Predictors of Student Loan Repayments: A Comparison Between Public, Private For-Profit and Private Nonprofit Schools

Mannat Bakshi

Arjun Ahluwalia

Professor Genna Miller, Faculty Advisor

Professor Michelle Connolly, Seminar Instructor

Professor Kent Kimbrough, Seminar Instructor

Honors Thesis submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University.

Duke University

Durham, North Carolina

2020

Acknowledgements

We would like to thank our advisor, Dr. Genna Miller, for her guidance and encouragement this year. We would also like to thank Dr. Kent Kimbrough with his help in the early stages of our thesis and Dr. Michelle Connolly for her all her time and guidance. Through their help and constructive criticism, we were able to expand our knowledge of existing student loan literature, gain stronger econometrics skills, and contribute a new piece of work to the literature.

Abstract

Using a sample of over 3,500 colleges from the College Scorecard Dataset[1], we investigate the association of average federal student loan repayment rates with institutional, regional, and student demographic characteristics of colleges. We consider educational cohorts from 2010 to 2016 at public, private for-profit, and private non-profit institutions. Our data do not allow us to see individual student characteristics, hence we control for traits of the average student in each college and focus on institutional traits that impact repayment rates. Our controls for demographics are consistent with prior research on student loan repayment rates (Lochner and Monge-Naranjo, 2014; Kelchen and Li, 2017).

We ran a Random Effects panel regression to determine how institutional, regional, and student demographic characteristics impact repayment rates. We see an important influence of the institution attended. Institution selectivity (lower admission and withdrawal rates) is associated with higher average repayment. Furthermore, the highest degree awarded is a more significant variable when it comes to describing the variation in repayment rates for public schools; private for-profit schools exhibit lower repayment rates and private nonprofit schools exhibit higher repayment rates regardless of the highest degree awarded. This could be due to a combination of signaling and screening effects. Local income and unemployment impact repayment for the average student in public and for-profit schools, but not in private non-profit schools.

A noticeable institutional finding is that, even after controlling for average school demographics, for-profit schools exhibit lower repayment rates across all types of degree-granting programs. Attending a for-profit school may be a negative signal of ability or value to potential employers. Median family income positively affects repayment twice as much for for-profit schools compared to other school

[1] https://collegescorecard.ed.gov/data/

types. These finding on for-profit institutions help explain Obama's "crack down on for-profit career training colleges" (Simon & Emma, 2014).

JEL classification: I2; I22; I26

Keywords: Education, Federal Loans

1. Introduction

Using institutional-level data from 2010 to 2016, we identify how college traits influence student loan repayment rates. This paper seeks to understand how institutional characteristics (admission rate, cost of attendance, highest degree granted and withdrawal rate), student demographics (family income, race, gender, major choice, part-time status, Pell-grant status), as well as regional attributes (city population, city income and state unemployment rate) of each institution impact repayment rates for public schools, private for-profit, and private nonprofit schools.

Ignoring moral hazard issues, student loan repayments are impacted by earnings after graduation. As such, they are influenced by many factors, including but not limited to personal ability, professional field, signaling perceived by employers based on the institution attended, value of the education received (and/or perceived by employers), and macroeconomic conditions. In times of financial hardship, the ability to rely on family for additional help is likely to be more significant.

1.1 U.S. Student Loan Market

Although it is widely accepted that higher education increases the productivity of the labor force (Moretti, 2004), there is increasing concern that students are taking on too much debt to finance their higher education (Lochner and Monge-Naranjo, 2016). In 2004, there were 23 million student loan borrowers in total (including both private and federal loans), whereas in 2012, there were 39 million

Private nonprofit four-year institution		Public four-year institution	
Tuition for 1987-1988:	$15,160	Tuition for 1987-1988:	$3,190
Tuition for 1997-1998:	$21,020	Tuition for 1997-1998:	$4,740
Tuition for 2007-2008:	$27,520	Tuition for 2007-2008:	$7,280
Tuition for 2017-2018:	$34,740	Tuition for 2017-2018:	$9,970

Source: CollegeBoard

Figure 1: Tuition Costs over Time (inflation adjusted 2017 USD)

(Brown Haughwot & Lee, 2014). This 70% increase in borrowers is coupled with an increase in average debt per borrower. Individual borrowings have also increased 67% between 2004 and 2012, from approximately $15,000 to $25,000 (Brown, Haughwot & Lee, 2014).

Data from the CollegeBoard (figure 1) shows us that over the period 1997-98 to 2007-08, tuition and fees for full time in-state students at private non-profit four-year colleges and universities grew by around 31%. For full time students at public four-year colleges and universities, published tuition and fees rose approximately 54%. These rising educational costs are forcing more students to borrow (and borrow higher sums of money) in order to finance their education.

