outcome of interest is often paramount in empirical research. As a result, researchers often either design or exploit random variation in the treatment in order to estimate such average treatment effects (ATEs). In such settings, researchers may recover the local average treatment effect (LATE), in which 'local' typically refers to a population that complies with the randomization (Imbens and Angrist, 1994). There is another way, however, in which these estimates are local - the sample into randomness. While much attention was given to generalize from the LATE to the experiment sample ATE, or the selection into compliance, relative less attention was given to the selection into experiment. If we wish to place bounds around the population ATE, both selection margins are important.

Experimental samples may differ from the population at large due to researcher site selection, would-be participants' decisions about whether or not to participate, or simply from micronumerosity (Allcott, 2015). For example, in conducting the Perry Preschool experiment, researchers recruited and enlisted children from surrounding neighborhoods in Ypsilanti, Michigan who had IQ scores that ranged from 75 to 85 (Weikart et al., 1978). Individuals randomized in the National Supported Work Demonstration used in the LaLonde (1986) evaluation of observational methods, the Moving to Opportunity housing voucher experiment (Goering et al., 1999), and the Oregon health insurance experiment (Finkelstein et al., 2012) are all self-selected into the experimental sample to some degree. More broadly, institutional review boards (IRB) largely require consent from participants for field experiments. Regardless of the source, researchers often worry about bias between the treatment effects of the experiment sample and that of the general population of interest arising from such differences. While the issue is most commonly discussed with experiments, sample selection is not unique to randomized control trials. Researchers often must make choices when compiling analysis samples, exogenous variation does not always occur in the populations of most interest, and often participants may choose whether to complete survey requests.

Absent compulsory random variation for the entire population of interest, what can researchers learn about the ATE? Andrews and Oster (2019) use observed variables to address selection into the sample on unobserved variables. Kowalski (2018) places bounds on the sample ATE from a well-identified LATE. Each approach utilizes a different set of assumptions to extrapolate away from the group of subjects for whom a causal effect is directly estimated. We explore bounds on the population ATE when there is both noncompliance with exogenous variation in treatment within the sample while allowing for nonrandom sampling from the population of interest. We extend each approach to address both margins threatening external validity, adopting each set of relevant assumptions. We finally introduce an additional reasonable set of assumptions to use to estimate the third set of bounds and compare all three approaches.

Andrews and Oster (2019) define the following "External validity bias arises here if the effects of the treatment on the volunteers [or selected samples] differ from effects in the overall population" and state that "Our goal is to provide a framework in which to consider selection on unobservables when studying external validity." However, their approach relies on observed variables and assumptions regarding the relationship between the observed, unobserved, and dependent variables, which may not always hold. Further, while providing a valuable framework of inquiry Andrews and Oster (2019) stops short of providing a way forward in the presence of noncompliance with a randomized assignment. The question of extrapolating from a local average treatment effect (LATE) to approximate or bound the average treatment effect (ATE) remains open.

Brinch et al. (2017) and Kowalski (2018) seek to use differences in outcomes across populations to address the external validity of LATEs. Under the assumption that potential outcomes are monotonically related to the probability of treatment, Brinch et al. (2017) and

Kowalski (2018) place bounds on the ATE within the estimation sample, but unfortunately provide no way forward to deal with sample selection. Further, this assumption is strong and leads us to reject external validity even when the differences in treated and untreated average outcomes are identical between experimental participants and nonparticipants.

We propose bounds that allow for both margins of selection into compliance and into the sample. In order to narrow the bound widths, we add an additional assumption: our monotonicity of potential outcomes assumption holds that on average, sub-populations are either positively or negatively selected on both potential outcomes. This assumption seems likely to hold in many instances. Though such a framework is not necessary, it is theoretically justified by dynamic complementarities as presented in Cunha and Heckman (2007).

We apply our method to a study of the efficacy of a first-year learning program on retention rates with a randomized control trial. Our bounds suggest that the ATE of the entire first-year student population is between -0.687 and 0.056. Despite the width of these bounds, we find the upper bound to be informative. In this data, the estimated lower bound for the population ATE following Kowalski (2016) is 0.046, close to the upper bound that we find when considering individuals' participation decisions in the study. When we weigh on observed variables according to Andrews and Oster (2019), we obtain a lower bound on the global average intent to treat (ITT) of 0.044, which is again just slightly lower than our upper bound. The reason that there is little to no overlap in the bounds by Kowalski (2016) or Andrews and Oster (2019), and by ours is that the assumptions made in Kowalski (2016) or Andrews and Oster (2019) is likely untrue in this data. Using the same data, Azzam et al. (2022) finds evidence indicating negative selection into the experiment sample based on observed variables and positive selection into the experiment sample based on unobserved variables.

We view our approach as an alternative to the existing tools in the literature in assessing the issue of external validity. We impose a set of assumptions that are weak and different from the existing ones, so that our method is generally applicable, and using which is at the choice of the researcher. While our bounds may be wide, it can be informative of the sign of the population ATE.

3.2 Setup and Method

We consider the following model:

$$y_i = b_i d_i + \epsilon_i,$$

where y_i is an outcome variable, d_i stands for treatment status, $b_i = \beta + e_i$ where e_i is the heterogeneous responsiveness to treatment, and ϵ_i represents unobserved individual heterogeneity. Typically we want to identify the average causal effect of the treatment β. In RCT where the treatment status is randomly assigned, assuming perfect compliance, estimating the above regression identifies the average treatment effect $E_S(Y_1 - Y_0)$, where we use E_S to emphasize that the average is taken over the experiment sample.

To inform policy, we are interested in the parameter $E_P(Y_1 - Y_0)$ where E_P is the expectation taken over the population of interest. When the experiment sample is not randomly drawn from the population and is determined through a selection process that can be related to the responsiveness to the treatment, as is the case in most of the field experiments, we can hardly expect $E_P(Y_1 - Y_0) = E_S(Y_1 - Y_0)$. Using the estimates for $E_S(Y_1 - Y_0)$ to inform policy can therefore be misleading. We propose three methods to provide bounds for $E_P(Y_1 - Y_0)$.

RCTs are often used to evaluate existing programs. In quasi-experimental research often exogenous variation is only a subset of the variation in treatment in the population. In both designs, however, researchers rarely make use of the data in which the variation in treatment is not exogenous. One exception is the the within-study design literature launched by the seminal work LaLonde (1986). In LaLonde (1986) and the many that have come after, results from observational approaches – such as OLS, propensity score matching (PSM), and regression discontinuity – are benchmarked against experimental results. Reviews of these studies in Glazerman et al. (2003), Bloom et al. (2005), and Cook et al. (2008) have found that in most cases there are significant differences between experimental and observational results, and these differences have served to elevate RCT relative to observational approaches.

Cook et al. (2008) comment that there is heterogeneity among within-study comparisons and give seven criteria for evaluating such comparisons. One criterion is that "The experiment and observational study should estimate the same causal quantity," such as ITTs or ATEs. We add that self-selection into experiments alters the causal quantity being estimated. This is because a program may have different effects on the type of people who sign up for the experiment than it does on people who do not. If the population of interest is the experimental participants, the RCT provides a valid estimate of the policy-relevant parameter, worthy of use for benchmarking. In that case, considering individuals from an observational setting outside the population of interest alters the estimand of the observational approach. It would be unsurprising were the observational approach to fail to estimate a parameter other than its estimand. On the other hand, if the population of interest is broader than the RCT participants, without further testing we do not know whether the differences in estimates arise from internal validity failings of the observational approach or external validity failings of the RCT.

