Raising Tuition Fee, Increasing Graduation Rate

Cameron Chang, Joy Jang
Econometrics I
December 20, 2023

Introduction

While income level has received a contentious amount of attention as it pertains to the availability of higher-level education level and its subsequent success in job search, and employment, the impact of tuition fees on higher-level education success has undergone the radar. This oversight prompts a necessary exploration into the nuanced relationship between tuition fees and educational success, aiming to provide a more comprehensive understanding of the multifaceted factors that contribute to equitable access and achievement in higher education.

In this paper, we will explore the effects of higher education tuition on the completion of a 4-year bachelor's program. Studying the correlation between increasing tuition fees and graduation rates holds significant implications for the accessibility and success of higher education. Elevated tuition costs can serve as a formidable barrier, particularly for students from lower-income backgrounds, potentially limiting their ability to pursue and complete a college education. This research sheds light on the intricate dynamics influencing student retention, as heightened financial strain may impede progress toward graduation.

Moreover, it allows policymakers to craft informed decisions regarding financial aid, funding, and education policies, fostering a balance between generating institutional revenue and ensuring equitable opportunities for diverse student populations. The investigation into tuition fees and graduation rates also addresses broader societal concerns, such as the impact of student debt on career choices, economic mobility, and workforce development. Ultimately, understanding this relationship is instrumental in creating an educational landscape that is both financially sustainable and conducive to the success of a diverse student body.

Survey of the Literature

A paper was published regarding the pricing at US public universities. According to Cook and Turner's "Progressivity of Pricing at US Public Universities," public research universities have increasingly adopted a high-tuition, high-aid pricing model. Between 2012 and 2018, the actual

amount students paid, known as net tuition, decreased more than expected based on the growth in state appropriations by using OLS regressions and IV regressions of tuitions and net tuition per student. During this period, although tuition levels continued to rise, the rate of increase was slower compared to previous years.

There have been multiple studies on income level on enrollment rate, as such, this paper will solely focus on the graduation aspect of the 4-year bachelor program. This paper aims to identify the primary mechanism/metrics associated with graduation rate according to Hemelt's "The Changing Landscape of Tuition and Enrollment in American Public Higher Education" graduation rates have had only a gradual incline over the last 30-40 years (Helmelt, 43).

Within this trend of increasing tuitions, using institutional data from US universities, we have identified a correlation between bachelor program completion and tuition fees, along sider factors such as race, gender, financial aid, and state among many others. Excluding enrollment rates, which through several research papers have been proven to correlate with bachelor completion rate, this paper aims to find the significance of other factors which will be expanded on later.

Data

The information was collected from a database of American Universities from IPEDS, a reputable source known for its information on post-secondary education. However, due to the large data set, certain metrics are missing due lack of reporting. Thus, many hypothesized significant variables were not included. Such examples include the percentage of foreign students, the number of students from each state, public and private sectors among others. Knowing this, the following study is on the effect of tuition fees on the graduation rate for a 4-year bachelor's degree program. Data collected included 1521 US universities, public and private, with data ranging from 2010-2013 depending on the metric.

Among the multitude of variable selections, the final selection of variables was selected through the following methodology. Foremost, as illustrated in Figure 1, a scatterplot was taken with said

variable against graduation rate, hence allowing us to check for a basic correlation as well as identify if any variables would be linearly or nonlinearly correlated. If said correlation was deemed significant, a multicollinearity test was taken to remove any doubt of undermining the statistical significance of our independent variable as seen in Figure 2. As seen in Appendix A, the variables were then classified into the types, which were either numeric or Alpha, with Alpha denoting binary variables. Additionally, the final condition was that each dependent variable has at least <1000 sample size to not skew our data set via unfair sample size comparisons.

