

EXPLORING THE DETERMINANTS OF HIGHER EDUCATION DEGREE PRODUCTIVITY IN MACHINE LEARNING

Qianhui Guo, Seattle University, guoqianhui@seattleu.edu
James J. Lee, Seattle University, leej@seattleu.edu

ABSTRACT

Studying the measures and determinants of institutional productivity is a critical field for policymakers and institutional leaders to identify an improvement strategy on how to allocate properly limited resources for higher productivity. Among them, degree completion is not only an ultimate outcome for students who are seeking a degree program, but also one of the essential measures that evaluate student success. Studies show that achieving higher degree productivity is identified as the most common concern for postsecondary institutions. This study reveals cost-related determinants play significant impacts on degree productivity by adopting a machine learning approach. The results suggest how these determinants worked in different scenarios which further explains the previous studies in degree productivity. With more online education being forcefully enforced with the lifestyle adaptation with COVID-19 pandemic, this explanation in degree productivity plays an important role to provide visions on how to allocate resources in an online learning environment.

Keywords: Degree Productivity, Machine Learning, Higher Education, Random Forest Regression

INTRODUCTION

Our study provides analytical guidelines for most institutions in the U.S. with evidence-based support, helping them make appropriate decisions toward improvement. With the threat of COVID-19, U.S. higher education has become unprecedentedly challenging as classes have moved to online in 2020. However, many universities failed to adapt to the rapid changes. It was reported that many students from over 25 universities in the U.S. had complained about the quality of online classes (Collin Binkley, 2020). Within a harder situation in delivering high-quality education online, an institution with originally high productivity is more likely to know how to pursue productivity even on an online basis. With this concern, our study will help higher education institutions lay a solid foundation for future education reform.

Institutional productivity has been one of the most pressing finance concerns among the higher education community (Kolbe and Kelchen, 2017). Institutional productivity refers to the return on institutional investment, measuring the effectiveness of resource management within an institution (Miller, 2007). Institutional productivity has been drawing increased attention from government, institutions, and researchers. RAND (1997), a nonprofit research institution, prepared an overview strategy on productivity improvement for California Education Roundtable enabling state policymakers to develop a productivity improvement program to adapt to the significant finance challenge. With a request from the National Postsecondary Education Cooperative (NPEC) and the National Center for Education Statistics (NCES), Kolbe and Kelchen (2017) conducted a report uncovering the reasonable measures of institutional productivity among a broad range of stakeholders. The recent report from Pew Charitable Trusts (2019) indicated that the federal government's funding had increased 24 percent to \$4544 for a full-time student between 2000 and 2015. With the increasing expenses in a higher education institution, productivity improvement has become unprecedentedly essential for stakeholders who aim to continue meeting public expectations without more spending.

This research, however, investigates the degree productivity using a machine learning method, revealing evidenced-based and robust recommendations to higher education without increasing their overwhelming governance in this field. To determine how general characteristics, student demographic characteristics and financial characteristics of institutions contribute to degree productivity and understand what actions higher education can take for further improvement, we imported 14 related potential determinants available in The Delta Cost Project and ArcGIS database. To test the model, we applied a random forest regression with the 100 trees algorithm to train a model with 80% of the total 34498 observations and adopted Mean absolute percentage error (MAPE) to assess the performance of the

model tested by the rest 20% of data. Additionally, we also employed the generalized additive model (GAM) smoothing with a 95% confidence level to uncover the tendency between predictor features and degree productivity.

The remainder of the paper is organized as follows: Sector 2 presents previous studies in degree productivity, and Sector 3 reveals data-preprocessing and machine learning methods. Sector 4 shows the results of our model, followed in Section 5, further discussing recommendations on the improvement of degree productivity. Finally, we indicate explications on the theoretical contribution in Sector 6 and summary the achievement of this study in Section 7.

DEGREE PRODUCTIVITY IN HIGHER EDUCATION

Researchers have been studying the measures and determinants of institutional productivity to help policymakers and institutional leaders to identify an improvement strategy on how to allocate properly limited resources for higher productivity. Among many goals of an institution, Miles et al. (2018) indicated that the number of graduates produced per dollar spent is the most common concern for a postsecondary institution. This measure has been applied in some literature for degree productivity improvement (Harris, 2013; Davis and Olga, 2013). Degree productivity plays an important role in the development of higher education institutions, with the potential to gain more enrollment, higher reputation, and more academic achievements. However, insufficient empirical studies make it unclear for an institution to take appropriate steps. The productivity was perceived impossible to control among a large number of postsecondary institution leaders and scholars as it has been accepted the fact that quantity increase in degree completion always came along with the sacrifice of quality (Harris, 2013).

