

Loan Outcomes Among Pell Grant Recipients: A Comparison of HBCUs and Non-HBCUs

1 Introduction

Student financial aid is one of the key factors that sustain the American higher education system. The Pell Grant is a need-based federal grant for undergraduate students with no repayment requirement, designed to help students from low-income families afford college. However, Pell Grant recipients often need to apply for loans to cover the remaining costs of their education (Ma et al., 2014). The growing reliance on student loans has raised concerns about the long-term financial stability of these students, as high student loan debt is linked to increased default rates. This research seeks to explore the disparities in loan default rates and repayment outcomes across different types of institutions, particularly comparing Historically Black Colleges and Universities (HBCUs) with non-HBCUs. By examining these differences, the study aims to uncover underlying economic inequalities and provide insights that could help reduce the financial burdens on affected students.

Studies have shown that higher borrowing rates, combined with insufficient post-graduation income, lead to an increased risk of loan default (Looney & Yannelis, 2015). However, there is a gap in the literature regarding the differences in outcomes for Pell Grant recipients attending different types of institutions. This research addresses this gap by focusing on whether the disparities in loan default rates between HBCUs and non-HBCUs can be explained by institutional differences. Through this study, we hope to gain insights into the role institutional context plays in shaping student financial outcomes and repayment success.

2 Motivation

Although Pell Grants provide vital financial relief to low-income students, they often do not cover the full cost of attending college, forcing students to supplement their aid with loans. This study examines the impact of additional borrowing, referred to as Pell Grant debt, on loan default and repayment rates. One key focus is whether the relationship between loan debt and

repayment outcomes differs between HBCUs and non-HBCUs. Historically, HBCUs serve a large population of low-income students who rely heavily on federal loans to finance their education (Lee et al., 2013). We aim to investigate whether there is a link between the amount of debt students incur and their success in repaying those loans, while considering how institutional support services and other factors might affect these outcomes.

Furthermore, we plan to incorporate key institutional characteristics, such as net cost, completion rates, and post-graduation earnings, to better understand the factors that influence loan default and repayment rates. By including or excluding these characteristics in our analysis, we hope to determine the direct relationship between loan debt among Pell Grant recipients and institutional type. Our goal is to identify the specific factors that contribute to financial instability for students, ultimately helping to create more equitable financial aid policies for all students. Understanding these disparities can lead to improved financial stability and greater academic success for low-income Pell Grant recipients, regardless of whether they attend HBCUs or non-HBCUs.

3 Research Questions

In this project, we aim to address the following research questions:

1. How does the loan default rate vary between Pell Grant students in HBCUs and non-HBCUs, or is it relatively the same across all universities?
2. How does the loan debt differ by HBCU status?
3. How is Pell Grant debt associated with loan default and loan repayment rates, and does this association differ for HBCUs versus non-HBCUs after controlling for other institutional characteristics such as net cost, admissions rate, completion rate, retention rate, and median earnings?
4. Can we predict loan default risk using student level data?

4 Expected Outcomes

We expect our analysis to show a clear association between high student loan debt among Pell Grant recipients and an increase in loan default rates

across institutions. Specifically, we aim to determine whether the loan default rate differs between Pell Grant students attending HBCUs and non-HBCUs. We hypothesize that students attending HBCUs will generally have higher loan default rates due to higher levels of debt, exacerbated by institutional underfunding. Additionally, we expect to find that the amount of loan debt influences repayment success, with students facing higher debt likely to have more difficulty in repaying their loans. Furthermore, we anticipate that by controlling for other institutional factors, we will gain a deeper understanding of the relationship between Pell Grant debt and financial outcomes, which will inform the creation of more equitable financial aid policies for low-income students across different educational settings.

5 Data Description

The data we used in this project is from the U.S. Department of Education College Scorecard website. College Scorecard is a project designed to increase transparency about postsecondary institutions in the United States and enable students and families to compare how well individual postsecondary institutions are preparing their students for success. It also helps them compare college costs and outcomes while accounting for their needs and educational goals. The data in this project were provided through federal reporting from institutions, data on federal financial aid, and tax information.