Figure 1: Scatterplot of Average Tuition Fee against Graduation Rate 4

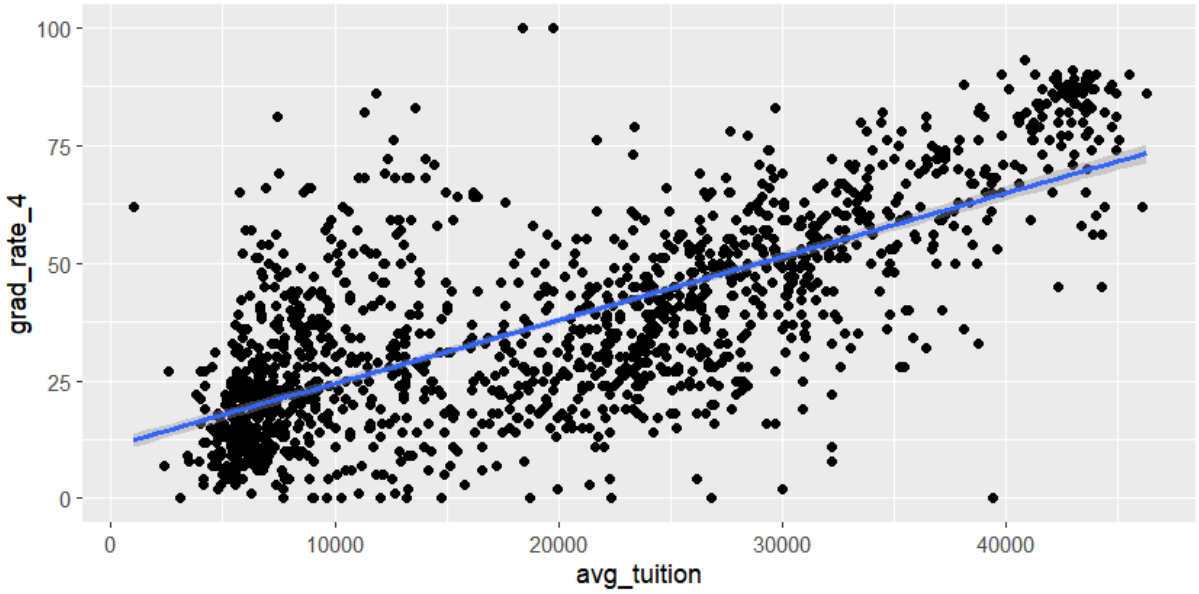


Figure 2: Variable Multi-Collinearity Test

	GVIF	Df	GVIF^(1/(2*Df))
avg_tuition	2.389975	1	1.545954
white	1.364719	1	1.168212
avg_act	2.582639	1	1.607059
women	1.142537	1	1.068895
highest_degree	1.432147	2	1.093948
full_enrollment	2.104382	1	1.450649
urbanization	1.340622	3	1.050069
international	1.355581	1	1.164294

Some of the most important variables included admission rate, race, highest degree as well as state and sector. We hypothesized that higher tuition fees would mean a higher socioeconomic status percentage of undergraduates, in which higher socioeconomic status is correlated with higher student grades. This is supported by a published paper titled "Test scores and educational opportunities: Panel evidence from five low- and middle-income countries" by Jishnu Das, Abhijeet Singh and Andres Yi Chang, where they state that "higher test scores at 12 are associated with higher college attendance at age 22", however there is an estimatedR^2 of 0.15 to 0.55 which that the association "explains approximately 15% to 55% of the SES GAP in completed years of schooling" (Das, 1).

Figure 3: Regession results (SE in parenthesis)