Degree Productivity

Institutional productivity, defined as the ratio of institutional outputs to inputs, is regarded as a standard to measure the effectiveness of resource management within an institution (Miller, 2007). Measuring institutional productivity was the highest priority among a broad range of stakeholders, including institutional representatives, policymakers, researchers, and representatives from professional organizations (Kolbe and Kelchen, 2017). However, finding an overall assessment that measures institutional productivity becomes quite complicated (Miles et al., 2018), as higher educations have a wide range of desired outputs, including research, teaching, student achievement, contributions to the public, and so on. Focusing on single or several institutional outcomes, many researchers have already adopted different measures of “productivity” (Massy et al., 2013), such as enrollments, credit hours, degrees, graduation rates, etc. The National Education Center proposed a multifactor measure of productivity using the idea of “sheepskin adjusted instructional activity” (Jaeger and Page, 1996; Massy et al., 2013), and Harris (2013) suggest the ratio of degrees granted to total sector expenditures as a productivity indicator in his research, to help institutional stakeholders to get control over college costs.

Degree completion is not only an ultimate outcome for students who are seeking a degree program, but also one of the essential measures that evaluate student success (Kahu and Karen, 2018). This indicator arouses great concerns from students, parents, institutional faculty, policymakers, and other related stakeholders (Astin and Leticia, 2005). Students and parents think highly of degree attainment as it helps achieve the students’ career goals. Additionally, it is a significant indicator that reveals how well the institutional performance is. Therefore, in this research, we define ‘**degree productivity**’ as the number of degrees granted as the institutional output. All educational and related expenses are measured into higher education productivity, in which degree productivity can be explained in terms of student success given limited available resources.

Institutional Attributes

Several institution-level characteristics significantly influence the degree completion and institutional costs (Baumol et al., 1995; Bowen et al., 2009). Those characteristics are divided into three groups based on different entities: institution, student, and finance.

General Institutional Characteristics

Student-Faculty Ratio. A small student-faculty ratio always refers to small class size. In a small class, students are expected to have more interaction and participation, and thus gain more learning. It is found that reducing student-faculty ratio is more likely to increase degree completion (Bound et al., 2010), but Harris (2013) mentioned there is a small impact of reducing student-faculty ratio on productivity.

Institutional Size. Different institutional size is always along with a different institutional environment and culture. A larger institution is more diverse but at the same time requires more spending on education and tends to be more challenging to manage. Several researchers uncovered that there was a negative relationship between institutional size and students' likelihood of persistence or degree completion (Pascarella & Terenzini, 2005; Bailey et al., 2006; Calcagno et al., 2008). But Titus (2004) found that 4-year larger institutions play an essential role in increasing students' likelihood of persistence.

Part-time Faculty Proportion. Increasing the proportion of part-time faculty is a great cost-saving strategy for higher education. There is a huge gap between the salary of full-time and part-time faculty. According to *The 2004 National Study of Postsecondary Faculty*, the average basic salary for a full-time faculty is \$65,407, while only \$11,160 for part-time faculty. However, the increase in the part-time faculty has a significantly negative impact on students' graduation rate at a four-year college (Ehrenberg et al., 2005). Jacoby (2006) also used regression analysis to find an adverse impact on students' graduation rates when institutions relied heavily on part-time faculty.

Retention Rate. Student retention has a direct influence on degree completion. Astin (2005) demonstrated that a retention rate is a useful approach to judge the productivity of the actual degree completion rate. She gave an example to show that if a college did a poor job in student retention, the actual degree completion rate would far behind the expected degree completion.

Sector. According to the National Center for Education Statistics, the total spending per full-time equivalent student is 30.7% higher at private non-profit 4-year institutions than a public 4-year institution in 2016-17. Despite the higher expenditure, a private university tends to provide better education quality in many aspects, such as lower class size, more networking opportunities, more specialized classes, etc. Thus, attending a private university tends to increase students' chance of persistence and degree completion compared with attending a public college or university (Astin, 1975; Oseguera, 2005).

Administrative Support. Administrative support is necessary for an institution to maintain a daily basis. But researchers demonstrated that administrative expenditure has a negative impact on student success in four-year institutions (Titus, 2004; Ryan, 2004). The outcome could be better if more spending is used in a more effective way like instructions instead of administrative services.

Student-related Characteristics

Student Demographic Characteristic. Previous research shows that students' demographic characteristics, including gender, race/ethnicity, and the proportion of part-time students, did affect the degree completion significantly. Female students gain more intolerance/awareness and communication skills (Toukoushian et al., 2001) and have a higher probability of completing a degree (Astin, 2005). Regarding race and ethnicity, White or Jewish students are more likely to complete a degree compared with other racial groups (Astin, 2005; Titus, 2006). Other research shows a consistent finding that the graduation rate is lower when there's a large share of the minority in a university (Calcagno et al., 2008). Additionally, a higher graduation rate, which usually comes along with higher instructional expenditures, can result from the smaller share of part-time students (Bailey et al., 2006).

Student Pre-college performance. SAT or ACT are both standardized tests for college admissions in the U.S., used as a measure of student pre-college performance by many researchers. Douglas and Ronald (2010) exhibited that higher graduation rates occur at high SAT institutions relative to low SAT institutions. But the SAT failed to predict who will drop out from an institution in Astin's (2005) research.