As of March 2019, about 92% of all student loans were federal — not private — loans (MeasureOne, 2019). Student loans are extended under Federal Student Loan programs administered by the Department of Education. These federal programs have historically offered loans at rates lower than those offered by most private lenders, on terms that are more attractive to student borrowers, and without adjusting the pricing on loans according to the risks inherent in different courses of study or lending to different types of borrowers (Simkovic, 2013). This highlights that individuals are neither screened for their ability nor the type of institution that they are attending or the degree that they will be pursuing. This imperfect knowledge in the market suggests that some students are borrowing more than they will be able to pay back.

There are rising delinquency rates (90 or more days late) on student loan payments (both federal and private loans) over the past ten years: among borrowers under the age of 30 still in repayment, the fraction of delinquent borrowers on student loans increased sharply from 20% in 2004 to 35% in 2012 (Brown, Haughwot & Lee, 2014). The delinquency rate of student loans has risen relative to rates on other types of consumer debt including credit card debt, auto debt, and mortgages. In late 2008, student loan debt surpassed credit card debt to have the highest rate of consumer loan delinquency, suggesting that the number of borrowers struggling with their student loan payments is increasing (Li, 2013).

1.2 Institution Type

According to the Brookings Institute, 52% of for-profit undergraduate borrowers defaulted on their student loan within 12 years of entering the school. This was followed by public 2-year colleges where 26% of borrowers defaulted on their loan. In a study that looked at New York schools, although just 6% of all undergraduate students in New York attended for-profit schools, these schools accounted for 41% of all student loan defaults after five years (Dvorkin, Bowles, & Shaviro, 2018). Furthermore, on average, for-profit schools are more expensive for every degree relative to other school types. According to Lee (2020), a certificate program (the most commonly offered program at private for-profit institutions) at a for-profit college costs $19,806 compared to $4,250 at a public college. Additionally, there is evidence that certificate-seeking students enrolled at for-profit schools are 1.5% less likely to be employed and, conditional on that employment, exhibit 11% lower earnings after graduation relative to those who attended public institutions (Cellini, 2016).

Despite information suggesting that for-profit schools are a larger and riskier investment, according to a summary of a senate hearing in 2012, for-profit colleges are rapidly increasing their reliance on taxpayer dollars. In 2009-10, the sector received $32 billion, 25 percent of the total Department of Education student aid program funds (U.S. Senate, 2012). More recently, senators and consumer advocacy groups have sought to use legislation to end funding for-profit schools. In October 2019, democrats Rep. Pramila Jayapal and Sen. Sherrod Brown introduced the “Students Not Profit Act”, which would ban the Department of Education from sending federal loans and other funding to for-profit colleges. Approximately one-third of for-profit institutions acquire almost all of their revenue from taxpayer money via government-backed student loans (Nova, 2019).

1.3 Institutional Characteristics

Previous studies (Lochner and Monge-Naranjo, 2014; Kelchen and Li, 2017) look at characteristics traditionally connected to economic drawbacks and their effect on loan default while controlling for certain institutional level characteristics, such as cost of attendance and university type. Traditional economically disadvantaged characteristics include being from an underrepresented minority group or from a lower income bracket. In this paper, we not only control for the aforementioned institutional level characteristics, but we also control for selectivity and withdrawal rate as well as regional characteristics, such as median city income and the local state unemployment rate. Additionally, we categorize the student bodies at the schools by race, family income, gender, part-time status, and major choice.

Lochner and Monge-Naranjo (2014) found that Black individuals have much higher nonpayment rates on debt relative to White individuals; interestingly this is a phenomenon not wholly explained by differences in post-school earnings, choice of major, type of institution, or student debt levels. Controlling for all other variables, including median family income and school type, we found that schools with greater relative Black and Hispanic borrowers populations had lower repayment rates. Regarding major choice, Lochner and Monge-Naranjo (2014) suggest variation exists in repayment/nonpayment depending on the student's degree. Unsurprisingly, engineering majors owe a much smaller portion of their debt relative to other majors after 10 years, whereas social science and humanities majors represent the greatest share of debt in nonpayment. In our paper, we find that having a larger portion of STEM degrees raises the repayment rates for both public schools and private non-profits.

Although existing literature and statistics do highlight that private for-profit institutions result in higher default rates, there is little explanation as to why this phenomenon exists and how for-profit institutions differ from public schools and private nonprofit schools. This paper aims to highlight some of the key institutional differences between the school types and repayment rates.