To illustrate this point, let us use the OLS estimates on the entire freshman class (which we hold here is the population of interest) for comparison against the RCT estimates. The results from columns 3 through 6 of Table 5 may be interpreted as a weighted decomposition of such OLS regression estimates in which retention is regressed upon treatment, ignoring the randomization. We acknowledge that this design is not an exemplar of observational approaches, but use it for ease of explanation. Note that $plim(\beta_{OLS}) = ATE + Bias_{OLS}$ and $plim(\beta_{IV}) = LATE$, assuming proper randomization. The local aspect of this LATE pertains to participation in the randomization, but it becomes more localized with noncompliance. The upshot here is that the two estimators may converge to different parameters, even if the controls included in OLS are sufficient to capture selection into treatment. We can relate the two parameters according to the following, maintaining perfect compliance with randomization within the experiment:

$$ATE = LATE \times P(L = 1) + E(b_i|L = 0) \times (1 - P(L = 1)) \tag{3.1}$$

where $E(b_i|L=0)$ is the average effect among those who do not participate in the experiment. It is typical in this literature to suspect bias in the observational approach. In comparing the two estimates, we observe the following:

$$plim(\beta_{OLS}) - plim(\beta_{IV}) = [1 - P(L=1)][E(b_i|L=0) - LATE] + bias_{OLS} \tag{3.2}$$

Equation 3.2 nicely demonstrates that the comparison in results provides a mixture of the possible external invalidity of the RCT ($E(b_i|L=0) - LATE$) and the possible bias in the OLS estimate. Thus, differences in estimates alone cannot point to the general failings of either approach.

3.2.1 Using Selection on Observed Variables

Andrews and Oster (2019) provides a way to correct the bias between the population ATE and the experiment sample ATE caused by observed variables. They assume that the individual treatment effects are a function of a set of observed and unobserved variables.

$$y_{i1} - y_{i0} = TE_i = \alpha + C_i'\gamma + U_i'\delta + \varepsilon_i \tag{3.3}$$

By obtaining an estimate of the coefficients for the observed variables ($\hat{\gamma}$) using the experiment sample and inserting the mean values of those observed variables from the population, they correct for the bias between the population ATE and the experiment sample ATE caused by differences in the observed variables.

In the cases where there is no bias between the population ATE and the experiment sample ATE caused by unobserved variables, the above correction can be used to generate (approximately) a valid estimate for the population ATE. They consider three special cases. In the first case, unobserved variables are unrelated to treatment effects heterogeneity ($\delta = 0$). In the second case, unobserved variables do not predict selection into the experiment. In the third case, the set of unobserved variables that predict treatment effects heterogeneity and that predicts selection into experiment are unrelated.

While in some settings one of the three special cases may be satisfied, generally they are too restrictive and unlikely to hold. In a more general setting, they provide a way to bound for the ATE in the population with an additional assumption. They assume that the bias in the ATE between the population and the experiment sample caused by differences in unobserved variables operates in the same direction and, is no larger than, the bias caused by differences in observed variables, or the corrections made as described above. With this assumption, we can calculate a bound for the population ATE based on the experiment sample ATE and the correction for bias made based on the difference in the observed variables between the experiment sample and the population.

3.2.2 Extension Beyond the ITT Using Selection on Observed Variables

Without perfect compliance, we can only identify either the LATE, or to say the ATE for the compliers, or the ITT for the experiment sample without further assumptions. To move forward, a common assumption made is the ignorability of the selection into treatment within the experiment sample. This assumption is more likely to hold when the number of observations that are not compliant with the treatment assignment is small.

Alternatively, if the data can pass the tests proposed in Black et al. (2017), which are the necessary conditions, we may also feel more confident in making this assumption.

With the ignorability of the selection into treatment assumption, we can proceed with the procedures described in the previous section and obtain the same types of conclusions based on the assumptions regarding the bias caused by selection on observed variables and the bias caused by selection on unobserved variables that we are willing to make. We consider this as an extension of the procedures described in Andrews and Oster (2019).

3.2.3 From the LATE to the Sample LATE

Brinch et al. (2017) and Kowalski (2018) examine what we might learn about the ATE within an analysis sample, given that we can estimate a LATE. With the objective of providing tests for whether the LATE generalizes to the remaining sample with minimally restrictive assumptions, Brinch et al. (2017) introduces two assumptions. They assume that outcomes in each potential state of the world are weakly monotonically related to the probability of treatment. Such an assumption follows by maintaining a Roy model of economic decision-making. For instance, those who have a high likelihood of participating in a retention program have a high chance of persisting in college relative to those who have a low probability of entering the program, in the state of the world in which both participate in the program. Conversely, those who have a low likelihood of participating in the program may have a relatively high chance of persisting in college relative to likely program entrants, in the state of the world in which both do not enter the program. While this example fits the economic theory, Brinch et al. (2017) make no restriction on the direction of potential outcomes as the probability of treatment, only that it is monotonic.

Kowalski (2018) applies those assumptions to place bounds on the sample ATE. Consider the case of a binary treatment and binary instrument for either an experimental or quasi-experimental setting. Let $T = N, C, A$ be the individual types, never-takers, compliers, and always-takers. Without further assumptions, we can say that the probability of treatment is lowest for the never-takers and highest for the always-takers with the compliers lying between. Let $E[Y_1|D = 1, Z = 0] = E[Y_1|T = A]$ be the observed potential outcome under treatment of always takers, and $E[Y_0|D = 0, Z = 1]$ be the observed potential outcome under no treatment of never takers. We can estimate the potential outcome under treatment for those who comply with the exogenous variation using the share of compliers among those for whom $D = 1$ and $Z = 1$, $p(T = C|D = 1, Z = 1)$ which we estimate in the first stage, and $E[Y_1|D = 1, Z = 1]$, which we observe in the data. Thus,

$$E[Y_1|T = C] = \frac{E[Y_1|D = 1, Z = 1] - (1 - p(T = C|D = 1, Z = 1))E[Y_1|T = A]}{p(T = C|D = 1, Z = 1)}. \quad (3.4)$$

Likewise, we can construct the expected potential outcome in the absence of treatment for those who comply with the exogenous variation using the share of compliers among those for whom $D = 0$ and $Z = 0$, $p(T = C||D = 0, Z = 0)$ which we estimate in the first stage, and $E[Y_0|D = 0, Z = 0]$, which we observe in the data:

$$E[Y_0|T = C] = \frac{E[Y_0|D = 0, Z = 0] - (1 - p(T = C|D = 0, Z = 0))E[Y_0|T = N]}{p(T = C|D = 0, Z = 0)}. \quad (3.5)$$

Suppose that we observe $E[Y_1|T = A] > E[Y_1|T = C] > E[Y_0|T = C] > E[Y_0|T = N]$. The $LATE = E[Y_1|T = C] - E[Y_0|T = C] > 0$. Under the assumption that outcomes in each potential state of the world are weakly monotonically related to the probability of treatment, $E[Y_1|T = N]$ (which we do not observe) is bounded above by $E[Y_1|T = C]$, such that the maximum treatment effect among the never-takers is $E[Y_1|T=C]-E[Y_0|T=N]$. Similarly, $E[Y_0|T=A]$ is bounded below by $E[Y_0|T=C]$, such that the maximum treatment effect for the always-takers is given by $E[Y_1|T=A]-E[Y_0|T=C]$. The maximum ATE for the sample is then given by the weighted sum of the LATE and the maximum treatment effects of the always-takers and never-takers with the weights determined by the shares of the three populations within the data.

We may obtain further tightening if we allow the probability of treatment to vary continuously as a function of the observed variables and assume that potential outcomes are linearly related to the probability of treatment. However, this assumption is strong and even the weak monotonicity in potential outcomes over the discrete probability of treatment may lead us to reject external validity even when the differences in treated and untreated average outcomes are identical between populations within the sample.

It is not obvious how to apply the method to the population recognizing selection into the analysis sample. We do so, by estimating the probability of treatment

3.2.4 Our Approach

Let L denote the experiment participation status. We write

$$\begin{aligned}E_P(Y_1-Y_0) &= P(L=1)E_P(Y_1-Y_0|L=1)\\&\quad+P(L=0,D=1)E_P(Y_1-Y_0|L=0,D=1)\\&\quad+P(L=0,D=0)E_P(Y_1-Y_0|L=0,D=0)\end{aligned}$$

where by definition we have $E_S(Y_1-Y_0)=E_P(Y_1-Y_0|L=1)$.

We make three assumptions regarding the selection into the experiment and the selection into treatment. Assumptions 1 and 2 allow us to derive a simple bound for the population ATE while adding assumption 3 allows us to tighten the bound.