College Scorecard provides two data files: institutional level data and field of study data. Due to the limited nature of field of study data, we focus on institutional level data for our analysis. To use this data, we downloaded the zip files of the dataset, extracted them as a CSV file, and used Python data libraries (NumPy, Pandas, Matplotlib, and Seaborn) to clean and prepare the data. College Scorecard provides documentation for the dataset, and we used this documentation to learn about the variables in the data and the research questions that can be answered from the data.

The institutional level data about institutions contains 3305 variables and 6484 observations. The variables in the dataset ranged from basic information about the college (name, location, degree type, etc) to information about cost, financial aid, loan default rate and repayment, earnings of students, completion rate, etc. Since we were only interested in loan debt, default and repayment rate for Pell Grant students in HBCUs and non-HBCUs, we

removed variables that weren't useful for our research purposes. Hence, we reduced the number of entries to 45. We kept variables, such as the loan repayment rate for Pell Grant students, completion rate for black students, loan default rate for Pell and non-Pell Grant students, etc., and we removed variables, such as count of students by race, location of school, URL of schools, etc.

Some of the variables (e.g., PELL_DEBT_MDN, ADM_RATE, etc.) had to be converted into numerical variables so we could make box and bar plots with them. Some of the variables had inconsistent observations. For example BBRR4_FED_PELL_DFLT variable had some missing values, inequality symbols, extraneous characters, range of figures, etc. To clean these variables, we defined a cleaning function that removed all of these inconsistencies. Moreover, since we are only dealing with HBCUs and non-HBCUs, we set every other school which were not HBCU (schools that had observation of zero under the HBCU column) to non-HBCU schools. We also changed the variables with vague names to easily understandable ones. For example, BBRR4_FED_PELL_DFLT and BBRR4_FED_PELL_PAIDINFULL variable names were changed to Default_Rate and Repayment_Rate, respectively.

The following is a list of the variables used in our analysis and their definitions:

HBCU Historically Black Colleges and Universities.

INSTNM Institution's name.

BBRR4_FED_PELL_DFLT Default rate for Pell Grant students on BBRR (Borrower Based Repayment Rate) for federal loans, measured after four years of entering repayment.

BBRR4_FED_PELL_PAIDINFULL Rate of Pell Grant students who paid their federal loans in full on BBRR, measured after four years of entering repayment.

PELL_DEBT_MDN Cumulative median student debt for Pell Grant recipients.

COSTT4_A Average cost of attendance, including tuition and fees, for academic year institutions.

ADM_RATE Admission rate.

C150_4_PELL Completion rate for Pell Grant students who graduate within 150% of the expected time (6 years for 4-year institutions).

RET_FT4 Retention rate for full-time students in 4-year institutions.

MD_EARN_WNE_P6 Median earnings of the entry cohort 6 years after enrollment.

6 Method

The first question we attempted to answer was how loan default and repayment rates vary between Pell Grant students in HBCUs and non-HBCUs. To answer this question, we first made a box plot showing the distribution of default rates among Pell Grant recipients in HBCU and non-HBCU institutions. We used box plots because they help us to easily understand the distribution of data and identify outliers. They also help us to visualize the variability in data, skewness, and measures of central tendency. Hence, we could get a good visualization of the difference in variability between these default rates. We also did the same for the repayment rates among Pell Grant recipients across these two categories of universities (HBCUs and non-HBCUs). For a clearer illustration, we also made a bar plot of the mean default and repayment rates for Pell Grant students between HBCUs and non-HBCUs. After seeing a difference in the average default and repayment rates, we wanted to know if this difference is statistically significant. Hence, we did a t-test to compare the mean of these two groups (HBCUs and non-HBCUs) for both the default and repayment rates. Next, we built a linear regression model to examine the relationship between HBCUs and default rates (as well as repayment rates). This helps us to know how these rates differ on average for HBCUs compared to non-HBCUs, just like the t-test. It also helps us to know, on average, the increase in these rates (in percentage points) between HBCUs and non-HBCUs. Unlike the t-test, we can draw insights from R-squared values and F-statistics values given from these models, and we can add more variables (e.g., median earnings, completion rate) to control more factors.