2. Literature Review

Although existing research has considered the ways that student demographics and institutional characteristics are related to student loan repayment and default rates, the combination of examining institutional factors by school type while controlling for average student demographics is something that has not been previously studied. The majority of the existing research is concerned with examining the impact of student demographics through the use of individual-level data. There is also little research on the regional effects of an institution on repayment rates, and this has not been studied in conjunction with the student demographics and institutional factors. Finally, using the multiple years of data provided in our College Scorecard dataset, we are also able to consider how institutional, student demographic and regional effects relate to student loan repayment rates over time.

2.1 Student Demographic Factors

The literature is consistent in finding that Black students have the highest rate of student loan default among other racial student groups (Hillman, 2015; Jackson and Reynolds, 2013; Lochner and Monge-Naranjo, 2014) even after controlling for post-graduation earnings (Lochner and Monge-Naranjo, 2004). Hillman (2015) also found that Hispanic students default on their loans at higher rates than White students. Lochner and Monge-Naranjo (2015) also found that the higher nonpayment rates by Black and Hispanic students are not wholly explained by differences in post-school earnings, choice of major, type of institution, or student debt levels. The reasons behind these discrepancies in repayment rates are not fully explained in the literature, although Volkwein et al. (1998) highlight that students of color are more likely to be unemployed. Additionally, Lochner and Monge-Naranjo (2004) point out that the post-college earnings of African Americans ar lower than that of all other racial groups, both of which could hinder their ability to repay their loans. With regard to gender, Chen and Wiederspan (2014) concluded that bachelor's degree recipients who are female graduated with more debt than males. Despite this finding, there is little information in the literature that can lead to a consistent conclusion regarding gender and student loan repayment.

Students from lower family incomes incur more debt than students from higher family incomes (Herr and Burt, 2005; Steiner and Teszler, 2005). There is also evidence that students from higher family incomes are less likely to default on their loans (Chen and Wiederspan 2014; Hillman 2015; Looney and Yannelis 2015). According to Gross, Cekic, Hossler & Hillman (2009), students who have higher family incomes benefit from having a financial safety net that can be particularly helpful when they are going through fluctuations in personal income. The same is not true for students with lower family incomes.

The literature highlights that STEM majors owe a much smaller portion of their debt relative to other majors after 10 years, whereas social science and humanities majors represent the greatest share of debt in nonpayment (Lochner and Monge-Naranjo, 2016). Moreover , Wiswall and Gemici (2011) highlight that there is an increasing importance of college major as a determinant of earnings.

2.2 Institutional Factors

Looney & Yannelis (2015) highlight that there have been increases in the number of borrowers at for-profit and public two-year community colleges. They find that these borrowers have led to a surge in student loan defaults as many of them had withdrawn from the institution, been unable to obtain a job post-graduation or were from economically disadvantaged backgrounds. This finding that for-profit colleges and two-year colleges have higher default rates than nonprofit or public four-year colleges is consistent across the literature (Hillman 2014; Lochner and Monge-Naranjo 2014). Ishitani and

McKitrick (2016) found that default rates at four-year colleges tend to be lower at more selective colleges and those with higher graduation rates. Kelchen and Li (2017) found that institutional characteristics exhibit a stronger association with repayment rates than with cohort default rates. This influenced our decision to use repayment rates as our dependent variables in our regressions.

2.3 Regional Factors

Webber and Rogers (2014) found that recovery from economic recessions varies across states, which can lead to differences in funding for public institutions and resources for student services programs. This can also affect local employment opportunities. Both factors may suggest that loan repayments would be dependent on the state's ability to recover from an economic recession. Ishitani and McKitrick (2016) found that institutions located in rural areas and areas with higher unemployment rates tend to have higher default rates: they found that a 1% increase in the state unemployment rate led to an increase in the institutional default rate by 0.36%.

3. Theoretical Framework

Our framework functions in two parts. First, we examine the relationship between schools recruiting students versus the value that the institution is providing through the degree, as well as employers' selection of workers. Next, we seek to identify how institution types influence the incentive compatibility of repaying loans.

3.1 Sorting Model

In order to examine the relationship between schools self-selecting students versus the value that an institution provides, we use the "sorting model" (Weiss, 1995). The sorting model combines both the signaling and screening models. In the labor market, in the signaling model the informed (students) move first whereas in the screening models the uninformed (firms) move first. Through the sorting model, both signaling and screening are taking place in order to sort workers according to their unobserved abilities. In the context of our paper, signaling would refer to the students signaling their productive capabilities by choosing to attend college and the screening refers to employers deciding the productive value of a certain degree when looking to hire a new graduate.

According to Weiss (2005), "in sorting models, schooling is correlated with differences among workers that were present before the schooling choices were made; firms make inferences about these productivity differences from schooling choices, and students respond to this inference process by going to school longer". Individuals can pursue more education as a way of signaling their ability. Stiglitz (1975) highlights that "an indirect effect of sorting may be to improve the match between workers and jobs".