Assumption 1: Ignorability of noncompliance with randomization

$$E(Y|L=1,D=0)=E(Y_0|L=1) \text{ and } E(Y|L=1,D=1)=E(Y_1|L=1)$$

This assumption states that on average the potential outcomes in the experiment sample are the same conditional or unconditional on the treatment status. This assumption is reasonable when the rate of noncompliance in the experiment sample is low, or when tests on the selection into treatment in the experiment sample are rejected.

Assumption 2.1: Weakly monotonic selection of participation by the potential outcome

$$\begin{aligned} & E(Y_0|L=1,D=0) \geq E(Y_0|L=0,D=0) \\ \Longleftrightarrow & E(Y_1|L=1,D=0) \geq E(Y_1|L=0,D=0) \end{aligned}$$

and

$$\begin{aligned} & E(Y_0|L=1,D=1) \geq E(Y_0|L=0,D=1) \\ \Longleftrightarrow & E(Y_1|L=1,D=1) \geq E(Y_1|L=0,D=1) \end{aligned}$$

This is the key assumption we make to obtain bounds on the population ATE. It states that, on average, if the potential untreated outcome for those who participate in the experiment is no smaller than those who do not participate in the experiment, then the potential treated outcome for those who participate in the experiment should also be at least as large as those who do not participate in the experiment, and vice versa, and it true for both the treated and untreated students. This assumption rules out the case where in the groups of students who eventually do and do not take up the treatment, those who self-select into the experiment are the ones less likely to benefit or more likely to get hurt from the treatment. This assumption may be violated when selection into the experiment is made based on criteria that are different from the outcome we study. For example, in our application, if students choose to participate in the experiment without the intention to obtain a chance to enroll in the program and complete the first year more smoothly, but with the intention to improve their social network so that they can find a part-time job more easily, our assumption may not hold.

Proposition 1: Assuming that Assumption 1 and Assumption 2.1 hold. Then

$$U_B \geq E_P(Y_1 - Y_0) \geq L_B \tag{3.6}$$

where L_B and U_B are two constants that can be calculated from the data. See Appendix C.1 for proof.

Proposition 1 is the main result of this paper. Although there is no guarantee that the bounds are tight and informative in all settings, they rely on weak assumptions and are applicable to many settings. The calculation is arithmetically extremely simple. We encourage researchers running RCT to calculate those simple bounds as long as the data permits and the assumptions are likely to hold.

The ignorability of noncompliance with randomization assumption is sometimes too strong to be made. We consider an alternative assumption.

Assumption 2.2: Weakly monotonic selection into noncompliance by the potential outcome

$$E(Y_0|AT) \geq E(Y_0|C) \Longleftrightarrow E(Y_1|AT) \geq E(Y_1|C)$$

and

$$E(Y_0|NT) \geq E(Y_0|C) \Longleftrightarrow E(Y_1|NT) \geq E(Y_1|C)$$

Assumption 2.2 is in the similar spirit to Assumption 2.1. It states that the potential outcomes are monotonic in the decision of noncompliance with the random assignment, and is less restrictive than Assumption 2.1.

Proposition 2: Assuming that Assumption 2.1 and Assumption 2.2 hold. Then

$$U_B \geq E_P(Y_1 - Y_0) \geq L_B \tag{3.7}$$

where L_B and U_B are two constants that can be calculated from the data. See Appendix C.2 for proof.

Based on either set of the assumptions, we provide bounds for the population ATE. If the estimated effects local to the experiment sample or the compliers in the randomization are lying outside those bounds, it is indicative that the effects identified in the RCT may not be generalized. It is then crucial to distinguish which parameter is most relevant to inform policy.

3.3 Application

We use as an application the study of the effectiveness of a first-year learning program on college retention at a large, selective, four-year, public institution, as studied in Azzam et al. (2022). Such programs are common in four-year institutions in the United States with the stated intended purpose often to increase college retention. This setting has many attractive features. First, it is a large RCT with over 1,500 students entering the RCT and the randomization appears to be carried out effectively. Second, the data consists of the entire freshman class over 8,500, which we define as the population of interest. Third, we can see students' decisions to enter the RCT, their decision to comply with the randomization, and their decision about whether to receive treatment in the absence of randomization with over 100 students receiving treatment without receiving randomization. Lastly, as the program was scaled up such that 90 percent of subsequent cohorts also received treatment the ATE was of direct policy importance.

Azzam et al. (2022) finds that there is no statistically significant effect of the program on the retention rates among the compliers in the experiment sample. Further, they find that the experiment sample is negatively selected based on observed variables and positively selected on unobserved variables. It is hence unlikely that the bias caused by selection due to observed variables and unobserved variables have the same direction, as the method proposed in Andrews and Oster (2019) will need to assume. Moreover, Figure 3.1 shows the average outcomes among the treated and untreated students who are and are not in the experimental sample. We observe that the average treated outcome in the experiment sample is higher than the average untreated outcome, while the average untreated outcome in the non-experimental sample is higher than the average treated outcome. The assumption made in Kowalski (2018) that the potential outcomes are monotonically related to the probability of treatment is therefore also under question.

In the need for an understanding of the population ATE that is of policy relevance, and in the absence of an existing method that could provide reasonable results, we apply the method we propose. The assumption we make that the noncompliance is ignorable in the experiment sample does not hold trivially. In panel A of Table 3.1, we show that the issue of selection within the experimental sample (or noncompliance) seems to be of minor concern, using the tests proposed in Black et al. (2017). We are hence more confident with applying our method to this setting and compare the results to those from the (extended) methods in Andrews and Oster (2019) and in Kowalski (2018).

3.3.1 Results

Table 3.2 reveals the estimated intent to treat estimates (ITT) of the treatment assignment on retention in panel A, the estimated LATE of the FYLC program in panel B, and panel C shows the first stage estimates of the randomized assignment of treatment on program take-up. It reveals a strong first stage in which treatment assignment increases the likelihood a participant enters the FYLC by 68 percentage points. However, neither the ITT nor the LATE reveals any statistically significant impact of the program, though the LATE point estimates of around 3 percentage points would be meaningful, particularly if those results were to scale to the remaining freshman class.

In Table 3.3, we present the bounds calculated using different methods. Column (1) shows the Manski's bounds with no assumption imposed. The upper bound is 11.9

percentage points and the lower bound is -68.9 percentage points. Unsurprisingly, the resulting bounds are wide barely informative. The original method proposed by Andrews and Oster (2019) can only be used to bound the parameter of the population intent to treat, which is between 4.4 to 9.6 percentage points, as is shown in column (2). In column (3), we apply the extended version of the Andrews and Oster (2019) by adding the assumption of ignorability of noncompliance in the experiment sample. The bounds for the population ATE are 6.8 and 14.8 percentage points. The AO method indicates a large and positive population effect, despite the statistically insignificant effect of 3 percentage points found for the experiment sample. In column (4), we apply the Kowalski (2018) method to extend from the LATE to the experiment sample ATE. The bounds for the experiment sample ATE are 2.5 and 17.2 percentage points. We then extend the Kowalski (2018) method by adding the assumption that the potential outcomes are weakly monotonically related to the probability of participation in the experiment. The implied bounds for the population ATE are 4.6 and 14.8 percentage points under this added assumption. Similar to the Andrews and Oster (2019) method, the Kowalski (2018) method also indicates a positive and potentially large treatment effect for the entire freshman student population. In column (6), we use Assumption 1 and Assumption 2.1, and find an upper bound of the population ATE of 5.6 percentage points and a lower bound of -68.9 percentage points.[1] While the bounds we produce are also wide, the upper bound restricts the possibility of a large positive population effect, in contrast to the other two methods. Also, there is little overlap between the bounds for the population ATE suggested by the Andrews and Oster (2019) method or the Kowalski

[1]The lower bound coincides with the Manski's bound because based on our assumptions and the data we have, we are effectively only bounding the upper end of the ATE. Under a different setting and dataset, our bounds could be different from the Manski's bounds on both ends.

(2018) methods and ours. In column (7), we use Assumption 2.1 and Assumption 2.2, which is less restrictive than the ignorability of noncompliance with randomization assumption. The implied population treatment effect becomes wider but is still bounded from above by 8.7 percentage points.