For our second question, we wanted to find out how the loan debt differs by HBCU status for Pell Grant students. To answer this question, we used the

median debt for Pell Grant students (PELL_DEBT_MDN). Again, we used a box plot to show the distribution of loan debt across these institutions. We saw some differences in these distributions, but they were not as large as the default and repayment rates. To find out if these differences were statistically significant, we conducted a Welch's t-test as we did for the first question. Then we built a linear regression model for the same reasons we gave for the above.

Next, we explored how the loan debt we investigated in question two is associated with the loan default and repayment rates (which we explored in question one). First, we used a scatter plot to show the relationship between loan debt and these rates. We used scatter plots so we could visually see relationships between these rates and loan debt and spot group differences. We also wanted to check if we could see a linear relationship between these variables before we used multiple linear regression with interaction terms. Next, we proceeded to building the multiple linear regression model. Before building this model, we had to control for other institutional variables, such as net cost, admission rate, completion rate, retention rate, and median earnings. Hence, our independent variable changed from just HBCU to the following:

$$\text{PELL_DEBT_MDN} + \text{HBCU} + \text{PELL_DEBT_MDN} * \text{HBCU} + \text{COSTT4_A} + \text{ADM_RATE} + \text{C150_4_PELL} + \text{RET_FT4} + \text{MD_EARN_WNEP6}$$

The definitions of these variables have been given in the **Data Description** section.

By using multiple regression, we were able to account for multiple factors that could affect the loan debt and default and repayment rates. More so, we were able to build an improved model that gave us a deeper understanding of the complex relationship between loan debt and default and repayment rates.

For the fourth research question, we wanted to build a predictive machine learning model that could predict the risk of loan default for students based on student level characteristics. We hypothesize that such a model would be used by loan providers in assessing the risk of a student loan default. To test the feasibility of such model, we intended to build a comprehensive model that would take in features such as student's academic performance, their family income, and the characteristics of the institutions they plan to at-

tend. We include variables such as Pell Grant Status, and whether a student is attending a HBCU in our model, as it relates with our previous research questions on whether these factors impact student loan default rates. Since we do not have student level data to build the proposed model, we decided to use synthetic data that is simulated with assumptions that match the real world data.

We use the student’s GPA, and SAT score as a proxy for academic performance. These variables are generated using a normal distribution of average GPA and SAT Scores of US students. We use the following income levels to classify students into different income groups as an indicator of their financial stability: (1) \$0–30,000 (2) \$30,001–48,000 (3) \$48,001–75,000 (4) \$75,001–110,000 (5) \$110,000+. The students are randomly assigned into one of these groups with each group’s probabilities being 30%, 25%, 20%, 15% and 10% respectively. A binary variable to indicate whether a student is a Pell Grant recipient. This variable is categorized as 1 with a 90% probability if the student’s family income level is \$0–30,000, otherwise, it is categorized as 1 with a 20% probability. Next, we use institutional-level variables for the institution the student is attending, such as net cost, admission rate, completion rate, and median earnings after graduation. Each student is randomly assigned to either an HBCU or a non-HBCU institution using weighted probabilities that reflect the real-world proportion of such schools. After classification, we generate institution-level variables using normal distributions. These distributions are set up differently for HBCU and non-HBCU schools, based on statistics from the College Scorecard dataset for each group. Once all the predictor variables are generated, we define a function to calculate the log odds and probabilities of student default based on the predictor variables ($\log_odds = (-3 + (2.50 - \text{synthetic_df}['GPA']) + 0.00015 * \text{synthetic_df}['COSTT4_A'] + 0.25 * (4 - \text{synthetic_df}['Income_Level_Encoded']) + 0.40 * \text{synthetic_df}['Pell_Recipient'] + 0.40 * \text{synthetic_df}['HBCU'])$). Then, the outcome variable, Loan Default, is stimulated as a binary variable based on each student’s risk probability of default. With a synthetic data ready, it is split into train and test the dataset and a logistic model is fit on the data. The logistic regression model is used to predict the risk of loan default risk for students based on the predictor variables. While we recognize the limitations of using synthetic data, this approach allowed us to explore the feasibility of predictive modeling in this context and simulate how such a model could be applied in real-world scenarios.