This theory provides the background for understanding that a student's choice to attend a for-profit school may be driven in part by their ability as well as the aggressive recruiting strategies from the for-profit institutions (screening). However, the value of their degree may not be seen favorably by employers (signaling). An alternative choice to attend a similarly competitive public or private nonprofit school would "sort" the student as having a higher ability as well as a higher-valued degree. In this way, we add to the "sorting model" by considering not just the choice of pursuing further education, but also once the choice to pursue more education is made, there is the choice of what type of institution to attend.

The mission of public institutions is to provide higher education opportunities at relatively low cost. As such, these institutions may be geared towards lower-income students. However, the for-profit structure of private, for-profit schools leads them to pursue enrollment in a rather aggressive and predatory manner. According to a senate hearing about for-profit schools, "in 2010, the for-profit colleges examined employed 35,202 recruiters compared with 3,512 career services staff and 12,452 support services staff, more than two and a half recruiters for each support services employee" (US

Senate, 2012). These schools put more resources into recruiting than into teaching and academic counseling. For-profit schools often offer the same types of degrees and certificates as public schools, acting as a substitute. Recruiters may convince students that the private for-profit school is a beneficial investment that will result in higher net earnings. Therefore, the recruiters “select” lower-income students who have already decided to “signal” to employers that they invested in education. Students from these backgrounds may have fewer resources to search and check that the recruiters are providing them with correct information. Eventually, as students attend private for-profit schools, the quality of education is sub-par and there are few academic supports to keep students engaged. Students end up withdrawing and maintaining lower earnings than if they had finished, resulting in likelihood to default on loans (Weiss, 1995).

Even if students finish their studies at a for-profit school, employers may “screen” based on the type of institution attended. Employers may believe that those who attend a private for-profit university are not receiving a quality education or that the students are not as productive. In this way, even if a student has high ability, the signal concerning their ability may be “distorted” due to the type of institution they attended. This would again lead to lower earnings and ability to repay loans. Moreover, by using the sorting model, we can see how the choice of what type of school to attend may give the wrong signal about ability, essentially sorting students by low-income status and resulting in lower earnings and repayment than if they had gone to public colleges.

3.2 Incentive Compatibility

An agreement can be called incentive compatible when all parties achieve the best outcome for them by simply acting according to the terms of the agreement. In any loan market, the rational lender should foresee temptations a borrower may have to default and determine the conditions under which default will take place. The lender should not lend more than the borrower would have the incentive to repay.

Based on Lochner and Monge-Naranjo (2016), we assume that once a borrower leaves school, he or she can always opt to default on a repayment D(z). For simplicity, in order to quantify the cost of borrowing, we assume that a defaulting borrower loses a fraction $\kappa \in (0, 1)$ of the individual's labor earnings, so the individual's post-school consumption is cD1 (z) = $(1 - \kappa)$ zaf (h), where zaf (h) represents labor earnings. These losses could reflect punishments imposed by lenders themselves (e.g. wage garnishments) or by others (e.g. landlords refusing to rent or employers refusing to hire). Alternatively, the borrower could repay D (z) yielding post- school consumption cR1 (z) = zaf (h) − D (z). The borrower's decision is straightforward: repay if the cost of defaulting exceeds the cost of repaying: D(z) $\leq \kappa$zaf (h). We will examine the relevance of incentive compatibility with regard to institution types.

We argue that due to the issues discussed above, the lowest repayment rates are likely to be at private for-profit institutions relative compared to other types. This is because κzaf (h) for the for-profit institutions is likely to be lower than at other schools, such that students end up defaulting. This suggests that many variables that affect repayment rates may be stronger (larger value/size) for the private for-profit schools.

4. Data

4.1 The Dataset

The College Scorecard Dataset from the Department of Education provides cohort data for the years 1996 through 2017. For this project, we examine cohort datasets between 2010 and 2016, paying attention to the differences in repayment rates over time and the effects of different variables in predicting the repayment rates. This data is collected at an institutional level with colleges self-reporting their statistics on measures, such as their highest degree awarded, percentage of students from a particular race, etc. We split our data up by three institution types: public, private for-profit, and private non-profit.

Repayment rates refer to whether a borrower makes a payment towards their loan during the period of time during which that borrower is required to make payments on a loan. A borrower can lengthen the repayment period by making smaller payments or shorten it by making additional payments. Additionally, a borrower can postpone a payment without defaulting and will still be counted as someone who is repaying their loan. Borrowers whose repayment periods have not yet begun (one reason for this could be that they are continuing their higher education) are not accounted for in this measure. A borrower would drop out of the repayment measure if they were to default on their loan.