3.4 Conclusion

In this paper, we tackle the issue of external validity in RCTs due to their policy relevance in many situations. By utilizing a special setting where some individuals not participating in the experiment self-selected to receive treatment and their outcomes were observed by the researcher, we propose a method of calculating bounds for the parameter of population ATE, relying on very weak assumptions. While such a setting is less common, our proposed method is still valuable. We view our method as a complement to existing methods in the literature to be applied to different cases. As we have illustrated in the paper, in the setting of our application that studies the effect of a first-year learning program on retention rates, the assumptions made in other existing methods are unlikely to hold. Indeed, the bounds we obtain using our method are very different from the bounds obtained using other methods. As a future step, we need to explore ways to tighten the bounds.

Table 3.1: Selection within and into the analysis sample

Panel A: Testing selection within the experimental sample				
	(1)	(2)	(3)	(4)
Won	0.009	0.005	0.000	0.005
	(0.025)	(0.029)	(0.026)	(0.029)
Same sign p-value		0.852		0.999
Observations	803	762	803	762
Controls	No	No	Yes	Yes
Sample	C + NS	T + CO	C + NS	T + CO
Treatment status	Untreated	Treated	Untreated	Treated

Panel B: Testing selection into the experiment among the untreated				
	(1)	(2)	(3)	(4)
Lottery	0.028**	0.067*	0.039***	0.063*
	(0.011)	(0.034)	(0.011)	(0.033)
Same sign p-value		0.051		0.062
Observations	7252	879	7252	879
Controls	No	No	Yes	Yes
Sample	C + NS	T + CO	C + NS	T + CO
	+ NET	+ NET	+ NET	+ NRT
Treatment status	Untreated	Treated	Untreated	Treated

Table 3.2: RCT estimates

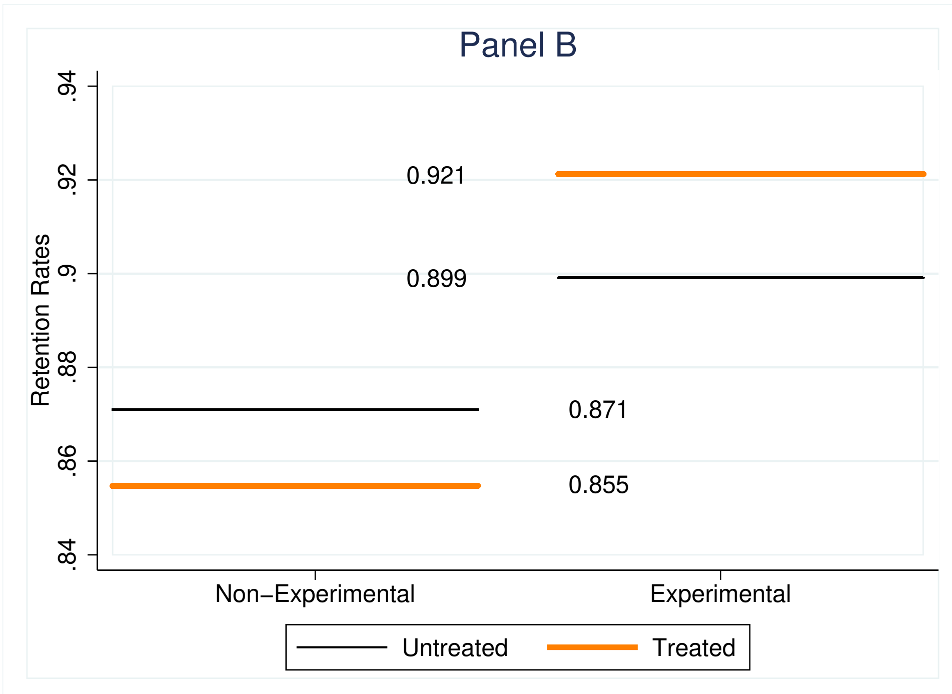
	(1)	(2)
Panel A: ITT on retention		
Won lottery	0.019	0.018
	(0.015)	(0.015)
Panel B: LATEs of FYLC on retention		
FYLC	0.029	0.027
	(0.022)	(0.022)
Panel C: First stage on FYLC		
Won lottery	0.648***	0.648***
	(0.019)	(0.019)
Observations	1565	1565
Retention Mean	0.91	0.91
Controls	No	Yes

Table 3.3: Bounds for population parameter by existing methods and our proposed method

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Manski	AO	AO*	K	K*	MZ1	MZ2
	ATE	ITT	ATE	SATE	ATE	ATE	ATE
Upper bound	0.119	0.096	0.148	0.172	0.148	0.056	0.087
Lower bound	-0.689	0.044	0.068	0.025	0.046	-0.689	-0.689

* Indicates extensions of the published methods.

Figure 3.1: Retention rates among the experimental and non-experimental populations by treatment status.



Chapter 4

Conclusions

This dissertation studies higher education effectiveness and its relationship with budget spending decisions. The first two chapters are motivated by the observation of changes in the composition of college spending decisions. In the first chapter, I investigate whether higher spending on educational inputs is effective in promoting student earnings. The empirical results suggest positive effects that are smaller and more precisely estimated than existing estimates in the literature. Several factors may contribute to the differences in the estimates. First, my data covers almost all institutions in the United States while earlier studies rely on data that are usually specific to a state or a school system. Second, the panel structure of my data allows me to control the unobserved institution effects that are likely to be correlated with both the spending decisions and student earnings. I further analyze the heterogeneity of the effects and find that the effects are mainly driven by private institutions and four-year institutions. In addition, the effects are likely to vary over the time dimension. The difference in the estimates between what I find and what is in the literature is not just numerical but also leads to vastly different policy implications concerning the cost-benefit analysis. My estimates suggest that if we view an increase in educational inputs as a monetary investment and the resulting higher student earnings as its return, students will need to work for over 40 years before such an investment is paid off, while the higher estimates in the literature require just less than 20 years.

In the second chapter, we investigate what drives the budget decision on instructional spending when it is an (in)effective educational input. In particular, we analyze whether competition for good students between colleges can explain the decision-making process. We construct a simple theoretical model where students value both the consumption of amenities and human capital production, and colleges value the outcome of their students that depends on the outcomes of the human capital production process. Colleges make a trade-off between spending more on educational inputs so that they better promote the human capital of the student body they admit, and spending more on consumption amenities so that they can attract a student body with higher endowments of human capital or learning abilities. Simulation results show that colleges respond positively to their competitors' decisions over the range where budget choices are commonly made, and such responsiveness declines as the competitive pressure from its competitors declines. Empirical evidence generally supports our predictions. We find that in terms of share of spending on educational inputs, a 10 percentage point increase by competitors is associated with a 4-8 percentage point increase. In terms of the absolute amount of spending on educational inputs, we find an estimated elasticity between 5 to 18 percent between a college and its competitors. The estimated elasticity is lowest for Doctoral universities and highest for special focus two-year colleges. We also find that the responsiveness declines as the distance between the colleges and their competitors increases.

In the third chapter, we generalize causal parameters that are identified through randomized control trials and are valid internally for the experiment sample to a more general population of interest. This is a crucial step in applied research work in determining whether we want to expand a small-scale program to its large-scale operations. The method we propose relies on a set of weak assumptions that are likely to hold in a general case. Our approach complements several other existing methods in the literature that aim at the same goal and depend on different sets of assumptions. With an application to evaluate the effectiveness of a first-year learning program on retention rates, our method suggests that despite a positive treatment effect found for the experimental sample, the average treatment effect for the entire freshmen student body is bounded from having a large positive effect, and may likely to have a negative one. We also show that assumptions required for two other methods are unlikely to hold in this particular setting, highlighting the importance of having an additional tool at hand.

Bibliography

Allcott, H. (2015). Site selection bias in program evaluation. *The Quarterly Journal of Economics 130*(3), 1117–1165.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of political economy 113*(1), 151–184.

Andrews, I. and E. Oster (2019). A simple approximation for evaluating external validity bias. *Economics Letters 178*, 58–62.

Azzam, T., M. D. Bates, and D. Fairris (2022). Do learning communities increase first year college retention? evidence from a randomized control trial. *Economics of Education Review 89*, 102279.