Financial Characteristics

Student Services. Student services, including health services, student counseling, career advising, and so on, are indispensable parts for a higher education institution to meet students' satisfaction and attract more enrollments. The growth of student services spending is more likely to increase student persistence and graduation rate, especially for those colleges where students have lower entrance test scores or have lower family incomes. It is also pointed out reallocating some funding from instruction to student services may positively influence students' persistence and graduation rates at some institutions (Douglas and Ronald, 2010).

Financial Aid. Financial difficulties become a big challenge for students to complete their degrees in an institution. Much previous research has concluded the positive direct and indirect effect of financial aid on students' persistence and degree completion, controlling for other related characteristics (Terkla, 1984; Perna, 1998). Perna (1998) also mentioned that the effectiveness of financial aid depends on the type and package of the aid. The grant is more effective than a loan in enhancing degree completion. Besides, it is presented that an institution with lower levels of Pell Grant dollars per FTE student has a higher graduation rate (Douglas and Ronald, 2010).

DATA AND METHOD

Data Sources

The data used in this paper are from *The Delta Cost Project*, a database that reveals what colleges do with their money in the period between 1987 and 2015. It aggregates the Integrated Postsecondary Education Data System (IPEDS) finance data to support comparisons over time and adjust for changes in accounting standards and the IPEDS survey questionnaire. A set of core data is collected and analyzed every academic year, including student services share of education and related expenses, education and related expenses per degree, and full-time student retention rate. In the projects, we extracted sample data from *The Delta Cost Project* between 2004 and 2015 with 58623 observations and 673 variables. A total of 7155 post-secondary institutions in the US are included in the dataset for 12 academic years. Regarding the geographical data, we extracted the percentage of US total GDP by State from ArcGIS with 51 states from 2004 and 2015.

Estimation Approach

To estimate the relationship between degree productivity and the cost-related determinants, we construct a predictive model using the Random Forest Regression solution. There is a tendency that a single decision tree model might overfit the training dataset but perform poorly on the training dataset. By aggregating results from large numbers of decision trees, the random forest can fix such an issue. Unlike a single decision tree, observing the nodes within each decision trees in a random forest become impractical. However, a random forest can provide extra information, including variable importance and proximity measure within the internal structure of the data (Andy and Matthew, 2002). A random forest model with a proper number of trees produces a high accuracy rate in short operation time and also works well with high dimensional data set and outliers, which is a good fit for our large dataset. In our research, after comparing many random forest models with different numbers of trees, we found that the accuracy rate of the random forest model apparently decreased with the increase of decision trees if the number of trees isn't big enough. Therefore, we adopted a random forest model with 100 trees in our project to improve the accuracy and robustness of the predictive model and to uncover the significance of predictor variables on determining degree productivity.

Measures

Measures of degree productivity. The Delta Cost Project database includes *Education Expenses Per Degree*, a measure of spending on direct educational costs per degree in an academic year. By employing educational and related expenses as input and the number of degrees granted in an academic year as the institutional output, we can measure how productive an institution can grant a degree. The more educational spending on a degree awarded, the lower degree productivity is within a postsecondary institution.

General institutional characteristics, student demographic characteristics and financial characteristics are three important parts for a postsecondary institution, and arouse much attention from researchers and institutional stakeholders (Bound et al., 2010; Astin, 2005; Titus, 2006; Terkla, 1984;). In this study we adopt the three parts above to provide actionable insights on these fields to the U.S. higher education institutions.

Measures of General Institutional Characteristics. To measure institutional size, we adopted the *12-month full-time equivalent (FTE) enrollments*, including both undergraduate and graduate students. As full-time students make up an essential part of an institution, we employed *full-time students' retention rates*, which displays the percentage of previous year's fall first-time, full-time students re-enrolled at the institution in the following fall, to represent retention rate in our project. Additionally, Full-time and part-time faculty were found to have a different impact on students' outcomes (Ehrenberg et al., 2005). To measure the influence of student-faculty ratio as well as part-time faculty proportion on degree productivity, we import two measures from the original database, including *full-time faculty per 100 FTE students* and *part-time faculty per 100 FTE students*. The *Delta Cost Project* database includes the measure of *the sector* with ten institutional categories, including "Public 4-year or above", "Private nonprofit 4-year or above", "Private for-profit 4-year or above", "Public 2-year", "Private nonprofit 2-year", "Private for-profit 2-year", "Public less-than-2-year", "Private nonprofit less-than-2-year", "Private for-profit less-than-2-year" and "Sector unknown". Considering the local economic condition would influence the education and related cost, so we integrated the percentage of total US GDP by State from ArcGIS into the institutional data according to the State and academic year.

Measures of Student Demographic Characteristics. Previous research exhibits that race and ethnicity influence the likelihood of completing a degree for a student (Astin, 2005; Titus, 2006). To study how racial characteristics affect degree productivity, we also imported the ratio of the White student enrollment to total student enrollment as the percentage of White students. Similarly, we imported the percentage of Black students, Asian students, Hispanic students, as well as non-resident students.