We downloaded all the institutional data for all the available years: 1996 to 2017. We went through each file and checked if the dependent variable had observations in each year. If there were more than 50% missing values, we then discarded observations from that year. We found that repayment rates were not recorded up until 2009, although they had 2-year cohort default rates, which could have been another possibility to examine instead of repayment rates. However, the problem with the cohort default rate variable was only available as a 2-year cohort default rate until 2012, and only as a 3-year cohort default rate after that. This discrepancy would have made it hard to observe patterns over several years and when looking at total missing values, there were significantly more missing values with the 2- or 3-year cohort default rate than with the corresponding 3-year repayment rate. This is how we chose to focus on the 3-year repayment rate. We also removed any observations that did not have a unique ID for the institution that it represented. This is important as when we are doing the panel regression, and by doing so, iterating over the institutions and the years.

4.2 Variable Construction

We created a share of STEM degrees awarded by institutions each year. We took the sum of the majors listed in the original dataset (figure 2) to create this variable. We predict that schools with a higher percentage of students graduating with a STEM degree will have higher repayment rates.

STEM
Natural Resources and Conservation
Computer and Information Sciences and Support Services
Engineering
Engineering Technologies and Engineering-Related Fields
Biological and Biomedical Sciences
Mathematics and Statistics
Military Technologies and Applied Sciences
Physical Sciences
Science Technologies/Technicians
Psychology

Figure 2: STEM majors

We also wanted to incorporate information about the city and state in which an institution is located, such as income, unemployment, and population. We interpreted the city population estimates as a proxy for urbanization. We suspect that this variable is related to repayment rates since we hypothesize that more urban cities would tend to have greater employment opportunities. However, the relationship would only be significant if students are likely to seek employment in the same area as their school. This is not always the case. It might be more relevant to assume this when considering public and private for-profit schools as these schools have historically filled their student body with local students. Through the United States Census Bureau (2018), we got the median population estimates for every city in the

United States from 2010-2017[2]. For each observation in our dataset, we had to match the observation's city and state with the city and state in the United States Census Bureau's dataset in order to obtain the population estimates. We also looked at the corresponding median household incomes in the United States from 2011 to 2016 (Census Bureau, 2020) for the cities that were nearest to each university's location[3]. This dataset provided the median income of each county, so we took the median value of all the county incomes in a city in order to get an estimate of the median income for a city.

For many of the observations in our dataset, there were missing values in several columns. Although we removed observations that were missing values for the dependent variable (3-year repayment rate), it would not be possible to do the same for the explanatory variables, since many of these variables were missing at least 20% of their values. Additionally, the observations that were missing values for a certain column often did not overlap with the observations that were missing values for a different column. Because of the lack of overlap, simply removing observations that were missing values would result in a much smaller dataset, and the parameters that we would estimate via a regression model would not be as accurate since the model was not exposed to enough variation in repayment rates. Instead of simply replacing missing values of a variable with the mean of that variable, which would significantly increase the bias of our model we implemented a data imputation method called Multiple Imputation by Chained Equations (MICE)[4]. The MICE algorithm works by running multiple regression models and pooling the results. Each missing value, which the models are trying to predict, are modeled conditionally on the observed (non-missing) values. We implemented this on the data or each year and transformed the dataset using the MICE package. We then got our transformed

[2] City And Town Population Totals. *United States Census Bureau*. *2010-2018*.

[3] Historical Income Tables: Households. *United States Census Bureau*. 2020.

[4] https://www.rdocumentation.org/packages/mice/versions/3.8.0/topics/mice.

dataset for each of the years and kept track of the year for each observation. Finally, we merged the datasets to create the panel dataset.

We wanted to check for correlations between the continuous variables in our dataset. We created a correlation matrix, which is visualized in figure 3. Utilizing this matrix, we were able to determine if our predictors were redundant and describing the same information. We observed that the percentage of

Figure 3: Correlation matrix for continuous variables

first-generation students was strongly negatively correlated to the median family income. The correlation coefficient was -0.82. Because of the high correlation, we decided to remove the percentage of first-generation students from our dataset before running the regression, since the regression model assumes the absence of multi-collinearity. Median family income was also correlated with the percentage of Pell-grant students and the cost of attendance, but we decided that the correlations (-0.64 and 0.62 respectively) were reasonable.

We then calculated the correlations for the dummy variables using Thiel's U index (the uncertainty coefficient)[5]. We took all the pair permutations of the categorical variables. We then applied the Thiel's U index on each of the pairs and created a correlation matrix for these variables. Through this correlation matrix, we were able to observe that there were no strong correlations between the dummy variables. This is an asymmetric method, and we accounted for this when we took permutations instead of combinations.