	Dependent variable:							
	grad_rate_4							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(intercept)	10.9624*	1.2406	-55.583*	-64.078*	-63.792*	-60.383*	-63.083*	-62.986*
	(0.7693)	(1.2033)	(2.3340)	(3.2044)	(3.2603)	(3.3080)	(3.3861)	(3.384)
avg_tuition	0.00137*	0.00131*	0.0008*	0.00078*	0.00075*	0.00086*	0.00087*	0.00088*
	(0.00003)	(0.00003)	(0.00003)	(0.00003)	(0.00003)	(0.00004)	(0.00004)	(0.00004)
white		0.17265*	0.6816*	0.07104*	0.06058*	0.07363*	0.0610*	0.0447*
		(0.0168)	(0.01538)	(0.01531)	(0.1559)	(0.0157)	(0.01603)	(0.0167)
avg_act			3.2107*	3.32687*	3.46154*	3.13116*	3.14423*	3.27840*
			(0.1152)	(0.11844)	(0.12465)	(0.14146)	(0.14108)	(0.1465)
women				0.10905*	0.12166*	0.12681*	0.13723*	0.12767*
				(0.02837)	(0.02834)	(0.02810)	(0.02820)	(0.02825)
highest_degree : Docter's					-3.9931*	-5.4942*	-4.8485*	-4.7636*
					(0.94876)	(0.9909)	(1.0007)	(0.9971)
highest_degree : Master's					-2.4842*	-2.7163*	-2.4576*	-2.3718*
					(0.95640)	(0.94893)	(0.9469)	(0.9436)
full_enrollment						0.00033*	0.00038*	0.0004*
						(0.00006)	(0.00007)	(0.00007)
urbanization : Rural							2.39871	2.35303
							(1.43921)	(1.43226)
urbanization : Suburb							3.1860*	3.0944*
							(0.7981)	(0.7954)
urbanization : Town							3.09750*	3.28014*
							(0.85524)	(0.8537)
international								-0.2749*
								(0.0850)
Observations	1475	1475	1195	1195	1195	1195	1195	1195
R^2	0.5226	0.5545	0.7306	0.7339	0.7379	0.7429	0.7474	0.7496
Adjusted R^2	0.5223	0.5539	0.7299	0.7333	0.7366	0.7413	0.7453	0.7473

RSE	15.17	14.66	10.98	10.92	10.85	10.75	10.67	10.62

Expanding on this paper, it is clear that socioeconomic status is vaguely connected to completed scores, however it isn't significantly correlated enough to be considered statistically significant. As such, we are inclined to think that there are other variables including race, sector, and state to have a high correlation with graduation rates, which are not directly associated with socioeconomic status. As such, the regression model illustrated in Figure 3 illustrates our results.

Furthermore, we assumed 'loan', a variable which denotes for the percentage of students receiving any sort of financial aid, could be a significant variable in the regression. The reason for this is that this metric serves as an essential indicator of the financial landscape faced by students pursuing higher education. A higher percentage of students receiving loans is often associated with increased financial strain, potentially impeding academic persistence and success. As students grapple with the demands of loan repayment and financial obligations, the allocation of time and resources for their studies becomes a critical consideration. Nevertheless, as illustrated in Appendix C, which presents a regression incorporating 'loan,' the estimated coefficient for the variable cannot be deemed statistically significant.

Model

The model used for this paper is a multiple linear regression model as none of the variables have a non-linear relationship as mentioned prior. Following the assumptions that data collected are each independent, have homoscedasticity, and are normally distributed, we plotted a model shown in Figure 3. Each independent variable was then plotted against the dependent variable of graduation rate in 4 years.

Empirical Application

As can be observed in Figure 3, model 1 and 2 are of little importance to us, as while the independent variables of average tuition and race white were significant, their respective R^2 is at a low 0.52 and 0.54 respectively, indicating that there is a correlation relationship, however nothing that can be considered incredible strong. Since the relative standard error is also significantly higher in models 1 and 2 compared to all the models that proceed with it. As such, the majority of the decisive relationship can be observed from model 3 to model 7 onwards. However, it has to be noted that, unlike our hypothesis, the dependent variable of average tuition has a very small coefficient of 0.0137 and 0.00131 respectively, which throughout all models, decreases as more variables are included. This implies that an increase in average tuition by $1 will increase the graduation rate by 0.00137 and 0.00131 percentage points respectively. This is decisive as despite average tuition being statistically significant with a test statistic greater than the critical value, as we look at later models, it is clear that other independent variables account for a large proportion of average tuition. This will be explained further when looking at model 3.