Behrman, J. R., M. R. Rosenzweig, and P. Taubman (1996). College choice and wages: Estimates using data on female twins. *The Review of Economics and Statistics*, 672–685.

Belfield, C. R. and H. M. Levin (2002). The effects of competition between schools on educational outcomes: A review for the united states. *Review of Educational research 72*(2), 279–341.

Black, D., J. Joo, R. LaLonde, J. A. Smith, and E. Taylor (2017). Simple tests for selection: Learning more from instrumental variables.

Black, D. A. and J. A. Smith (2006). Estimating the returns to college quality with multiple proxies for quality. *Journal of labor Economics 24*(3), 701–728.

Bloom, H. S., C. Michalopoulos, and C. J. Hill (2005). Using experiments to assess nonexperimental comparison-group methods for measuring program effects.

Brinch, C. N., M. Mogstad, and M. Wiswall (2017). Beyond late with a discrete instrument. *Journal of Political Economy 125*(4), 985–1039.

Cook, T. D., W. R. Shadish, and V. C. Wong (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management 27*(4), 724–750.

Cunha, F. and J. Heckman (2007). The technology of skill formation. *American economic review 97*(2), 31–47.

Dale, S. and A. B. Krueger (2011). Estimating the return to college selectivity over the career using administrative earnings data. Technical report, National Bureau of Economic Research.

Dale, S. B. and A. B. Krueger (2002). Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *The Quarterly Journal of Economics 117*(4), 1491–1527.

Deming, D. J. and C. R. Walters (2017). The impact of price caps and spending cuts on us postsecondary attainment. Technical report, National Bureau of Economic Research.

Epple, D., R. Romano, and H. Sieg (2006). Admission, tuition, and financial aid policies in the market for higher education. *Econometrica 74*(4), 885–928.

Finkelstein, A., S. Taubman, B. Wright, M. Bernstein, J. Gruber, J. P. Newhouse, H. Allen, K. Baicker, and O. H. S. Group (2012). The oregon health insurance experiment: evidence from the first year. *The Quarterly journal of economics 127*(3), 1057–1106.

Foster, E. and J. Rodgers (1980). Quality of education and student earnings. *Higher education 9*(1), 21–37.

Gale, D. and L. S. Shapley (1962). College admissions and the stability of marriage. *The American Mathematical Monthly 69*(1), 9–15.

Glazerman, S., D. M. Levy, and D. Myers (2003). Nonexperimental versus experimental estimates of earnings impacts. *The Annals of the American Academy of Political and Social Science 589*(1), 63–93.

Goering, J., J. Kraft, J. Feins, D. McInnis, M. J. Holin, and H. Elhassan (1999). Moving to opportunity for fair housing demonstration program: Current status and initial findings. *Washington, DC: US Department of Housing and Urban Development*.

Griffith, A. L. and K. N. Rask (2016). The effect of institutional expenditures on employment outcomes and earnings. *Economic Inquiry 54*(4), 1931–1945.

Grogger, J. (1996). School expenditures and post-schooling earnings: evidence from high school and beyond. *The Review of Economics and Statistics*, 628–637.

Hanushek, E. A. (1981). Throwing money at schools. *Journal of policy analysis and management 1*(1), 19–41.

He, Y., S. Sinha, and X. Sun (2021). Identification and estimation in many-to-one two-sided matching without transfers. *arXiv preprint arXiv:2104.02009*.

Hoekstra, M. (2009). The effect of attending the flagship state university on earnings: A discontinuity-based approach. *The Review of Economics and Statistics 91*(4), 717–724.

Hoxby, C. M. (2000). Does competition among public schools benefit students and taxpayers? *American Economic Review 90*(5), 1209–1238.

Hyman, J. (2017). Does money matter in the long run? effects of school spending on educational attainment. *American Economic Journal: Economic Policy 9*(4), 256–80.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Jackson, C. K., R. C. Johnson, and C. Persico (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *The Quarterly Journal of Economics 131*(1), 157–218.

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

Kowalski, A. E. (2016). Doing more when you're running late: Applying marginal treatment effect methods to examine treatment effect heterogeneity in experiments. Technical report, National Bureau of Economic Research.

Kowalski, A. E. (2018). How to examine external validity within an experiment. *Journal of Economics & Management Strategy*.

Lafortune, J., J. Rothstein, and D. W. Schanzenbach (2018). School finance reform and the distribution of student achievement. *American Economic Journal: Applied Economics 10*(2), 1–26.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *The American economic review*, 604–620.

MacLeod, W. B., E. Riehl, J. E. Saavedra, and M. Urquiola (2017). The big sort: College reputation and labor market outcomes. *American Economic Journal: Applied Economics 9*(3), 223–61.

Mountjoy, J. and B. R. Hickman (2021). The returns to college(s): Relative value-added and match effects in higher education. Technical report, National Bureau of Economic Research.

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

Papke, L. E. (2005). The effects of spending on test pass rates: evidence from michigan. *Journal of Public Economics 89*(5-6), 821–839.

Robins, G. B. (1973). Understanding the college budget.

Rothschild, M. and L. J. White (1993). The university in the marketplace: Some insights and some puzzles. In *Studies of supply and demand in higher education*, pp. 11–42. University of Chicago Press.

Rothschild, M. and L. J. White (1995). The analytics of the pricing of higher education and other services in which the customers are inputs. *Journal of political Economy 103*(3), 573–586.

Solomon, L. C. (1975). The definition of college quality and its impact on earnings. In *Explorations in Economic Research, Volume 2, number 4*, pp. 537–587. NBER.

Wachtel, P. (1976). The effect on earnings of school and college investment expenditures. *The Review of Economics and Statistics*, 326–331.

Webber, D. A. (2012). Expenditures and postsecondary graduation: An investigation using individual-level data from the state of ohio. *Economics of Education Review 31*(5), 615–618.

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

Weikart, D., A. Epstein, L. Schweinhart, and J. Bond (1978). The ypsilanti preschool curriculum demonstration project: Preschool years and longitudinal results. ypsilanti, mi: High.

Wooldridge, J. (2021). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Available at SSRN 3906345*.

Appendix A

Appendix for Chapter 2

A.1 Oster's Sensitivity Analysis

In this section, I describe the Oster's approach in more detail. Following the formulation in Oster (2019), suppose that the true relationship can be written as

$$\ln(Earning_{ic}) = \alpha + \beta \cdot \ln(InExp_{ic}) + \boldsymbol{W}_{1ic} + W_{2ic} + v_{ic} \tag{A.1}$$

The term W_{1ic} here corresponds to the collection of observed variables that are controlled in the regression. In my case, $\boldsymbol{W}_{1ic} = \boldsymbol{X_{ic}} \cdot \boldsymbol{\gamma} + \boldsymbol{\kappa_i} + \boldsymbol{\eta_c} + \boldsymbol{\theta_{s(i)}} \cdot c$. The term W_{2ic} represents a combination of all possible unobserved confounders. It also takes care of measurement errors, but not model misspecification. It is assumed that after controlling for W_2, we have $\mathbb{E}(v|X, W_1, W_2) = 0$.

Oster (2019) formalizes the sensitivity analysis first proposed in Altonji et al. (2005). In the framework she described, there are two parameters to be considered. The first parameter $\delta = \frac{\text{cov}(W_2,\ln(InExp))/\text{var}(W_2)}{\text{cov}(W_1,\ln(InExp))/\text{var}(W_1)}$ is conceptualized to be the degree of selection on unobserved confounders relative to the degree of selection on observed variables. The second parameter R_{max} is the maximum amount of variation in the outcome variable that could be explained when all possible confounding variables are controlled for. $R_{max} = 1$ when we assume $\text{var}(v|InExp, \boldsymbol{W}_1, W_2) = 0$. For given values of δ, R_{max}, and the estimation model, Oster (2019) established the relationship so that the coefficient adjusted for such a level of confounding factors $\tilde{\beta}$ can be calculated. Alternatively, fixing R_{max} and $\tilde{\delta} = 0$, I can calculate the corresponding value of δ, which is presented in Table 1.10.

