Proceedings of the 5th International Conference on Education Innovation and Philosophical Inquiries
DOI: 10.54254/2753-7048/62/2024PE0002

Analysis of Factors in College Enrollment Using Regressions

Shaoming Zhi[1,a,*]

[1]*College of Arts and Science, New York University, New York, US*
a. sz3378@nyu.edu
**corresponding author*

Abstract: College enrollment, the process in which families make investments for the future, individuals receive higher education and seek self-improvement, and an aspect closely related to the lives of everyone in a modern society. The trends of college enrollment have been fluctuating due to various factors throughout history, mostly due to socioeconomic factors. In this article, multiple linear regression models are constructed to analyze and evaluate how some possible factors are correlated with the fluctuations of college enrollment in the United States each year from 1973 to 2022. Data for dependent and independent variables were gathered from government and non-profit organization websites such as the Federal Reserve Bank of St. Louis, the United States Census Bureau, and the Educational Data Initiative. The results have shown that many factors are influencing the fluctuations, with GDP and income being the main factors; this article provides a supporting role in making admission policies and country developments.

Keywords: College, College Enrollment, United States

1. Introduction

Analyzing college enrollment patterns is one of the most direct and effective ways of testing the education that a society has. The amount of people attending college can be affected through many aspects, just as it reflects many aspects as well. Modeling college enrollment can assist in analyzing economic developments, understanding how accessible higher education can be, seeing how willingly households are to send their family members for investments during different socioeconomic statuses, and making policies on education, admission, or funding.

The topic of college enrollment has been a continuous topic discussed by the academic world, and numerous studies have been conducted to understand it more. For instance, research on college enrollment and decision-making during recessions of 2008 and 2020 was done by Barr and Turner, and Dworak, respectively [1, 2]. Jung and Shrestha discussed the effects of financial acts of government on college enrollment, and Manoli and Turner gathered income data and its impacts on college enrollment [3, 4]. Juszkiewicz broke down the details of college enrollment, gender ratio, and completion rate trends for a time series of ten years, and Perna quantified the racial and ethnic group distributions in college enrollment [5, 6].

Most of the existing articles regarding college enrollment focus on only one or two impacting variables. This article aims to examine college enrollment in the entire United States for a time series of fifty years, from 1973 to 2022, with a balanced data set. The results have shown that among five

DOI: 10.54254/2753-7048/62/2024PE0002

independent variables, GDP, income, and unemployment rate have statistically significant impacts on college enrollment, and GDP explains most of the variances in college enrollment.

2. Data and Methods

Data in the entire United States from 1973 to 2022 on GDP, median income, tuition, the proportion of the age 5-19 population to the whole population, and unemployment rate were collected from websites including but not limited to the Federal Reserve Bank of St. Louis, the American Institute for Economic Research, and the United States Census Bureau. The reason behind selecting these types of data as independent variables is that the wealth and affordability of colleges directly affect the number of people enrolled in them. Knowing the relative affordability of college with respect to social wealth for fifty years will be beneficial for households, educational institutions, and the government.

Using the data gathered, a time series was constructed in an Excel worksheet. A process of removing the factor of inflations was also applied. With the integration of R, linear regressions were conducted using said data: package "readxl" was utilized to convert the Excel worksheet into a data frame [7]. The built-in "lm" function was applied, and four models were designed. The number of college enrollments was set as the dependent variable; GDP, median income, tuition, the proportion of the age 5-19 population to the whole population, and unemployment rates were set as independent variables. The outcomes, including coefficients, standard errors, p-values, and significance were obtained by using the package "stargazer" [8]. The equations are shown below.

$$log(CollegeEnroll) = \beta_0 + \beta_1 log(GDP) + \varepsilon \tag{1}$$

$$log(CollegeEnroll) = \beta_0 + \beta_1 log(GDP) + \beta_2 log(Income) + \beta_3 log(Tuition) + \varepsilon \tag{2}$$

$$\log(CollegeEnroll) = \beta_0 + \beta_1 \log(GDP) + \beta_2 \log(Income) + \beta_3 \log(Tuition) + \beta_4 YouthPercent + \varepsilon \tag{3}$$

$$\log(CollegeEnroll) = \beta_0 + \beta_1 \log(GDP) + \beta_2 \log(Income) + \beta_3 \log(Tuition) + \beta_4 YouthPercent + \beta_5 Unemploy + \varepsilon \tag{4}$$

Each variable represents their respective following: CollegeEnroll is the number of people enrolled in college in the United States. GDP is the Gross Domestic Product of the United States, in billions of dollars. Income refers to the median household income of the United States, in dollars. Tuition is the average tuition of all private and public colleges across the United States, in dollars. YouthPercent is the proportion of the age 5-19 population to the whole population in the United States, in percentage. Unemploy is the unemployment rate across the United States, in percentage.

