

The Effect of Students' Debts on Post-Baccalaureate Earning

Linh Tang

May 5, 2020

Abstract

People in the US often either consider higher education as a life-changing opportunity or a life-long burden. Many students choose to take debts to help alleviate the financial burden of college. However, there has been endless controversy on whether the loans are really worth it. This paper aimed to analyze the effect of college's loan on how much the students earns on average at an institutional level. The result indicated that for every \$1000 increase in median debt, a student from the same institution would earn \$163 more annually. The research also came to a conclusion that taking on more debts to attend costly, more selective schools might pay off with higher post-graduation earning.

Keywords: Post-Baccalaureate, Income, Earning, College Debt, Student Loan, Education

JEL Classification: I2

1 Introduction

1.1 Motivation & Background

In 2017–18, the average total cost of attendance per year for first-time, full-time undergraduate students living on campus at 4-year institutions was \$50,300 at private nonprofit institutions, \$32,200 at private for-profit institutions and \$24,300 public institutions, according to the National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS). Meanwhile, the median American household income was \$62,626 in 2017, as reported in "Current population survey annual social and economic supplement" by the Census Bureau. Considering the level of income with the increasingly expensive college education, many students and their families had to depend on financial aid to cover part of the expenses. A short summary from the U.S. Department of Education reported that on average, the average net price of attendance (after aid) accounts for two-third of the family income for private institution while the ratio is one-third for public institution. Therefore, many students take loans, from both government and private sectors, to help alleviate the financial burden. But is the loan really worth it? According to a study conducted by YouGov, more than a quarter (27%) of undergraduate in the United States reported to have an outstanding student loan. More alarming, 37% of these borrowers are pessimistic about the day they can complete the repayment. Notably, only half of those with a student loan (50%) agree that the education they received was worth the debt they accrued and spent the rest of their life for repayment, while 34% "strongly" or "somewhat" agree that college's loan wasn't worth it.

1.2 Research Question

This paper aims to analyze the effect of college's loan on student's life after graduation. The subjects being researched are first-time, full-time undergraduate students at 4-year accredited institutions in the States. We measure the impact on how much the student achieves their student's post-graduation plan at an institutional level.

1.3 Overview

The data for this research project comes from the pooled mega database College Scorecard Data, the online server managed by the U.S Department of Education, containing data at the institution-level and data by field of study within each school across multiple school years. The site aims to increase transparency in the business of higher education by providing insight to all aspects to evaluate an institution, ranging from total cost to attend, graduation rate, employment rate, to average amount borrowed, loan default rate, etc.

The model being used in this research is linear regression model with 4 different specifications. After being examined, debts are shown to significantly impact the level of income after graduation. Nonetheless, we also concluded that there are multiple factors that also matters, such as school type, size, tuition, prestige, etc.

The paper is organized as follow: In section 2, we discuss the existing work and their limitations, as well as introduce relevant economic theories. In section 3 we describe the data set and assumptions in building our model. We describe the model in details and discuss its performance and our findings in section 4. Finally, in section 5, we discuss the scope of this work and how this can be expanded in future.

2 Literature Review

There has been several studies looking at the long-term financial consequences of student loan debt on the economy and on the individuals themselves. Researchers have long focused on gathering and analyzing patterns of students' life decisions whom graduate with repayment responsibility. There has been controversies over whether students should take out loans for a degree. According to Staklis and Bentz (2016), the the percentage of first-time four-year graduates employed in a full-time job within one year of graduation has gradually decreased over the last decade. Meanwhile, the proportion of students who take on loan for their education has been constantly rising along with the costs of college. However, research by Abel and Deitz (2014) has found out that on average, people with bachelor's degree earns about \$25,000 more per year than high school graduate. Cumulatively, bachelor's degree holders make \$1 million in additional earnings over their lifetime. Most of these studies

solely measure students' success through their level of income after graduation. However, there remains some literature questions about these models, such as how long after graduation and what factors should we use to measure student's success, and whether the level of income is indeed related to the student going to graduate schools or to the job market after finishing their bachelor's degree, and if so does there exist multicollinearity between these variables. Some concerns have been raised recently at the statistics that 40% of the current nation's student debt balance, which has been increasing over the past two decades up to \$1.5 trillion, belongs to graduate students although they represent only 15% of the borrowers, as reported in the Graduate Student Debt Review of New American Foundation. According to the Current Population Survey by U.S. Bureau of Labor Statistics, people with advanced degrees on average earn \$58,000 more annually than those with only a high school degree. It's claimed that students decide to postpone the repayment and take on more debt to advance their credentials in order to pursue a career in a field with significantly higher pay, thus make it easier to repay the loan in shorter time. This paper tests 2 following hypotheses. First is whether student debt positively impacts the income afterwards as it motivates students to perform better while in school and actively seek for high-paid jobs once graduate. We also want to test if students debt is correlated with the highest level of education one would obtain. We will examine the patterns of those with outstanding debts on their immediate plan after earning the degree, whether they tend to go for graduate school or coming out to industry.