A.2 Heterogeneous Effect by Cohort and Institution

Section 4 of Wooldridge (2021) provides a method to test whether the treatment effects are heterogeneous across units and time when applying two-way fixed effects estimation, which I briefly describe here. The test involves estimating the following two regression equations

$$y_{it} = \alpha + \mathbf{x_{it}}\beta + \bar{\mathbf{x}}_{\mathbf{i.}}\lambda + \bar{\mathbf{x}}_{\mathbf{.t}}\xi + (\bar{\mathbf{x}}_{\mathbf{i.}} \otimes \bar{\mathbf{x}}_{\mathbf{.t}})\pi + e_{it} \tag{A.2}$$

and

$$y_{it} = \alpha + \mathbf{x_{it}}\beta + \bar{\mathbf{x}}_{\mathbf{i.}}\lambda + \bar{\mathbf{x}}_{\mathbf{.t}}\xi + (\bar{\mathbf{x}}_{\mathbf{i.}} \otimes \bar{\mathbf{x}}_{\mathbf{.t}})\pi + \mathbf{x_{it}} \otimes (\bar{\mathbf{x}}_{\mathbf{i.}} - \bar{\mathbf{x}})\zeta_{\mathbf{1}} + \mathbf{x_{it}} \otimes (\bar{\mathbf{x}}_{\mathbf{.t}} - \bar{\mathbf{x}})\zeta_{\mathbf{2}} + e_{it} \tag{A.3}$$

where $\bar{x}_{i.} = \frac{1}{T}\sum_{t=1}^{T} x_{it}$, $\bar{x}_{.t} = \frac{1}{N}\sum_{i=1}^{N} x_{it}$, and $\bar{x} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}$. Testing the null $H_0 : \pi = 0$ in Equation A.2 is a simple test of whether having additive unit and time effects is sufficient, and testing the null $H_0 : \zeta_1 = 0$ and/or $H_0 : \zeta_2 = 0$ in Equation A.3 tests for heterogeneous slopes across the unit and/or the time dimension. I apply the above method to my main estimation model with a subset of the control variables, because otherwise with all the control variables and the interaction terms, the number of regressors explodes. I use the five control variables that are most correlated with instructional spending and student earnings at the same time, or to put in other words, the control variables that will cause the greatest bias if omitted. The five variables are the average age at entry, the share of female students, the average SAT score, the average family income, and the median household income.

Table A.1 shows the results for the above exercises. Overall, the results suggest that the additive structure of the institution and cohort fixed effects are sufficient and the interaction terms of the institution-specific cohort averages and the cohort-specific institution averages are not necessary. In addition, there is little evidence of heterogeneous effects across institutions but some evidence of heterogeneous effects across cohorts. Note that the estimated coefficients are much larger because only five control variables are included, for the purpose to examine effect heterogeneity.

To further examine the heterogeneous effects across the time dimension, I interact the cohort indicators with the instructional spending. Table A.2 presents the results for earnings six years after entry.[1] The effects are small and even negative before the 2000 cohort and are positive and increasingly larger in later cohorts. Such a pattern of effect heterogeneity across cohorts is consistent with the diminishing return to educational input as institutions have been lowering their inputs on instructional spending.

[1]The very similar pattern is observed for earnings eight and ten years after entry, and hence not presented.

Table A.1: Test for Heterogeneous Effects Across Institutions and Cohorts

	(1) Mean	(2) Pct25	(3) Pct50	(4) Pct75
Log earnings six years after entry	0.028***	0.056***	0.031***	0.022***
	(0.0055)	(0.0092)	(0.0060)	(0.0051)
Number of Institutions	4808	4808	4808	4808
p-value for $H_0 : \pi = 0$	0.92	0.69	0.43	0.77
p-value for $H_0 : \zeta_1 = 0$	0.90	0.46	0.57	0.96
p-value for $H_0 : \zeta_2 = 0$	0.00	0.00	0.01	0.07
Log earnings eight years after entry	0.048***	0.076***	0.054***	0.043***
	(0.0041)	(0.0065)	(0.0044)	(0.0039)
Number of Institutions	4402	4402	4402	4402
p-value for $H_0 : \pi = 0$	0.23	0.05	0.04	0.19
p-value for $H_0 : \zeta_1 = 0$	0.96	0.45	0.37	0.95
p-value for $H_0 : \zeta_2 = 0$	0.00	0.00	0.06	0.15
Log earnings ten years after entry	0.044***	0.067***	0.050***	0.039***
	(0.0034)	(0.0056)	(0.0038)	(0.0031)
Number of Institutions	4069	4069	4069	4069
p-value for $H_0 : \pi = 0$	0.20	0.38	0.35	0.27
p-value for $H_0 : \zeta_1 = 0$	0.09	0.05	0.01	0.07
p-value for $H_0 : \zeta_2 = 0$	0.01	0.01	0.13	0.22

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.2: Heterogeneous Effects Across Cohorts

Depvar: Log earnings six years after entry	(1) Mean	(2) Pct25	(3) Pct50	(4) Pct75
cohort=1996 × ln(InExp/FTE)	-0.023***	-0.015	-0.017**	-0.022***
	(0.0052)	(0.0079)	(0.0055)	(0.0049)
cohort=1998 × ln(InExp/FTE)	-0.011**	-0.00075	-0.0065	-0.011**
	(0.0043)	(0.0068)	(0.0046)	(0.0042)
cohort=2000 × ln(InExp/FTE)	0.012**	0.034***	0.017***	0.0063
	(0.0045)	(0.0072)	(0.0048)	(0.0042)
cohort=2002 × ln(InExp/FTE)	0.014***	0.010	0.0097*	0.011**
	(0.0036)	(0.0057)	(0.0038)	(0.0035)
cohort=2004 × ln(InExp/FTE)	0.0094*	0.016**	0.0074	0.0053
	(0.0038)	(0.0061)	(0.0040)	(0.0036)
cohort=2005 × ln(InExp/FTE)	0.025***	0.048***	0.024***	0.021***
	(0.0047)	(0.0076)	(0.0050)	(0.0045)
cohort=2006 × ln(InExp/FTE)	0.046***	0.077***	0.046***	0.043***
	(0.0062)	(0.011)	(0.0067)	(0.0059)
cohort=2007 × ln(InExp/FTE)	0.054***	0.072***	0.052***	0.052***
	(0.0086)	(0.013)	(0.0091)	(0.0084)
Number of Institutions	4721	4721	4721	4721

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix B

Appendix for Chapter 3

B.1 Heterogeneous Effect by Cohort and Institution

We apply the Wooldridge (2021) test for heterogeneous response across time and institution for our estimation equation 2.12. We estimate the following two regression equations

$$y_{j,t,d} = \alpha + x_{-j,t,d}\beta + \bar{x}_{-j,.,d}\lambda + \bar{x}_{.,t,.}\xi + (\bar{x}_{-j,.,d} \otimes \bar{x}_{.,t,.})\pi + \varepsilon_{j,t,d} \tag{B.1}$$

and

$$\begin{aligned} y_{j,t,d} =& \alpha + \mathbf{x_{-j,t,d}}\beta + \mathbf{\bar{x}_{-j,.,d}}\lambda + \mathbf{\bar{x}_{.,t,.}}\xi + (\mathbf{\bar{x}_{-j,.,d}} \otimes \mathbf{\bar{x}_{.,t,.}})\pi \\ &+ \mathbf{x_{it}} \otimes (\mathbf{\bar{x}_{-j,.,d}} - \mathbf{\bar{x}})\zeta_{\mathbf{1}} + \mathbf{x_{it}} \otimes (\mathbf{\bar{x}_{.,t,.}} - \mathbf{\bar{x}})\zeta_{\mathbf{2}} + \varepsilon_{\mathbf{j,t,d}} \end{aligned} \tag{B.2}$$

where $y_{j,t,d} = \ln(E_{j,t,d})$, $\mathbf{x_{-j,t,d}} = (\ln(\mathbf{E_{-j,t,d}}), \ln(\mathbf{E_{-j,t,d}}) * \mathbf{d})$, $\mathbf{\bar{x}_{-j,.,d}} = \frac{\mathbf{1}}{\mathbf{T}} \sum_{\mathbf{t=1}}^{\mathbf{T}} \mathbf{x_{-j,t,d}}$, and $\mathbf{\bar{x}_{-j,.,d}} = \frac{\mathbf{1}}{\mathbf{J \times D}} \sum_{\mathbf{j=1}}^{\mathbf{J}} \sum_{\mathbf{d=1}}^{\mathbf{D}} \mathbf{x_{-j,t,d}}$.