7 Results

Our analysis examined key differences in loan outcomes and debt levels among Pell Grant recipients between HBCUs and non-HBCUs. We focus on four research questions: two comparing default and repayment rates, one assessing median debt differences, and one exploring the association between debt and loan outcomes while controlling for other institutional characteristics, concluding with a proof-of-concept predictive model based on synthetic student data.

7.1 Differences in Default and Repayment Rates by Institution Type

The box plots and bar charts illustrate that Pell Grant recipients at HBCUs exhibit notably higher default rates and lower repayment rates compared to their non-HBCU counterparts. A t-test confirmed a statistically significant difference in default rates ($t = 11.624$, $p < 0.001$). An OLS regression of the default rate on the HBCU indicator yielded a coefficient of 0.100 ($p < 0.001$), meaning that, on average, HBCU students experience a 10 percentage point higher default rate relative to non-HBCU students. Similarly, the repayment rate was significantly lower in HBCUs, as shown by a t-test ($t = -6.271$, $p < 0.001$) and the corresponding regression demonstrated a statistically significant negative coefficient (-0.0508 , $p < 0.001$). These results underscore clear discrepancies in loan outcomes between the two groups.

7.2 Differences in Median Loan Debt

The analysis of median Pell Grant debt revealed that although the absolute differences in borrowing between HBCUs and non-HBCUs are modest, they are statistically significant. A t-test indicated a significant difference in median debt ($t = 3.148$, $p = 0.003$). The OLS regression analysis estimated that the median debt at HBCUs is approximately \$1,660 higher than that at non-HBCUs ($p = 0.006$). This modest debt difference, however, contrasts with the larger gap observed in default rates, suggesting that factors beyond debt magnitude may contribute to the heightened risk among HBCU students.

7.3 Association Between Pell Grant Debt and Loan Outcomes

To further explore how debt levels are associated with loan default and repayment rates while controlling for institutional characteristics (e.g., net cost, admissions rate, completion rate, retention rate, and post-graduation earnings), we estimated multiple linear regression models. For default rates, the model ($R^2 = 0.495$) reveals that, controlling for these factors, the association between increasing debt and default rate is statistically significant. Here the main effect of HBCU status is not statistically significant (coef = 0.0234, $p = 0.378$) while the interaction between Pell debt and HBCU status shows a positive coefficient of 2.42×10^{-6} and is marginally significant ($p = 0.078$). A similar pattern emerged in the repayment rate model, where the model ($R^2 = 0.497$) reveals that an increase in median Pell debt reduces the repayment rate by about 2.87×10^{-6} ($p < 0.001$). Here, the direct effect of HBCU status (coef = 0.0337, $p = 0.346$) and its interaction with Pell debt (coef = -1.649×10^{-6} , $p = 0.372$) are not statistically significant, indicating that the moderating role of institution type does not extend to repayment outcomes.

7.4 Predicting Loan Default Risk Using Synthetic Student-Level Data

Recognizing that individual-level data are ideal for predicting default risk, we generated synthetic data to simulate key student and institutional characteristics. A logistic regression model was fitted to predict the probability of loan default using predictors such as GPA, SAT score, family income level (encoded), Pell Grant status, and institutional attributes including HBCU status. The model achieved a 90% accuracy on the test set, and the coefficient estimates indicated that higher GPA is associated with a lower likelihood of default. Notably, both Pell Grant receipt and HBCU status were identified as significant predictors, reinforcing our findings from the aggregate analyses.