Measures of Financial Characteristics. We consider two indicators to represent student service, including the expenditures for student services in an academic year and the share of education and related spending on student services. As for Financial aid, we import the measure of the total of undergraduates receiving any aid from the *Delta Cost Project* database to find out whether financial aid significantly affects degree productivity.

Data processing

Outliers. The sector is the only categorical variable among all predictor variables. According to the descriptive statistics, three categories have less than ten observations, respectively, including "Private nonprofit less-than-2-year", "Private for-profit less-than-2-year", and "Sector unknown". As the number of each category is too small to train a model, we decide to remove these three categories as outliers. In terms of numerical variables, we consider the data that is above the 90th percentile or below the 10th percentile as outliers. If a value is greater than the 90th percentile, we replaced it with the value of the 90th percentile. Similarly, if a value is less than the 10th percentile, we replaced it with the 10th percentile.

Missing Value Treatment. We first eliminate all the missing values from the independent variable *degree productivity*. In terms of the missing value in the numerical dependent variables, we first removed all the missing values from *institutional size* as we found that there is always a missing in the *Retention Rate* if the value of *institutional size* is missing. Then we replaced the missing value with the mean value of the same institution ID. As there were only less than 1% missing among the whole dataset, we finally removed the few rests of the missing values.

Descriptive Statistics

After making the adjustments above, our sample size is 34498. According to the descriptive statistics shown in Table 1, on average, an institution spent \$73,347.33 on education and related expenses per degree with the range between \$33,720.98 and \$153,727.84 vary from different institutions. Besides that, an institution spent on average 14.75% of its total education and related spending around \$5,414,880 on student service per academic year, but some intuitions unreasonably allocated their resources on student service up to 25.16% of their total education and related spending.

On average, there are around four full-time faculty per 100 FTE students and seven part-time faculty per 100 FTE students in a postsecondary institution. On average, there are around 3346 FTE enrollments for 12 months among all the institutions involved in our sample, about 66.64% of the previous year’s fall first-time, full-time students that re-enrolled at the institution in the following fall. Besides, around an of average 438 aids are granted to students in an academic year. In terms of student demographic characteristics, the majority race in the postsecondary institution in the United States is White students, average taking up 59.17%, followed by Black students 12.35%. The minor group of Asian students, Hispanic students, and non-resident students take up an average of less than 10% in an institution. According to the GDP information, the richest state shared 9% of the US total GDP, while the most impoverished state shared only 0.7%. Figure 1 shows the distribution of sectors, the only categorical variable in our dataset. The primary sectors of the institutions in our sample data are private nonprofit 4-year or above, public 2-year, public 4-year or above, private for-profit 4-year or above, private for-profit 2-year, public 4-year or above.

Table 1. Descriptive Statistics of Numerical Variables

	Full-time faculty per 100 FTE student	Part-time faculty per 100 FTE student	FTE 12-month enrollment	Number of aid	Student service cost	Education spending per degree
mean	4.47	6.64	3345.52	438.47	\$5,414,880.00	\$73,347.33
std	2.05	3.91	3335.41	404.22	\$5,042,739.00	\$35,437.17
min	1.88	1.75	180.00	26.00	\$267,886.40	\$33,720.98
25%	2.73	3.50	746.00	111.00	\$1,371,725.00	\$46,593.56
50%	4.13	5.80	1966.00	296.00	\$3,529,619.00	\$62,786.19
75%	5.81	9.01	4900.75	656.00	\$8,207,031.00	\$90,066.95
max	8.61	14.72	10170.00	1262.00	\$15,450,760.00	\$153,727.84

	GDP percent	White	Black	Asian	Hispanic	Non-resident	Full-time retention rate	Student service share
mean	3.50	59.17%	12.35%	2.38%	6.20%	1.43%	66.64%	14.75%
std	2.73	20.25%	10.82%	1.78%	5.11%	1.58%	13.26%	6.05%
min	0.70	23.28%	1.88%	0.44%	1.07%	0.00%	46.00%	6.00%
25%	1.50	43.01%	3.63%	0.82%	1.89%	0.00%	56.00%	9.76%
50%	2.70	60.54%	8.53%	1.76%	4.38%	0.74%	66.00%	13.89%
75%	4.50	76.53%	16.89%	4.14%	8.70%	2.92%	78.00%	19.49%
max	9.00	86.17%	35.61%	5.76%	17.03%	4.57%	88.00%	25.16%