Table B.1 presents the estimation results for the above equations. While the estimated coefficients remain mostly qualitatively the same, their magnitudes have altered slightly. The results of the test suggest the responses to competitors' budget changes are likely to be heterogeneous across both the time dimension and the institution dimension. It remains to be explored the factors contributing to those heterogeneities which are currently not modeled.

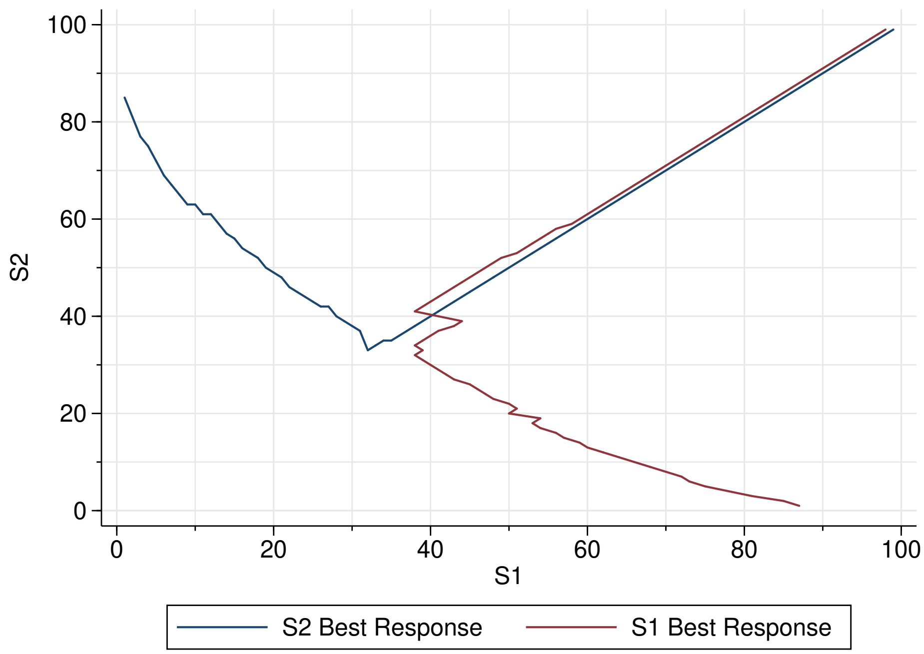
Table B.1: Test for Heterogeneous Effects

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A (Equation B.1)						
Ln(Avg_InExp)	0.20***	0.20***	0.14***	0.19***	0.21***	0.08***
	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.01)
Ln(Avg_InExp)	-0.31***	-0.27***	0.41***	0.15**	-0.36***	-0.25***
×Distance (1,000 miles)	(0.03)	(0.05)	(0.08)	(0.06)	(0.09)	(0.03)
p-value for $H_0 : \pi = 0$	0.00	0.00	0.00	0.00	0.00	0.00
Panel B (Equation B.2)						
Ln(Avg_InExp)	0.16***	0.18***	0.05***	0.09***	0.16***	0.06***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)
Ln(Avg_InExp)	-0.21***	-0.19***	0.01	-0.11***	-0.25***	-0.10**
×Distance (1,000 miles)	(0.02)	(0.04)	(0.02)	(0.03)	(0.07)	(0.03)
p-value for $H_0 : \zeta_1 = 0$	0.00	0.00	0.00	0.00	0.00	0.00
p-value for $H_0 : \zeta_2 = 0$	0.00	0.00	0.00	0.00	0.00	0.00
Observations	1075744	248611	422652	670528	642949	716366

Cluster-robust standard errors in parentheses.
Column (1): Associate's Colleges
Column (2): Special Focus Two-Year
Column (3): Doctoral Universities
Column (4): Master's Colleges and Universities
Column (5): Baccalaureate Colleges
Column (6): Special Focus Four-Year
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

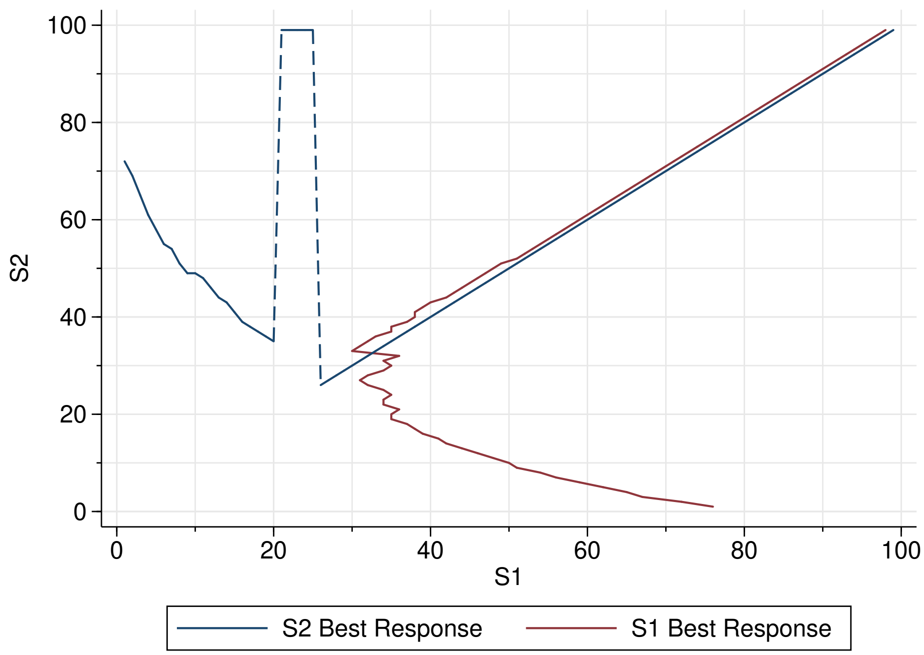
B.2 Robustness Appendix

Figure B.1: Best response functions between two colleges, positive correlation between π_i and H_{i1}



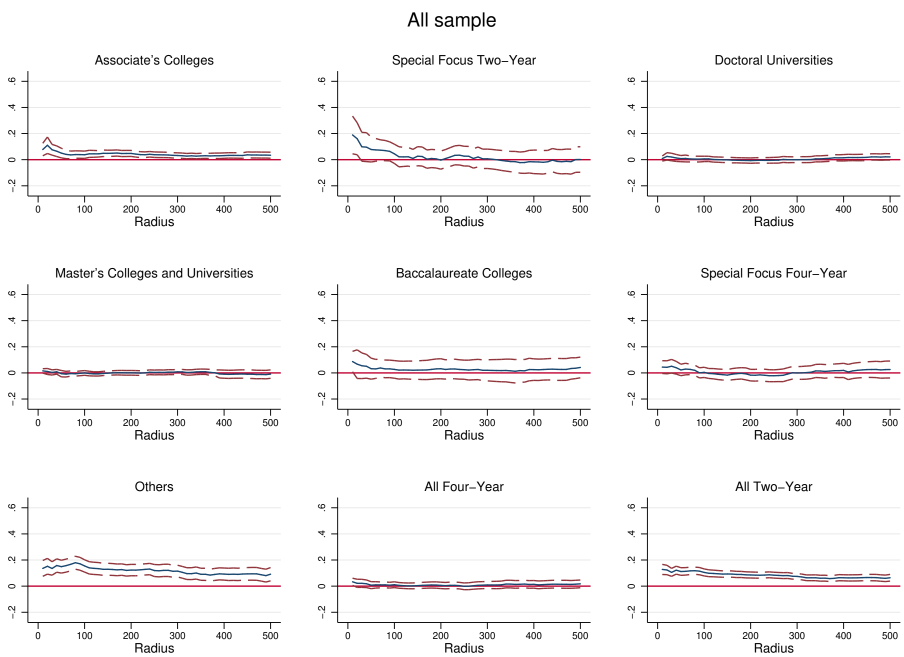
Notes: This figure plots the best response functions of two colleges in setting their share of spending on educational inputs s_j to each other's decisions. Characteristics for college 1 is technology in human capital production $\zeta_1 = 1.1$, total spending per student $B_1 = 8010$, and capacity $N_1 = 800$. Characteristics for college 2 is $\zeta_2 = 1.101$, $B_2 = 8000$, and $N_2 = 1,500$. Total number of students is $N = 10,000$ and is drawn from the distribution in Equation 2.7 with $\rho_{23} = 0.5$.