3 Data & Empirical Model

3.1 Data Sources

The data being used for this research is cross-sectional data collected in the school year 2015-2016. Although the original data set is enormous and contain hundreds of variables, I realized a great deal of abundant information overfits our model. Thus, variables are then selected manually to best align with the target of this research. Nonetheless, the lack of necessary variables in each set mentioned above, as well as observations with missing values requires us to filter and combine them for a final sample. The sample is initially created

at institution-level, then the data by field of study is combined after being averaged. This process leads to a big loss of data, and consequently, a limited number of observations might decrease the statistical power of the research or lead to potential bias.

3.2 Sample Selection & Statistical Summary

Higher education in the US varies vastly amongst different types of schools. Specifically, there is a huge gap between 2-year institutions and 4-year institutions, as well as between different degree programs (associate, bachelor, master, PhD, postdoc) in terms of quality of education, value of degree, expectation for outcome, type of students and their family income, attendance cost i.e the financial burden, and earnings after graduations, etc. For example, students who are qualified to go to a 4-year university might still go to the 2-year one because of their financial situation and they are not able or not willing to take student's loans. Moreover, the statistics are also very different between an undergraduate program and a graduate program, since graduate students usually don't have to take much loan to pay for their study but instead get paid or are financially-prepared before entering. It is also important to note that employers and companies often pay lots of attention a person's highest degree during their hiring process, especially in the salary offer. In addition, many jobs, in particular those high-paid, would have specific qualifications in term of education.

Since it's barely possible to find a way to standardize these unquantifiable differences, this research will solely focus on 4-year undergraduate colleges and universities that are accredited across the US. Moreover, in order to reduce noise and possibility of multicollinearity, the sample included only data from the main campus of the institutions in case they have multiple locations.

Table 1 displays the statistical summary of the institutions in the sample by variables. Noticeably, the mean of loan principle (\$16,467) is about two third of that of the actual debt owned after graduation (\$23,115). The difference between these number indicate the effect of the loan, leading to an all-time controversial question of will taking loan to do college worth the money. Looking at the statistics, almost 60% of the students take loan to fund their undergraduate study, but we can be positive that on average, the annual earnings afterwards is still slightly higher than the cost of attendance (\$37859 to \$34557). However,

this might not be a valid answer since the number can varies across public and private, non-profit and for-profit, inclusive and selective schools. Generally, the standard deviations for most variables are huge, shedding light upon the noticeable gap between institutions. To illustrate this point, Table 4 and 5 display the difference in statistics between public and private schools, which can be found in the Appendix.

Variable	Mean	Standard Deviation	Observation
Average Earning	37859.37	11046.95	1339
Admission Rate	0.6556777	0.1943812	1339
Enrollment	5871.214	7572.315	1339
Total Cost	34557.46	14474.36	1339
% Loan Takers	0.5768924	0.1727939	1339
Median Debt (all)	16467.02	4857.75	1339
Median Debt (graduated)	23115.28	5086.595	1339
Median Debt (female)	16979.39	4951.97	1339
Median Debt (male)	15651.93	4984.167	1339

Table 1: Statistical summary of all sample data

3.3 Empirical Model & Assumptions

The final sample used for analysis consists of 1339 observation, each unit of observation is an institution with 14 unique factors covering the location (city & state), administration type (public, non-profit private, for-profit private), admission rate, enrollment, cost of attendance, percentage of loan takers, and amount of debt after graduation (both overall & categorized by students' gender).