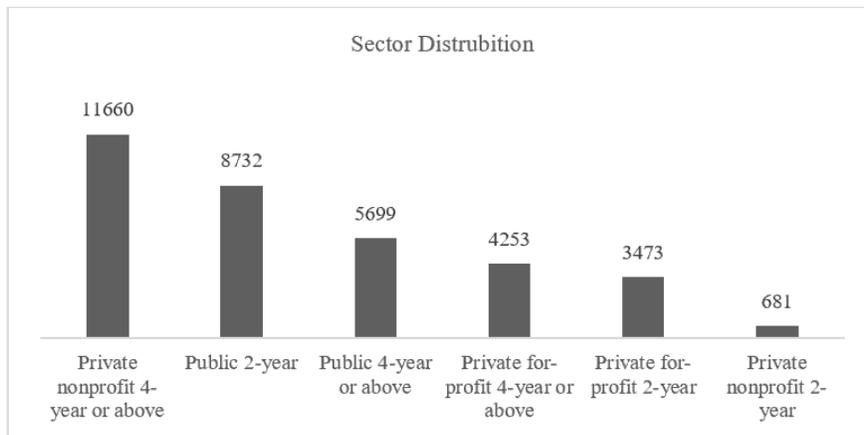


Figure 1. Sector Distribution

RESULTS

Measures

Mean absolute percentage error (MAPE) is commonly used in predicting quantitative data. With the advantage of scale-independent, MAPE is a good measure for forecasting applications across different datasets, especially when

the available data is enough (Armstrong and Collopy,1992; Hyndman and Koehler, 2006). Therefore, We adopted MAPE in our random forest model to assess the model performance. The smaller the MAPE is, the more accurate a model is. MAPE Formula:

$$M = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

Model

According to the MAPE shown in Table 2, we found that some of the predictor variables in our hypothesis are insignificant in determining the institutional degree productivity. Specifically, the MAPE of the full model (1), including all predictor variables, is 24.12%. The MAPE becomes slightly smaller to 22.55% when we reduced the *full-time retention rate, percentage of Asian students, the percentage of Hispanic students and the percentage of non-resident students* variables from the full predictor variables. We found that the reduced model (2) produced a higher accuracy in predicting the *educational spending per degree* compared with the slight change of MAPE between two models. Variables including *full-time retention rate, percentage of Asian students, the percentage of Hispanic students, and the percentage of non-resident students* insignificantly affect the institution’s degree productivity.

Table 2. Random Forest Model Comparison

No	Model	Independent Variables	Dependent Variables	MAPE
(1)	Random Forest Full Model	Education spending per degree	Full-time faculty per 100 FTE students, Student service cost, FTE 12-month enrollment, White, Student service share, Number of aid, GDP percent, Part-time faculty per 100 FTE students, Sector, Black, Full-time retention rate, Asian, Hispanic, Non-resident	24.12%
(2)	Random Forest Reduced Model		Full-time faculty per 100 FTE students, Student service cost, FTE 12-month enrollment, White, Student service share, Number of aid, GDP percent , Part-time faculty per 100 FTE students , Sector, Black	22.55%

Figure 2 exhibits significant features of the full model and reduced model, showing consistent importance among features. Both models indicate that full-time faculty per 100 FTE students is the most significant determinant of degree productivity, followed by student service expenses, FTE 12-month enrollment, and the percentage of White students. Besides that, by comparing the important features of the full model and the reduced model, the result is consistent with MAPE. Variables including *full-time retention rate, the percentage of Asian students, the percentage of Hispanic students, and the percentage of non-resident students* are the least significant among 14 dependent variables in the full model. The importance of other features shown in the reduced model is improved by up to 3% by excluding these insignificant variables. Besides, according to the reduced model, the share of student service, number of aids, GDP percent by state, part-time faculty per 100 FTE student, sector, and the percentage of Black students are slightly significant to determine degree productivity.

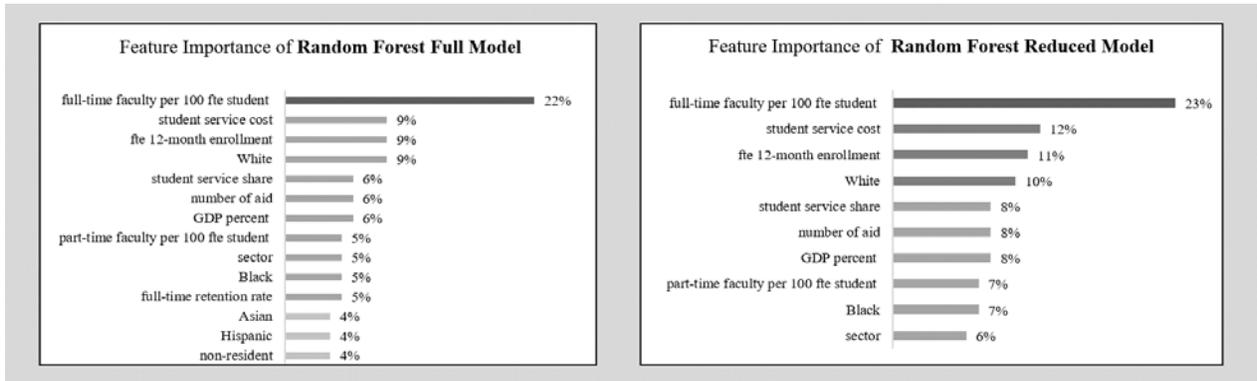


Figure 2. Difference Between Random Forest Full Model And Reduced Model

DISCUSSION

We can identify the significant features of our random forest reduced model that influence institutional degree productivity. To understand how the predictor variables related to the degree productivity, we introduced partial dependence techniques using the generalized additive model (GAM) smoothing with a 95% level of confidence to capture important patterns between significant features and degree productivity.