The transition between model 2 and model 3 is the most distinct throughout the Regression results. Despite the decrease in observations caused by included variables that had omitted certain metrics, model 3 has an increase in R^2 of 30% from 0.5545 to 0.7306 which this correlation would remain throughout all other models. This trend follows as the relative standard error also ranges from 10.98 to 10.62. In model 3, we have the inclusion of the average ACT dependent variable which is not only statistically significant but according to the model, it is suggested that the average ACT score accounts for affects graduation rate the most. Not only does it have one of the highest coefficient values in our data, ranging from 3.13116 to 3.46154, but also the increase from model 2 to model 3 increasing the R^2 the most in our regression model. Our suspected reason for this correlation is that higher ACT scores are correlated with higher GPA scores in university which leads to a higher percentage of students completing their bachelor's degree.

Looking from Model 3 to Model 8, whilst R^2 is bound to increase due to R^2 inflation, the Adjusted R^2 also increases by approximately a little less than 0.1 with each proceeding model. The variables that are the most significant indicated by their test statistical significance, but also

their coefficient value include, the highest degree offered by doctors and master's respectively, as well as the urbanization of the university's location. Other variables such as gender, and race, while being statistically significant don't impact graduation rate as much as other values.

Interesting to note that is the urbanization of suburb or town areas is around the same as indicated by the very similar coefficient values hovering around 3.09 and 3.28 with rural areas hovering around 2.39871 to 2.35303. Some of the geographical differences may provide more advantages such as resource availability or support availability which may include technological advantages or more research opportunities, however, it is to note that since suburbs and towns are very similar in coefficient which may mean that there are diminishing returns when it comes to geographical location. It is also important to note that the geography and urbanization of a region can be broken down into many other factors such as which include other variables such as demographics, race, and culture among others. Whilst we would have liked to have had concise the variables more specific variables, variables such as cultural or demographic impact tend to be qualitative than quantitative.

As for the highest degree awarded, we believe that the highest degree award is most likely due to more people being inclined to take on programs that combine master's or doctorate degrees with bachelor's degrees. As such factors such as an increase in academic rigor, and extended time to completion in conjunction with the financial burden of additional years in higher education may increase the number of students that would drop out. In our regression model, the highest degree awarded, doctorate and master's respectively, is the only independent variable to have a negative impact on the graduation rate.

Conclusion

Comparing our papers to that research by those in the past, the idea of there being a relationship between tuition fees is similar to the findings found regarding socioeconomic status and its relationship with graduation rates. While other studies tend to focus primarily on the combined effect of completion rate and enrollment rate, even in such studies, the maximum variance of

economic status that explained the completion rate was 55% which is very similar to our data collected from R^2 in model 1. This could potentially be a systematic issue rather than a specific individual variable issue. Data collected in this study in addition to other studies are solely focused on US universities and thus are not indicative of university success rate globally. While significant, the effect of tuition fees on graduation rates is too minute as compared to other factors such as average tuition and the highest degree awarded. While the ability to afford high-average tuition is connected to one's high socioeconomic status, all other studies suggest that there is no definite relationship between high monetary worth and graduation rates. It is far more likely that test score-related statistics such as ACT would have a more definitive relationship to graduation rate.