In general, income, although it might not be the most holistic and optimal, is the most explicit and widely-used measurement of a person's success. Therefore, we will take the average earning to be the dependent variable. Since we are focusing on the effects of debt, we will treat the actual amount of debt after graduation as the primary explanatory variable in our model. Our hypothesis is that model with additional variables shows better result. This is justifiable as there are many other factors that might correlated with earnings, thereby not including them can lead to omitted variable bias. The selectivity varies greatly between

institutions, and thus the what students often expect to get out of their degree is likely to be a huge difference as well. Therefore, we would include admission rate in our model, as more selective schools are likely to have higher average students' earnings than inclusive schools. Moreover, the cost of attendance might play an important role since it can indicate the type of school and type of students there. Usually, students at expensive colleges are either from middle to high income family, or are very talented to get scholarships, or if not both, they are at least motivated and risk-seeking enough to take huge loan for their education. Moreover, as we notice a pattern in our sample, it's necessary to consider the type of schools in our model, whether the school is public or private, and if private, is it non-profit or for-profit.

The full model to predict **Average Earning**:

$$\text{Earn}_i = \beta_0 + \beta_1 \text{Admit}_i + \beta_2 \text{Cost}_i + \beta_3 \text{Debt}_i + \beta_4 \text{Public}_i + \beta_5 \text{Profit}_i + \alpha_i$$

where Admit_i , Cost_i , Debt_i and α_i are the average admission rate, cost of attendance, student debt and error terms, respectively. Public_i and Profit_i are the dummy variables indicate the type of the institutions.

As mentioned in the sample selection part, this model would only work for undergraduate four-year program in the United States. It's also important to acknowledge that the data being estimated is from only one school year, there might be possible unnoticed bias to our estimates. Our model is currently set up on an assumption that there isn't any extraneous variable hit that year specifically.

In addition, average earnings of students in the past might factor into a person's choice of school. It's not unusual to see students taking lots of loan to get into prestigious expensive college with the hope of a high-paid job post graduation to pay back the debt. As a consequence, reverse causality might possibly threaten the validity of this research. In that vein, a difference in difference model can be a potential solution, such as comparing colleges that gives lots of aid and those that don't, or colleges with high rate of loan takers and college with low rate. However, in this model, we are assuming that there doesn't exit endogenous variables.

4 Results & Discussions

Table 2 reports the primary results (number of observation = 1339). The first column presents the most basic specification, including only average amount of debt at school-level. The specification in the second column integrates the another financial factor that is likely to cause omitted variable bias: cost of attendance. The third model being examined consists of 3 independent variables: debt, cost and admission rate. The big difference in the coefficients of these 3 variables might due to the difference in their units of measurement. Finally, incorporating the type of schools into the last model gives us a full picture of how each factor impact the estimation of earning after graduation. All models are tested at 95% confidence level, the coefficients are highlighted in italics are insignificant, t-statistics are included in parentheses.

Variable	First	Second	Third	Full
Median Debt	0.335 (5.78)	<i>0.107</i> (1.42)	0.181 (2.37)	0.163 (2.121)
Total Cost		0.142 (5.64)	0.100 (3.75)	0.177 (4.79)
Admission Rate			-7000.2 (-4.33)	-6698.6 (-4.11)
School type (public/ private)				2838.2 (3.06)
School type (profit/ non-profit)				5164.0 (1.85)
Constant	32010.6 (30.35)	31190.4 (29.63)	36000.1 (23.60)	32268.8 (1.85)

Table 2: Primary Results

The first model estimates a statistically significant impact of debt on the average earning. Contrarily, the second specification suggest otherwise. This might be due to the omitted variable bias caused by unaccounted characteristics of the school. By controlling for the difference in cost of attending, we notice an insignificant impact of debt on post-graduation

earning. The third model consider 3 main empirical features of a school: debt, total cost and admission rate with all variables being statistically significant. However, these two models give badly performed results with high RMSE and low R-squared. I regard the last (full) model to be the preferred specification as it most completely captures the nature of the included variables and their impacts: median debt, total cost of attendance, admission rate, and the type of school (public, for-profit private, non-profit private). The full model suggests 95% confidence level for most variables except for whether the school is for-profit or non-profit. Interpreting the results, we can conclude that for every \$1000 increase in cost of attendance, students from the same school are likely to earn \$176 more, while every \$1000 increase in median debt would lead to \$163 increase in average earning. If the school's admission rate decrease by 1%, the students are likely to earn \$67 more. On average, students from public school earn \$2838 more than private schools' students. A heterogeneity analysis has also been conducted between school types. Since the effects of whether a school is for-profit or non-profit is not significant, we will study the variation between public and private colleges only. The result of each school type separately is presented in table 3 below.