Among ten variables in the reduced model, full-time faculty per 100 FTE students is the most significant in determining degree productivity. By looking into the trend in Figure 3, the increase in full-time faculty per 100 FTE students leads to greater education and related expenses per degree when there are over five faculty per 100 FTE students. In a small class, students are expected to gain more learning with more interaction and participation. Bound et al. (2010) indicated that higher faculty-student ratio increases the likelihood of degree completion, but the higher faculty-student ratio requires more institutional spending in faculty. Therefore, it is impractical for an institution to hire too much faculty in order to increase faculty-student ratio and thus increase their degree completion rate. Our research tends to show that average four to five full-time faculty per 100 FTE students is a proper investment for an institution, as it finds a great balance between student degree completion and faculty spending to produce higher degree productivity.

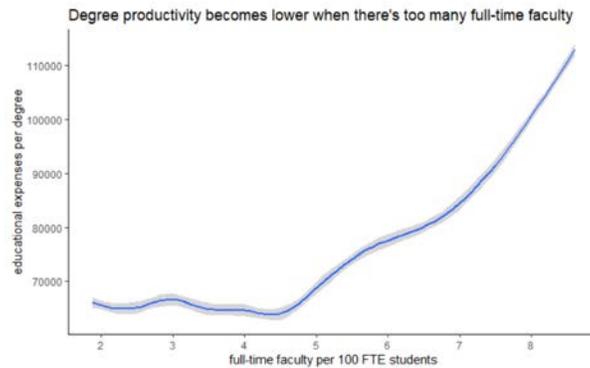


Figure 3. Trend Between Full-Time Faculty Per 100 FTE Students And Degree Productivity

The second most significant feature in our model is the expenses of student services. Besides that, the share of education and related spending on student services is slightly significant to degree productivity. Student services provide students with support in many aspects, such as admissions, recruitment, and student-related activities. It is an essential part of an institution to improve students' experience in an institution, but it is not directly related to students' academic achievement. According to Figure 4, spending too little or too much on student services results in lower degree productivity. When student services spending is over \$1,000,000 and less than around \$4,000,000 per academic year, the growth of student services spending tends to increase degree productivity, which is also consistent with results of studies conducted by Douglas and Ronald (2010). They found a significant result that the growth of student

services spending is more likely to increase student’s graduation rate, and reallocating some funding from instruction to student services may positively influence students’ persistence and graduation rates at some institutions. However, when the budget is over \$4,000,000, the increase in student services may have a negative effect on degree productivity. Figure 4 also displays that 7% to 10% of education and related spending on student services is more appropriate for institutions to pursue high degree productivity. The degree productivity becomes increasingly lower with the growth of education and related spending on student services if the share is over 10%. A possible explanation for this result is that given a limited budget on education and related expenses, more sharing of spending on student service would result in fewer budgets and lower quality on other parts directly related to student's academic outcome and degree completion, such as faculty and research. Therefore, we will suggest an institution to properly allocate its spending on student services at no more than 10% of its educational spending.