4.3 Summary Statistics

According to the summary statistics, which can be found in the Appendix (images C-E), the mean repayment rate is lower for private for-profit schools (35%). The standard deviation of the repayment rates is also the smallest for the private for-profit schools (14.2%). On the other hand, private nonprofit schools have the greatest mean repayment rate (61.1%) but also the highest standard deviation (19.3%).

The mean cost of attendance is lowest for public schools ($16,043) and highest for private nonprofit schools ($34,666). Withdrawal rates are lowest for public schools and private for-profit schools: 24.8% and 25.1% respectively. The mean withdrawal rate is 19.2% for private nonprofit schools. There is also a significant difference between the percentage of STEM degrees awarded by institution type: private for-profit schools award 7%, public schools award 14% and private non-profit schools award 16%. There are also almost double the mean number of black students at private for-profit institutions (24%) compared to private nonprofit and public schools. The mean number of

[5] https://www.statisticshowto.com/uncertainty-coefficient/

Hispanic students is also higher for private for-profit schools (17.5%). The private for-profit schools also appear to be situated in the most populated cities.

For the categorical variables, we can observe that the mode of the highest degree awarded for private for-profit schools is a certificate degree, whereas it is an associate degree for public schools and a graduate degree for private nonprofit schools. We can also observe that the most common region for all types of schools is in the Southeast.

4.4 Limitations

Limitations of our dataset include the fact that we only have data on institutions and not individuals. We can see the average demographics of the students at a school, but we do not have individual information on demographics, loan size and repayment. This means that we cannot draw conclusions on an individual level and we cannot see variations in the student-body demographics.

5. Empirical Specification

5.1 Regression Equation

$$y_{it} = \beta_1 + \beta_2 x_{it2} + \dots \ + \beta_k x_{itk} + \varepsilon_{it}$$

where y_{it} is the average 3-year repayment rate for institution i in year t (2010-2016).

The explanatory variables include

University Characteristics:

Institution Type, Cost of attendance, Withdrawal rate, Admission rate, Highest degree awarded

Student Demographic Characteristics:

Percentage of women, Percentage of part-time students, Percentage of students receiving a Pell grant, Percentage of STEM degrees awarded, Median family income, Percentage of Black students, Percentage of Hispanic students, Percentage of Asian students, Percentage of AIAN students, Percentage of NHPI students,

Regional Characteristics:

Median city income estimates, State unemployment rate, City population estimates

5.2 Dependent Variable

We are using three-year loan repayment rates for our dependent variable which the data dictionary for the Scorecard dataset defines as, "the share of borrowers paying off at least one dollar of their loan principal within three years of entering repayment"[6].

The repayment rate improves upon the cohort default rate as a metric for ability to pay back student loans by including students who are struggling to make payments. Cohort default rates significantly understate the proportion of students who have been unable to reduce their student loan amounts relative to the repayment rate, as the cohort default rate does not account for students who have not defaulted but also not started their repayment plan.

5.3 Independent Variables

There are three main categories for our independent variables; those include student demographic, institutional, and regional characteristics. Albeit all of the data available in our dataset is at the institutional level, we have labeled specific data groups "demographic", because they are more descriptive of the student profile attending these institutions. The practice of using institutional level data to gain insights on individual student borrowers has been employed by other literature (Kelchen and Li, 2017; Lochner and Monge-Niranjo, 2015) in the academic field of student loan markets.

5.3.1 Student Demographics

[6] https://collegescorecard.ed.gov/data/

In our student category, we control for the percentage of students who are first-generation, percentage of female undergraduates at an institution, percent of undergraduates by race. Students from economically disadvantaged backgrounds, which include first-generation, and Black and Hispanic students, tend to be associated with lower repayment rates (Kelchen and Li, 2017). Regarding gender, we expect the female indicator to have a positive relationship with repayment rate. On average, women take about two years longer than men do to repay student debt (AAUW, 2020).

With regards to major choice, we expect percentage of STEM degrees awarded to have a strong positive association with our dependent variable, as STEM majors owe a much smaller portion of their debt relative to other majors after 10 years (Lochner and Monge-Naranjo, 2016).

We also chose to control for family income level, looking at median family income by institution as a continuous variable. We expect that higher family income positively correlates to lower repayment rates (Kelchen and Li, 2017).

5.3.2 Institutional Traits

There is evidence to suggest for-profit colleges yield the lowest repayment rates amongst all institution types (Hillman, 2015; Lochner and Monge-Naranjo, 2014). We also control for the highest degree awarded by an institution. Graduates coming out of four-year universities are 18% more likely to begin paying off their loan principal relative to graduates of two-year programs at least one year after completion (Kreighbaum, 2020). We predict schools that offer bachelor’s and graduate degrees for “Highest Degree Awarded” will have the highest loan repayment rate.