3. Results

The regression summary using stargazer is summarized as follows in Table 1. The data for college enrollment, GDP, median income, average tuition, youth percentage, and unemployment were extracted from Educational Data Initiative, Federal Reserve Bank of St. Louis, United States Census Bureau, BestColleges, USA Facts, and Federal Reserve Bank of St. Louis, respectively. [9, 10, 11, 12, 13, 14]

DOI: 10.54254/2753-7048/62/2024PE0002

Table 1: Regression of College Enroll on Variables

	Dependent Variable: log(CollegeEnroll)			
	(1)	(2)	(3)	(4)
log(*GDP*)	0.264***(0.011)	1.061***(0.109)	0.924***(0.233)	0.904***(0.195)
log(*Income*)		-1.120***(0.153)	-1.038***(0.197)	-0.857***(0.193)
log(*Tuition*)			0.070(0.104)	0.002(0.081)
YouthPercent				0.007(0.007)
Unemploy				0.019***(0.003)
Constant	14.171***(0.096)	18.592***(0.607)	18.396***(0.678)	16.982***(0.967)
Observations	50	50	50	50
R^2	0.928	0.966	0.967	0.982
Adjusted R^2	0.926	0.965	0.964	0.980
Residual Std. Error	0.060 (df = 48)	0.042 (df = 47)	0.042 (df = 46)	0.032 (df = 44)
F-Statistic	615.695***	672.387***	443.124***	469.584***

Notes: The numbers in parentheses next to a value represent standard error, or degrees of freedom if noted. Trailing asterisks represent the level of significance: one asterisk represents $p<0.1$, two represent $p<0.05$, and three represent $p<0.01$.

The logarithm of some variables is taken before being used in the regression, as doing so has many benefits. When a variable, such as GDP, is numerically large, the amount of change for increasing one unit of that variable is not enough to be shown by three decimal places. When the logarithm of a large variable is used in a regression, it can show the impact of percent changes in the variable, which makes the coefficients easier to interpret. Moreover, not all variables are perfectly linear and have constant residuals. A logarithmic regression can reduce the negative effects of nonlinearity and heteroskedasticity [15].

The main independent variable GDP shows a statistical significance and positive correlation in model (1), and GDP alone was able to provide an adjusted R2 of 0.926. Then, as the second variable Income was introduced in model (2), both are significant, and income has a negative correlation and GDP maintains its positive correlation. In models (3) and (4), the introduction of new variables does not alter the significance and sign of GDP and Income. Though YouthPercent and Tuition have some effect on college enrollment, those effects are statistically insignificant. Unemployment, on the other hand, shows a significant, positive correlation with college enrollment.

Note that interpreting the constant (the intercept when holding all independent variables zero) is meaningless because even though having no youths and zero GDP, income, tuition, and unemployment rate at the same time for a society makes sense mathematically, it is almost impossible that any society would be in such a situation by common sense. And regarding this research, the United States is unlikely to undergo such a situation.

DOI: 10.54254/2753-7048/62/2024PE0002

4. Discussion

GDP and college enrollment in the United States are positively correlated. GDP, or Gross Domestic Production, measures the number of consumptions, investments, government spending, and net exports over a given period [16]. In other words, GDP measures all economic activities within a country and reflects how much the government is thriving overall. An increase in GDP means that the economic activities mentioned above are increasing, leading to the citizens having better living conditions. And since the living conditions are improving, more families are capable and are willing to send their children to college for better education.

Median income and college enrollment in the United States are negatively correlated. Firstly, there have been high-salary jobs that do not require a college degree throughout time, such as firefighters, truck drivers, and mechanics. Besides that, many new types of jobs began to exist; as technology became more and more advanced, social wellness became better and better, and online resources became more and more accessible. Social media content creators, digital art designers, web designers, and developers are examples of newly emerged, high-salary jobs if one has talent or can self-study.

There is a nonzero but statistically insignificant correlation between tuition and college enrollment in the United States. The coefficient value is 0.002, which means that tuition for college is quite inelastic. Families that are capable and determined to send their children to college will not change their decision with a change in the amount. Additionally, as GDP and income have been in rising trends from 1973 to 2022, it is likely that spending used on subsidies, financial aid, or scholarship programs for college students has increased as well, which makes college more affordable despite having the same tuition.