Overall, our results demonstrate that while median debt differences between HBCU and non-HBCU institutions are relatively modest, significant discrepancies exist in both default and repayment rates. The analyses suggest that factors associated with HBCU attendance, potentially reflecting broader resource and support disparities, are linked with higher default risk among Pell Grant recipients. Finally, our predictive modeling underscores the feasibility of using student-level data to anticipate loan default, provid-

ing a foundation for further investigation into the causal mechanisms that underlie these outcomes.

8 Discussion

The research found that Pell Grant Recipients at HBCUs had greater loan and debt amounts, as well as default rates, compared to their counterparts in non-HBCUs. The research found an interesting discrepancy. The average median debt for Pell Grant Recipients at HBCUs was only slightly higher than that of students at non-HBCUs, but HBCU student default rates were almost twice as high.

The differences between the default rate and the repayment rate among Pell Grant recipients in HBCUs versus non-HBCUs were both statistically significant. Although the AIC values for the chosen models were satisfactory, the adjusted R^2 values are of concern. The median amount of debt for Pell Grant recipients in HBCUs was also found to be statistically significant; however, the constructed regression model may require further investigation, as both the AIC values and the adjusted R^2 values suggest possible overfitting.

When controlling for other variables such as net cost, admissions rate, completion rate, retention rate, and median earnings, model fits improved significantly. Nonetheless, a notable concern is that in both models the HBCU indicator was not statistically significant, which potentially points to an area for further research.

Synthetic student-level data were used to predict the odds of defaulting, and the model yielded favorable results. However, given that many confounding characteristics may affect students at HBCUs versus non-HBCUs, the predicted statistic may not be accurate. As expected, the schools with more financially needy students had more students struggle to repay their debt amounts and often received higher amounts of loans. Although there are clear discrepancies in debt amounts and default rates between Pell Grant Recipients in HBCUs and non-HBCUs, this analysis could not pinpoint a causal relationship between HBCU enrollment and signs of post-graduate financial struggles. It does, however, suggest that the lack of resources at HBCUs may result in lower amounts of financial aid awards to students already receiving a Pell Grant. For example, the median debt amount for a Pell Grant Recipient at Vassar College is \$12,531, significantly less than the average median debt amount for Pell Grant Recipients at HBCUs, even though the average

cost to attend an HBCU is nearly \$60,000 less than that of Vassar College.

Ideally, student-level information regarding debt amounts and postgraduate income is available. However, given the proprietary nature of an income dataset, only school-level information was available. Even if student-level data were available, debt repayment is a matter of financial intelligence and awareness, which cannot be quantified without highly proprietary information. As we also do not have intimate knowledge regarding how schools give aid and how their resources are used, it is difficult to pin responsibility on the school or the endowments that they receive. Based on the data available, however, we discovered that there was a significant discrepancy in endowment amounts between HBCUs and PWIs, as they averaged approximately 58millionand 385 million, respectively.

Ultimately, further research would have to delve into student-level data to find a causal relationship between a student's enrollment status at an HBCU and their ability to repay their loans. That does, however, face additional difficulties regarding proprietary information laws and maintaining data privacy. However, if research can find a reason why students at HBCUs struggle to repay their loans, it could potentially improve the lives of millions of Americans in helping them find financial security and decrease financial burden post graduation.

References

- [1] Looney, A., & Yannelis, C. (2015). A crisis in student loans?: How changes in the characteristics of borrowers and in the institutions they attended contributed to rising loan defaults. *Brookings Institution*.
- [2] Lee, J. M., & Keys, S. W. (2013). Repositioning HBCUs For the Future: Access, Success, Research & Innovation. *APLU Office of Access and Success Discussion Paper 2013-01*. Washington, DC: Association of Public and Land-grant Universities.
- [3] Ma, J., Baum, S., Pender, M., & Bell, D. (2014, November 30). Trends in college pricing, 2015. *Trends in Higher Education Series*. College Board. Retrieved from <https://eric.ed.gov/?id=ED572540>