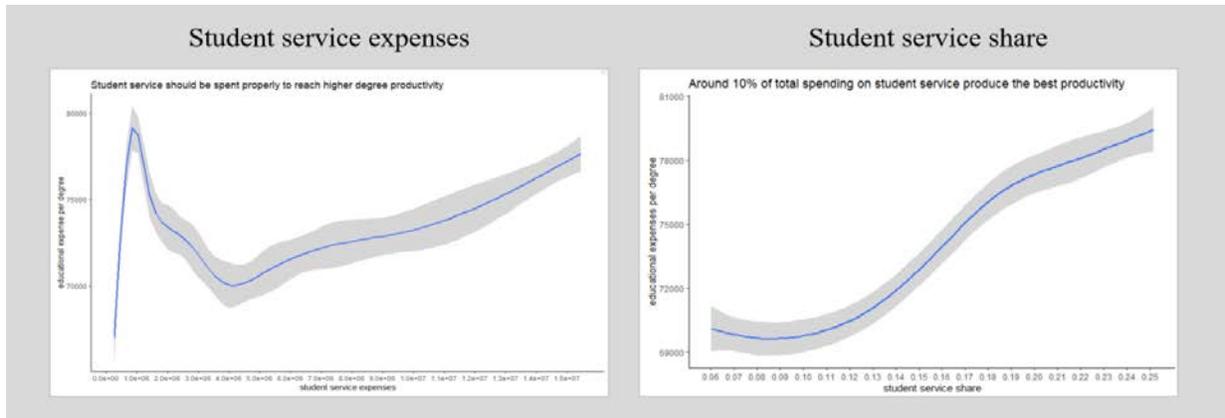


Figure 4. Trend Between Student Service And Degree Productivity

Moreover, institutional size has been studied in previous research on degree completion, but there are inconsistent results concluded from different researchers (Titus, 2004; Pascarella & Terenzini, 2005; Bailey et al., 2006; Calcagno et al., 2008;). Our results tend to support the findings of Titus (2004), who found a positive impact of institutional size on the students’ likelihood of persistence for 4-year institutions. Figure 5 shows that larger institutions are more likely to produce higher institutional degree productivity relative to small institutions. However, when there are too many students enrolled in an institution, the educational expenses per degree tend to rise. Many factors could explain the decrease in degree productivity, such as poor management, lower retention, poor teaching quality, and enormous fiscal spending. Therefore, to reach higher degree productivity, proper institutional size is around 6000 to 9000. A large institution should control its enrollment to avoid low productivity.

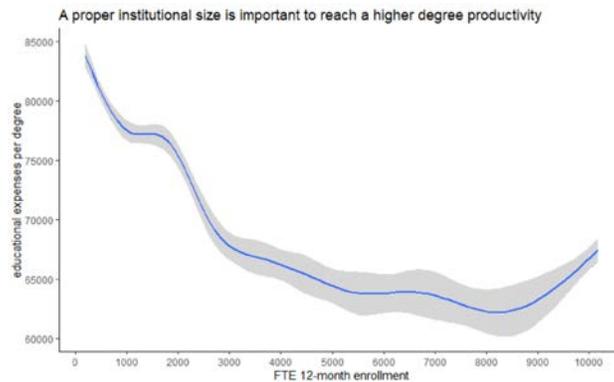


Figure 5. Trend Between Institutional Size And Degree Productivity

According to our findings, White students are the majority among all racial groups, taking up average 59.17% of students in a higher educational institution, according to *The Delta Cost Project Database*. Additionally, the percentage of White students is moderately significant to degree productivity. As the graph is shown in Figure 6, with

the increase in the proportion of White students, the degree productivity tends to increase. The results of previous research can explain this finding. It is found that White students are more likely to complete a degree than other groups (Astin, 2005; Titus, 2006). Larger percentages of White students lead to more degree completion in an institution and thus increase degree productivity.

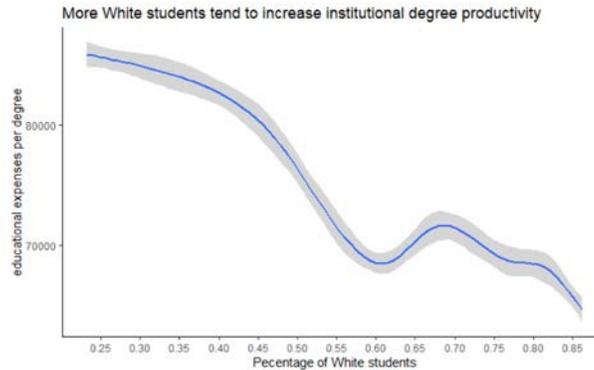


Figure 6. Trend Between Proportion Of White Students And Degree Productivity

EXPLICATIONS

This study contributed to the literature review in several aspects. First, institutional productivity always aroused great concern from a broad range of stakeholders and has been studied for decades in literature. For example, Harris (2013) applied cost-effectiveness analysis to address degree productivity issues. Adopting machine learning shows the patterns in degree productivity. Using a random forest model with 100 decision trees creates a robust predictive model with 77.45% accuracy rate. This study also contributes to the studies of degree productivity by applying data visualization techniques to provide deep insights into the factors that influence degree productivity, including feature importance and partial dependence.

Second, this study further explains results of previous empirical studies by highlighting the impact of full-time faculty per 100 FTE students, part-time faculty per 100 FTE, and GDP percent by the state on degree productivity. Previous research tends to consider the student-faculty ratio as a whole (Harris, 2013). However, in our study, we differentiated full-time faculty and part-time faculty, and thus we imported two measures, including full-time faculty per 100 FTE students and part-time faculty per 100 FTE to represent student-faculty ratio together. Our research shows a significant difference between full-time faculty per 100 FTE students and part-time faculty per 100 FTE as the importance of full-time faculty per 100 FTE students on determining degree productivity (23%) is three times larger than that of part-time faculty per 100 FTE (7%). Moreover, we also found the GDP percent by state extracted from ArcGIS database slightly impacted degree productivity. This geographical characteristic is among the first imported in the studies of degree productivity, indicating how a state's economic condition influences degree productivity. This study supplements new features that determine degree productivity and enriching the nomological network among general institutional characteristics and degree productivity.