We also control the cost of attendance across institutions. Institutions that have higher costs of attendance average lower non-repayment rates (Kelchen and Li, 2017). We expect as the tuition of a school increases, the quality of education likely increases as well, thereby having a positive relationship with repayment rates. According to Ishitani and McKitrick (2016), more selective schools have lower default rates. We use admission rates as a proxy for selectivity, and expect admission rate to have a negative relationship with repayment rate.

5.3.3 Regional Characteristics

Given that this paper uses data from 2010 through 2016, we included regional variables to control for economic differences over time. Our regional variables include state unemployment, median household income of the nearest city to the university, and population estimates of the nearest city. Ishitani and McKitrick (2016) found state unemployment rate and urban location to have a positive relationship with repayment rate. Based on their findings, we predict that schools in states with lower unemployment rates, in cities with higher populations and in cities with higher incomes would lead to higher repayment rates.

6. Results

Since we had time-series data, we utilized a panel regression to better understand which variables contributed to the variation in repayment rates while accounting for time. We ran three separate panel data regressions for each of the school groups: Pooled OLS, Fixed Effects (FE) and Random Effects (RE). In order to compare the FE and RE models, we conducted a Hausman test. This test detects endogenous regressors in a regression model. Endogenous variables have values that are determined by other variables in the system. Having endogenous regressors in a model will cause ordinary least squares estimators to fail, as one of the assumptions of OLS is that there is no correlation between a predictor variable and the error term. The null hypothesis of the test supports the claim that a RE model provides a better estimation of the parameters and is the preferred model, whereas the alternate hypothesis is that the preferred model is the Fixed Effects model. From the Hausman test, we derived the following p-values when comparing the FE and RE models for public, private nonprofit and private for-profit schools respectively: 0.99928, 0.99999 and 1.0. None of these p-values were significant, so we do not need to reject any of the null hypotheses. This shows us that we should use the RE models since there are no systematic differences in the coefficients between the FE and the RE models. Because of this, all

	Public	Private_nonprofit	Private_for-profit
Constant	0.59 (0.0)	0.58 (0.0)	0.49 (0.0)
Admission_rate	-0.03 (0.0)	-0.02 (0.0)	-0.12 (0.0)
Cost_of_attendance	0.0 (0.0)	0.0 (0.0)	0.0 (0.02)
Non-degree-granting	-0.02 (0.0)	-0.06 (0.0)	0.0 (0.4)
Certificate_degree	-0.0 (0.66)	-0.01 (0.08)	0.02 (0.0)
Associate_degree	-0.02 (0.0)	-0.02 (0.0)	0.01 (0.01)
Graduate_degree	0.04 (0.0)	0.01 (0.06)	0.01 (0.08)
Withdrawn_by_4yrs	-0.13 (0.0)	-0.12 (0.0)	-0.26 (0.0)
Median_family_income	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Demographics.race_ethnicity.black	-0.21 (0.0)	-0.31 (0.0)	-0.14 (0.0)
Demographics.race_ethnicity.hispanic	-0.02 (0.06)	-0.17 (0.0)	-0.04 (0.0)
Demographics.race_ethnicity.asian	0.14 (0.0)	0.08 (0.0)	0.11 (0.0)
Demographics.race_ethnicity.aian	-0.08 (0.03)	-0.3 (0.0)	-0.09 (0.01)
Demographics.race_ethnicity.nhpi	-0.15 (0.03)	-0.17 (0.0)	-0.02 (0.54)
Share_women	-0.0 (0.93)	0.03 (0.0)	-0.0 (0.45)
Percentage_STEM_degrees	0.04 (0.0)	0.03 (0.0)	-0.01 (0.2)
Share_pell-grant	-0.04 (0.0)	-0.1 (0.0)	-0.03 (0.0)
Share_part-time	-0.07 (0.0)	0.0 (0.77)	0.04 (0.0)
City_population_estimates	-0.0 (0.9)	0.0 (0.02)	0.0 (0.01)
Median_city_income_estimates	0.0 (0.01)	0.0 (0.15)	0.0 (0.0)
State_unemployment_rates	-0.0 (0.0)	0.0 (0.64)	-0.0 (0.0)
Year.2011	-0.04 (0.0)	-0.04 (0.0)	-0.06 (0.0)
year.2012	-0.07 (0.0)	-0.06 (0.0)	-0.11 (0.0)
year.2013	-0.1 (0.0)	-0.08 (0.0)	-0.12 (0.0)
year.2014	-0.11 (0.0)	-0.08 (0.0)	-0.12 (0.0)
year.2015	-0.12 (0.0)	-0.09 (0.0)	-0.12 (0.0)
year.2016	-0.11 (0.0)	-0.1 (0.0)	-0.11 (0.0)

Figure 4: Random effects panel regression's coefficients and p-values (in parentheses). P-values that are not significant at a 95% confidence level are colored in red. R-squared values for public, private nonprofit and private for-profit schools respectively: 59.65%, 52.34%, and 52.95%.

of the conclusions that we present in this section are taken from the RE models. The estimator coefficients and p-values across school groups from this model are pictured in figure 4.