References

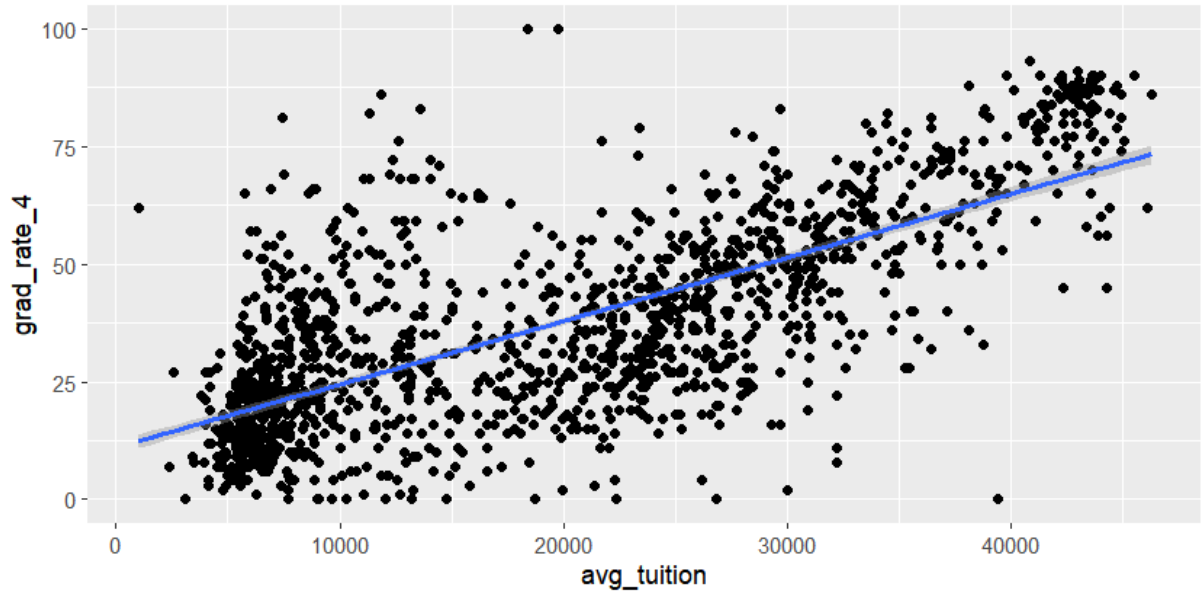
1. Cook, & Turner. (2022). *Progressivity of Pricing at US Public Universities*. National Bureau of Economic Research.
2. Das, Singh, & Chang. (2022). *Test scores and educational opportunities: Panel evidence from five low- and middle-income countries*. Sciencedirect.com. https://www.sciencedirect.com/science/article/pii/S0047272721002061?casa_token=4Z9Gg7tEQsQAAAAA:Xf6_X1s4m25D6nJbhxR4tlS4sEUH_9rG0c2PXGPvhN5YyF5xxS5RGZBEEiJDuKERsfExSJo
3. Helmet, & Marcotte. (2012). The Changing Landscape of Tuition and Enrollment in American Public Higher Education. *JSTOR*, *2*(1), XXVII-XXXVI. https://doi.org/10.1515/9781400843114.xxvii
4. IPEDS. (2017). *"American University data" IPEDS dataset*. Kaggle: Your Machine Learning and Data Science Community. https://www.kaggle.com/datasets/sumithbhongale/american-university-data-ipeds-dataset

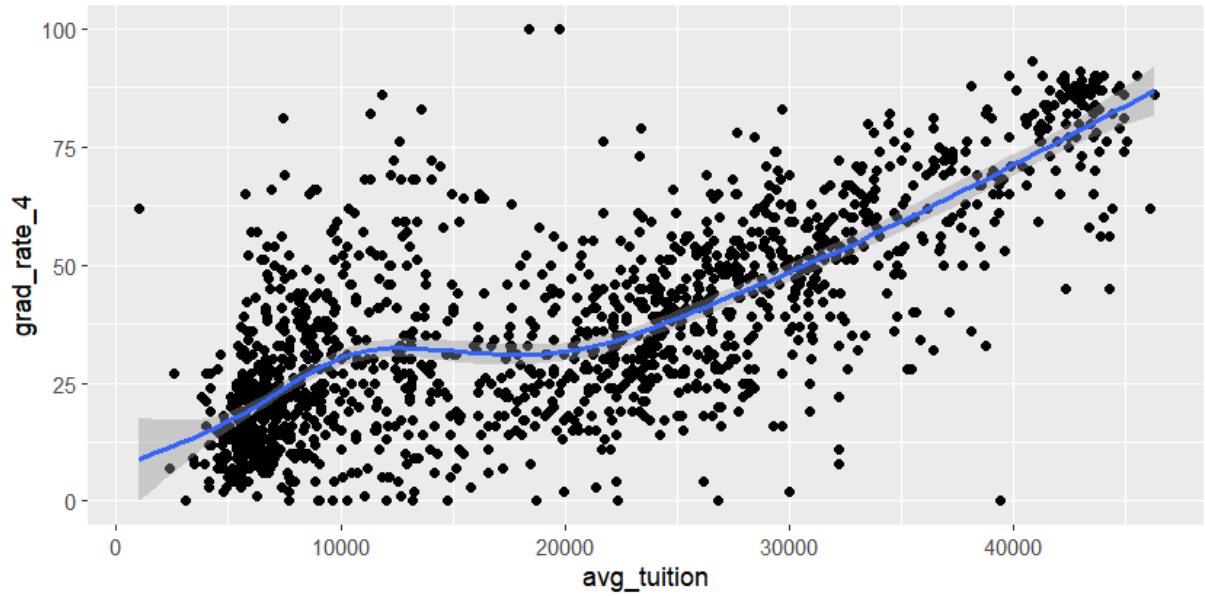
Appendix A.