Variable	Public (N = 493)	Private (N = 846)
Median Debt	<i>0.224</i> (1.41)	<i>0.141</i> (1.05)
Total Cost	0.455 (4.65)	0.123 (2.14)
Admission Rate	<i>2580.2</i> (1.20)	-10581.9 (-3.77)
Constant	22098.8 (7.66)	37666.4 (12.70)

Table 3: Heterogeneity Analysis of Private vs. Public Schools

5 Conclusion & Future Work

Level of debts play a statistically significant role in the future income of students. However, being examined in our model, there are more than just debt in what determine how much

you will earn after graduation. As expected, going to expensive selective schools increases the likelihood of getting a higher income for students, even though it means that students have to take bigger debt to afford the prestige. One surprising finding is that attending public school is likely to increase your future income, given the same cost, debt and admission rate. However, this isn't likely to happen, especially there's often a big gap between the cost of attendance between private and public that even though the model accounts for differences in cost, it's still hardly possible in real life. It's also important to note that non-profit and for-profit colleges might not bring a difference. This can be explained by the close similarities between these two types of schools, or maybe due to the very small (relatively) number of for-profit schools available in the data set as well in our normal life. In conclusion, we found out that debt, the leading variable in our study, does have a positive correlation with the earning after graduation. Therefore, it might be advisable to take on more debts to attend costly, more selective schools with the hope of earning more with that degree.

What is more, there stands a chance of multi-collinearity between the 5 independent variables. As a matter of fact, private schools are often more expensive and might have less money to give as they aren't funded by the federal, thus pushing students to take more loans. Also, many school with low admission rate are costly and need-based, thus might have some correlation with both cost of attendance and amount of debt. More hypotheses and tests must be conducted to give evident conclusion, but this concern should highly be considered.

Moreover, there is a likelihood that the extend and effects of debt depends on students' gender. Our sample suggests that there is a slight difference in the median debt between female and male. However, since the model is being examined at institutional level and we don't have information about the percentage of each gender in the whole enrollment, I doubt that taking gender into account would improve the performance of the model. Nonetheless, more tests can be done to verify this assumption and necessary adjustment. Obviously, earning might not completely reflect personal success and this has been usually discussed all over the places. However, there has been no concrete, universal formula to measure this abstraction. Moreover, since we are and due to the limited sources of data available, annual income should be enough for our model for now. But it is important that we take this drawbacks into account when looking at future steps of this research.

References

- Abel, J. R., & Deitz, R. (2014). Do the benefits of college still outweigh the costs? *Current Issues, Economics and Finance*, 20(3).
- College scorecard data. (2020). U.S. Department of Education. <https://www.collegescorecard.ed.gov/data>
- Current population survey annual social and economic supplement. (2019). US Census Bureau, Department of Commerce. <https://catalog.data.gov/dataset/current-population-survey-annual-social-and-economic-supplement>
- Delisle, J., Phillips, O., & Linde, R. (2014). The graduate student review debt. *New America Education Policy Program*.
- Nguyen, H. (2018). A third of those with student loans say taking on the debt wasn't worth it. <https://today.yougov.com/topics/education/articles-reports/2018/10/10/third-those-student-loans-say-taking-debt-wasnt-wo>
- Price of attending an undergraduate institution. (n.d.). National Center for Education Statistics. https://www.nces.ed.gov/programs/coe/indicator_cua.asp
- Staklis, S. (2016). Employment and enrollment status of baccalaureate degree recipients 1 year after graduation: 1994, 2001, and 2009. National Center for Education Statistics. <https://www.nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2017407>

6 Appendix

Variable	Mean	Standard Deviation	Observation
Average Earning	38693.09	12548.56	846
Admission Rate	0.6355214	0.2020878	846
Enrollment	2738.294	3790.628	846
Total Cost	42663.09	11936.9	846
% Loan Takers	0.6135739	0.171121	846
Median Debt (all)	1751.38	4885.907	846
Median Debt (graduated)	24275.72	4840.593	846
Median Debt (female)	18399.73	4954.272	846
Median Debt (male)	16915.84	51830791	846

Table 4: Statistical summary of the non-profit private schools

Variable	Mean	Standard Deviation	Observation
Average Earning	36428.68	7627.867	493
Admission Rate	0.6902665	0.1752592	493
Enrollment	11250.1	9240.668	493
Total Cost	20647.99	4265.345	493
% Loan Takers	0.5139462	0.1569161	493
Median Debt (all)	14091.42	3771.4970	493
Median Debt (graduated)	21123.95	4881.732	493
Median Debt (female)	14542.06	3887.306	493
Median Debt (male)	13483.02	3734.394	493

Table 5: Statistical summary of the public schools