Figure B.2: Best response functions between two colleges, negative correlation between π_i and H_{i1}



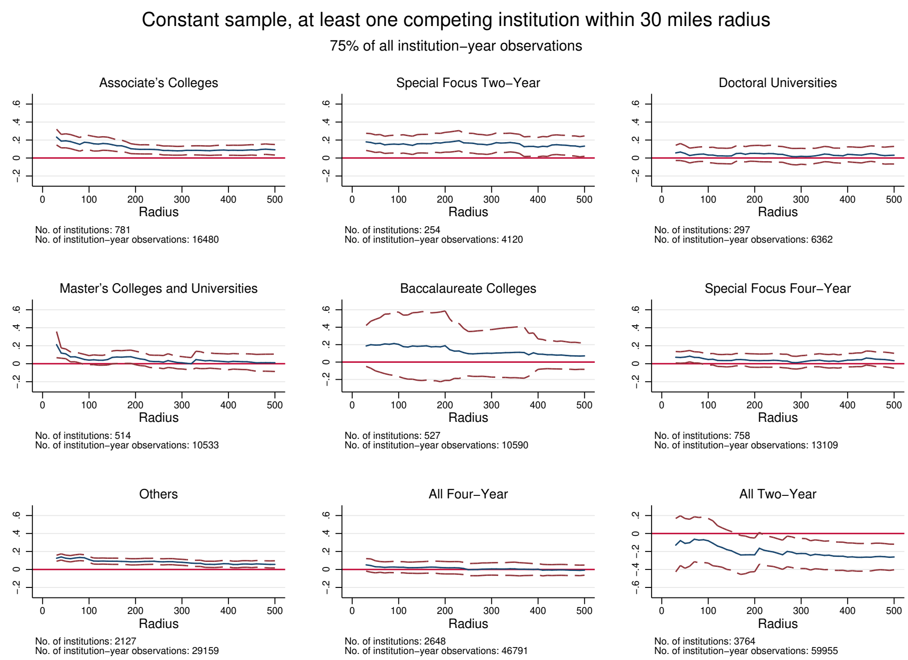
Notes: This figure plots the best response functions of two colleges in setting their share of spending on educational inputs s_j to each other's decisions. Characteristics for college 1 is technology in human capital production $\zeta_1 = 1.1$, total spending per student $B_1 = 8010$, and capacity $N_1 = 800$. Characteristics for college 2 is $\zeta_2 = 1.101$, $B_2 = 8000$, and $N_2 = 1,500$. Total number of students is $N = 10,000$ and is drawn from the distribution in Equation 2.7 with $\rho_{23} = -0.5$.

Figure B.3: Response to change in spending on educational inputs by competing institutions



Notes: This figure plots the estimated coefficients and the 95% confidence intervals as responses of a college to changes in its competing institutions' spending on educational inputs. Competing institutions are defined as institutions located within the specified miles radius.

Figure B.4: Response to change in spending on educational inputs by competing institutions



Notes: This figure plots the estimated coefficients and the 95% confidence intervals as responses of a college to changes in its competing institutions' spending on educational inputs. Competing institutions are defined as institutions located within the specified miles radius and of the same Carnegie classification. The sample is restricted to colleges with at least one competing institution within 30 miles radius so that the analysis sample is constant as radius changes in each subgraph.

Appendix C

Appendix for Chapter 4

C.1 Proof of Proposition 1:

Without lose of generality, suppose in the data, we observe $E(Y|L=0,D=1) < E(Y|L=1,D=1)$ and $E(Y|L=0,D=0) > E(Y|L=1,D=0)$.

Given that $E(Y|L=0,D=1) < E(Y|L=1,D=1)$, we have $E(Y_1|L=0,D=1) < E(Y_1|L=1,D=1)$. By assumption 2, we have $E(Y_0|L=0,D=1) \leq E(Y_0|L=1,D=1)$. By assumption 1, $E(Y_0|L=1,D=1) = E(Y_0|L=1,D=0)$. Therefore, we have $0 \leq E(Y_0|L=0,D=1) \leq E(Y_0|L=1,D=0)$.

Similarly, when we observe in the data that $E(Y|L=0,D=0) > E(Y|L=1,D=0)$, we have $E(Y_0|L=0,D=0) > E(Y_0|L=1,D=0)$. By assumption 2, we have $E(Y_1|L=0,D=0) \geq E(Y_1|L=1,D=0)$. By assumption 1, $E(Y_1|L=1,D=0) = E(Y_1|L=1,D=1)$. Therefore, we have $E(Y_1|L=1,D=1) \leq E(Y_0|L=0,D=1) \leq 1$.

The population ATE can be written as

$$
\begin{aligned}
E(Y_1 - Y_0) = P(L=1)E(Y_1 - Y_0|L=1) \\
+ P(L=0, D=1)E(Y_1 - Y_0|L=0, D=1) \\
+ P(L=0, D=0)E(Y_1 - Y_0|L=0, D=0)
\end{aligned}
$$

Substitute the upper bounds and lower bounds of the two terms $E(Y_0|L = 0, D = 1)$ and $E(Y_0|L = 0, D = 1)$ that was derived above. All other terms are observed in the data. Hence we obtain the upper bound and lower bound for $E(Y_1 - Y_0)$.

C.2 Proof of Proposition 2:

Without lose of generality, suppose in the data, we observe $E(Y|L = 0, D = 1) < E(Y|L = 1, D = 1)$, $E(Y|L = 0, D = 0) > E(Y|L = 1, D = 0)$, $E(Y|L = 1, D = 1, Z = 0) > E(Y_1|L = 1, C)$, and $E(Y|L = 1, D = 0, Z = 1) > E(Y_0|L = 1, C)$, where $E(Y_1|L = 1, C)$ and $E(Y_0|L = 1, C)$ are two terms that can be calculated from the data.

By Assumption 2.2, since $E(Y_1|AT) = E(Y|L = 1, D = 1, Z = 0) > E(Y_1|L = 1, C)$, we have $E(Y_0|AT) > E(Y_0|L = 1, C)$, and similarly, since $E(Y_0|NT) = E(Y|L = 1, D = 0, Z = 1) > E(Y_0|L = 1, C)$, we have $E(Y_1|NT) > E(Y_1|L = 1, C)$.

We can write

$$
\begin{aligned}
E(Y_0|L = 1, D = 1) = &P(Z = 1|L = 1, D = 1)[P(C|Z = 1, L = 1, D = 1)E(Y_0|C)+ \\
&P(AT|Z = 1, L = 1, D = 1)E(Y_0|AT)]+ \\
&P(Z = 0|L = 1, D = 1)E(Y_0|L = 1, D = 1, Z = 0)
\end{aligned}
$$

$$
\begin{aligned}
E(Y_1|L=1,D=0) = & P(Z=1|L=1,D=0)E(Y_1|NT)+ \\
& P(Z=0|L=1,D=0)(P(C|Z=0,L=1,D=0)*E(Y_1|C)+ \\
& P(NT|Z=0,L=1,D=0)*E(Y_1|NT))
\end{aligned}
$$

All terms in the above two expressions are either observed or bounded, so that the above two terms are also bounded.

By Assumption 2.1, we can calculate the upper bounds and lower bounds of the two terms $E(Y_0|L=0,D=1)$ and $E(Y_0|L=0,D=1)$ based on their values in relation to $E(Y_0|L=1,D=1)$ and $E(Y_1|L=1,D=0)$. Thus, under the same reasoning and calculation as in the proof of Proposition 1, we provide upper bound and lower bound for the term $E(Y_1-Y_0)$.