There is a nonzero but statistically insignificant correlation between the proportion of the age 5-19 population to the whole population and college enrollment in the United States. Peer competition is a significant factor, and the increase in population outruns the increase in capacity of colleges. As more youths are present, in both numbers and proportions, the criteria for entering college are rising as well. Therefore, as entering college is becoming more and more challenging, many families may choose to have their children pursue different futures in life if their children cannot enter college.

The unemployment rate and college enrollment in the United States are positively correlated. An individual is considered unemployed when that individual is jobless and searching for a job [16]. There are many college students who are actively looking for part-time jobs and interns aside from college work. Those students are considered unemployed, as they are following the same procedure as formally applying for a full-time job. Besides, when the unemployment rate is high, it can also mean that the economy of the country is not ideal. In such cases, attending college instead of working or looking for a job can be better off for many people.

GDP is the main factor contributing to the fluctuations of college enrollment, as it has a statistically significant and positive correlation, and as shown in model (1), it alone has an adjusted R2 of 0.926. Other factors such as income and unemployment rates are also statistically significant, with the former having a larger impact. The proportion of the age 5-19 population to the whole population and tuition have a nonzero but statistically insignificant correlation.

5. Conclusion

In summary, the usage of regression models to test the trends of college enrollment sufficiently shows statistical significance that reflects socioeconomic factors. Affecting variables such as GDP, income and unemployment rate explains the fluctuations of college attendance throughout the United States during the late 20th century and the early 21st century. It is essential that the data are well-analyzed by households, colleges, and governments to make decisions and make the future of humanity brighter.

DOI: 10.54254/2753-7048/62/2024PE0002

References

[1] *Barr, A., & Turner, S. (2015). Out of work and into school: Labor market policies and college enrollment during the Great Recession. Journal of Public Economics, 124, 63-73.*

[2] *Dworak, A. (2020). United States university enrollment numbers during the COVID-19 pandemic recession. Perspectives on the New Normal: Post COVID, 19, 67.*

[3] *Jung, J., & Shrestha, V. (2018). The Affordable Care Act and college enrollment decisions. Economic Inquiry, 56(4), 1980-2009.*

[4] *Manoli, D., & Turner, N. (2018). Cash-on-hand and college enrollment: Evidence from population tax data and the earned income tax credit. American Economic Journal: Economic Policy, 10(2), 242-271.*

[5] *Juszkiewicz, J. (2020). Trends in Community College Enrollment and Completion Data, Issue 6. American Association of Community Colleges.*

[6] *Perna, L. W. (2000). Racial and ethnic group differences in college enrollment decisions. New Directions for Institutional Research, 2000(107), 65-83.*

[7] *WICKHAM H, BRYAN J. (2023). Readxl: Read Excel Files. Tidyverse, Readxl. Retrieved from https://readxl.tidyverse.org, https://github.com/tidyverse/readxl.*

[8] *HLAVAC M. (2022). Stargazer: Well-Formatted Regression and Summary Statistics Tables. The Comprehensive R Archive Network, R package version 5.2.3. Retrieved from https://CRAN.R-project.org/package=stargazer.*

[9] *HANSON M. (2024). College Enrollment & Student Demographic Statistics. Educational Data Initiative. Retrieved from https://educationdata.org/college-enrollment-statistics.*

[10] *U.S. BUREAU OF ECONOMIC ANALYSIS. (2024). Gross Domestic Product - GDP. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/GDP.*

[11] *GUZMAN G, KOLLAR M. (2023). Income in the United States: 2022. United States Census Bureau, Retrieved from https://www.census.gov/library/publications/2023/demo/p60-279.html.*

[12] *BRYANT J. (2024). Cost of College Over Time. BestColleges, Retrieved from https://www.bestcolleges.com/research/college-costs-over-time/.*

[13] *USA FACTS (2022). Our Changing Population: United States. Retrieved from https://usafacts.org/data/topics/people-society/population-and-demographics/our-changing-population/.*

[14] *U.S. BUREAU OF ECONOMIC ANALYSIS. (2024). Unemployment Rate - UNRATE. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/UNRATE.*

[15] *Stock, J. H., & Watson, M. W. (2020). Introduction to econometrics. Pearson.*

[16] *Agénor, P. R., & Montiel, P. J. (2015). Development macroeconomics. Princeton university press.*