CONCLUDING REMARKS

The finding of the study will help higher education administrators and policymakers to identify essential determinants that could improve degree productivity and understand how to effectively allocate limited educational and related spending to achieve a better outcome on student degree completion. One of findings suggests that institutional administrators can keep its ratio of full-time faculty to students at the level of average four to five faculty per 100 FTE students, to achieve a higher efficiency of its investment in faculty. Additionally, institutional administrators and college policymakers can control the spending on student services at around 7% to 10% of its total education and related spending in order to reach higher degree productivity. In terms of institutional size, to achieve higher education degree productivity, the small institutions can increase their student enrollment, and large institutions can control their enrollment at a stable level, around 6000 to 9000 FTE enrollment per academic year.

REFERENCES

- Astin, A. (1975). Preventing Students from Dropping Out. San Francisco, Jossey-Bass Publishers.
- Astin, A. W. (2005). Making sense out of degree completion rates. *Journal of College Student Retention: Research, Theory & Practice*, 7(1), 5-17.
- Astin, A. W., & Oseguera, L. (2005). Pre-college and institutional influences on degree attainment. *College student retention: Formula for student success*, 245-276.
- Bailey, T., Calcagno, J. C., Jenkins, D., Leinbach, T., & Kienzl, G. (2006). Is student-right-to-know all you should know? An analysis of community college graduation rates. *Research in Higher Education*, 47(5), 491-519.
- Baumol, W. J., & Blackman, S. A. B. (1995). How to think about rising college costs. *Planning for Higher Education*, 23(4), 1-7.
- Binkley, C. (2020, May 4). Unimpressed by online classes, college students seek refunds. Retrieved from <https://www.fox5atlanta.com/news/unimpressed-by-online-classes-college-students-seek-refunds>
- Berger, J. B. (2000). Organizational behavior at colleges and student outcomes: A new perspective on college impact. *The Review of Higher Education*, 23(2), 177-198.
- Bound, J., Lovenheim, M. F., & Turner, S. (2010). Why have college completion rates declined? An analysis of changing student preparation and collegiate resources. *American Economic Journal: Applied Economics*, 2(3), 129-57.
- Bowen, W. G., Chingos, M. M., & McPherson, M. S. (2009). Crossing the finish line: Completing college at America's public universities. *Princeton University Press*.
- Calcagno, J. C., Bailey, T., Jenkins, D., Kienzl, G., & Leinbach, T. (2008). Community college student success: What institutional characteristics make a difference?. *Economics of Education review*, 27(6), 632-645.
- Jenkins, D., & Rodriguez, O. (2013). Access and success with less: Improving productivity in broad-access postsecondary institutions. *The Future of Children*, 187-209.
- Ehrenberg, R. G., & Zhang, L. (2005). Do tenured and tenure-track faculty matter?. *Journal of Human Resources*, 40(3), 647-659.
- Harris, D. N. (2013). Addressing the declining productivity of higher education using cost-effectiveness analysis.
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4), 679-688.
- Armstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International journal of forecasting*, 8(1), 69-80.
- Jacoby, D. (2006). Effects of part-time faculty employment on community college graduation rates. *The Journal of Higher Education*, 77(6), 1081-1103.
- Kahu, E. R., & Nelson, K. (2018). Student engagement in the educational interface: Understanding the mechanisms of student success. *Higher Education Research & Development*, 37(1), 58-71.
- Massy, W. F., Sullivan, T. A., & Mackie, C. (2013). Improving measurement of productivity in higher education. *Change: The magazine of higher learning*, 45(1), 15-23.

- Miles, P. C., Peterson, M., Miles, G., & Bement, D. (2018). Higher education: exploring productivity over time. *Journal of Applied Research in Higher Education*.
- Oseguera, L. (2005). Four and six-year baccalaureate degree completion by institutional characteristics and racial/ethnic groups. *Journal of College Student Retention: Research, Theory & Practice*, 7(1), 19-59.
- Perna, L. W. (1998). The contribution of financial aid to undergraduate persistence. *Journal of Student Financial Aid*, 28(3), 25.
- Pew's analysis of Fiscal Federalism (2018), "Two Decades of Change in Federal and State Higher Education Funding".
- Terkla, D. G. (1984). Does Financial Aid Enhance Undergraduate Persistence?.
- Titus, M. A. (2004). An examination of the influence of institutional context on student persistence at 4-year colleges and universities: A multilevel approach. *Research in higher education*, 45(7), 673-699.
- Titus, M. A. (2006). Understanding college degree completion of students with low socioeconomic status: The influence of the institutional financial context. *Research in Higher Education*, 47(4), 371-398.
- Toutkoushian, R. K., & Smart, J. C. (2001). Do institutional characteristics affect student gains from college?. *The Review of Higher Education*, 25(1), 39-61.
- U.S. Department of Education, National Center for Education Statistics (2019). "Postsecondary Institution Expenses" The Condition of Education 2019 (NCES 2019-144).
- Webber, D. A., & Ehrenberg, R. G. (2010). Do expenditures other than instructional expenditures affect graduation and persistence rates in American higher education?. *Economics of Education Review*, 29(6), 947-958.