List of Variables

Field	Variable	Type	Description
1	id	Numeric	Serial numbers of each university
2	name	Alpha	Name of university
3	highest_degree	Alpha	The highest degree that the university offers = Bachelor's Degree / Master's Degree / Doctor's Degree
4	acceptance_rate	Numeric	Total accepted number of students
5	enrolled	Numeric	Total enrolled number of students
6	avg_sat_reading	Numeric	Average SAT reading score of enrolled students
7	avg_sat_math	Numeric	Average SAT math score of enrolled students
8	avg_sat_writing	Numeric	Average SAT writing score of enrolled students
9	avg_act	Numeric	Average ACT score of enrolled students
10	num_bach	Numeric	Number of students receiving a Bachelor's degree
11	admitted_rate	Numeric	Percent of students admitted
12	avg_tuition	Numeric	Average tuitions of year 2010, 2011, 2012, 2013 (dollars)
13	state	Alpha	State in which the university is located
14	sector	Alpha	Sector of institution = Public / Private
15	urbanization	Alpha	Degree of urbanization = City / Rural / Suburb / Town
16	full_enrollment	Numeric	Number of full-time undergraduate enrollment
17	alaska	Numeric	Percent of total enrollment that are American Indian or Alaska Native
18	asian	Numeric	Percent of total enrollment that are Asian
19	black	Numeric	Percent of total enrollment that are Black or African American
20	latin	Numeric	Percent of total enrollment that are Hispanic/Latino
21	island	Numeric	Percent of total enrollment that are Native Hawaiian or Pacific Islander
22	white	Numeric	Percent of total enrollment that are White
23	tworace	Numeric	Percent of total enrollment that are two or more races
24	unknown	Numeric	Percent of total enrollment that are Race/ethnicity unknown
25	international	Numeric	Percentage of international students
26	women	Numeric	Percentage of women
27	grad_rate_4	Numeric	Percentage of student who graduates in 4 years
28	loan	Numeric	Percentage of students who receive student loans

Appendix B


grad_rate_4
100
75
50
25
0
0
10000
20000
30000
40000
avg_tuition


grad_rate_4
100
75
50
25
0
0
10000
20000
30000
40000
avg_tuition

Appendix C.

	grad_rate_4	t-value
(intercept)	-64.58532063* (4.865)	-13.110
avg_tuition	0.00086931* (0.00004)	20.202
white	0.06157057* (0.01735982)	2.771
avg_act	3.18364854* (0.18078817)	18.022
women	0.13571402* (0.0282336)	4.474
highest_degree : Docter's	-4.88268843* (1.00072073)	-4.749
highest_degree : Master's	-2.53128852* (0.94530493)	-2.552
full_enrollment	0.000389* (0.0000702)	5.631
urbanization : Rural	2.38737822 (1.4323203)	1/631
urbanization : Suburb	3.08092545* (0.7965176)	3.785
urbanization : Town	3.07481558* (0.85401142)	3.832
international	-0.27135030* (0.08606903)	-3.153
loan	0.01094238 (0.0251606)	-0.081

Observations: 1195, R^2: 0.7468, Adjusted R^2: 0.7454