Using the RE models, we got the following R-squared values for public, private nonprofit and private for-profit schools respectively: 59.65%, 52.34%, and 52.95%. This means that model was the best at describing the variation in repayment rates for public schools and worst for private nonprofit schools. Our first observation from the RE models is that the constant value was the smallest for private for-profit schools. We mean-centered all of the continuous variables, so the constants provide more information. This observation shows us that if we assume averages for the explanatory variables, private for-profit schools have lower repayment rates.

6.1 Repayment Rates Across School Groups

We mean-centered all of the continuous variables, so the constants in the regression highlight that if we assume averages for the explanatory variables, private for-profit schools have lower repayment rates. Visualizing the repayment rates for each school group in a histogram (figure 5) allows us to observe that the repayment rate distributions for each school group are varied. With respect to private nonprofit schools, they have the highest mean repayment rate and their repayment rates vary widely. The distribution is also negatively skewed. Below a repayment rate of 40%, there seems to be a significant increase in frequency, suggesting that the distribution could be bimodal. On the other hand, the distribution of repayment rates for public schools more likely resembles a symmetrical curve. This result suggests that educational institutions belonging to this group fall below or above the mean in equal proportions. The distribution of repayment rates for private for-profit schools is centered at the smallest repayment rate among the distributions. The curve is thinner than the other two curves and has a large frequency at its mean. It appears that the student bodies belonging to private for-profit schools have a harder time on average paying back their loans.

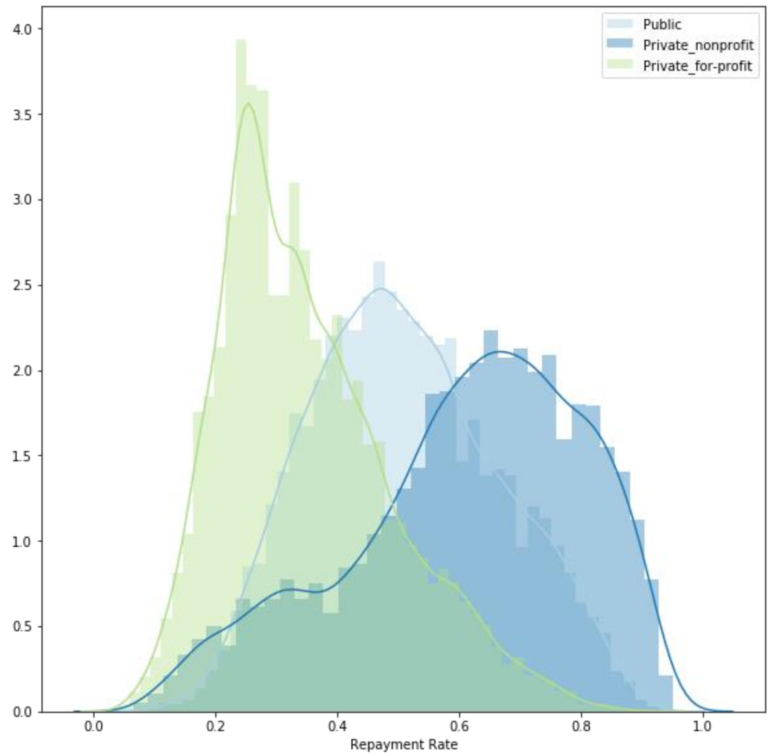

Figure 5: Histograms of repayment rates across school groups

6.2 Student Demographics

According to the RE models, the coefficient for median family income is positive and significant for each school group at a 95% confidence level. The coefficient is twice as large for private for-profit schools as it is for private non-profit schools. This suggests that students from private for-profit schools are more dependent on family income when it comes to paying back their student debt. This may suggest that the value added from their degree is less valuably received by employers in the job market. Figure 6 visualizes the relationship between median family income and repayment rates by school type. The results from the RE models and the figure suggest that schools with student bodies that come from wealthier backgrounds have historically been better at repaying their loans, and this is especially true for students from private for-profit schools.

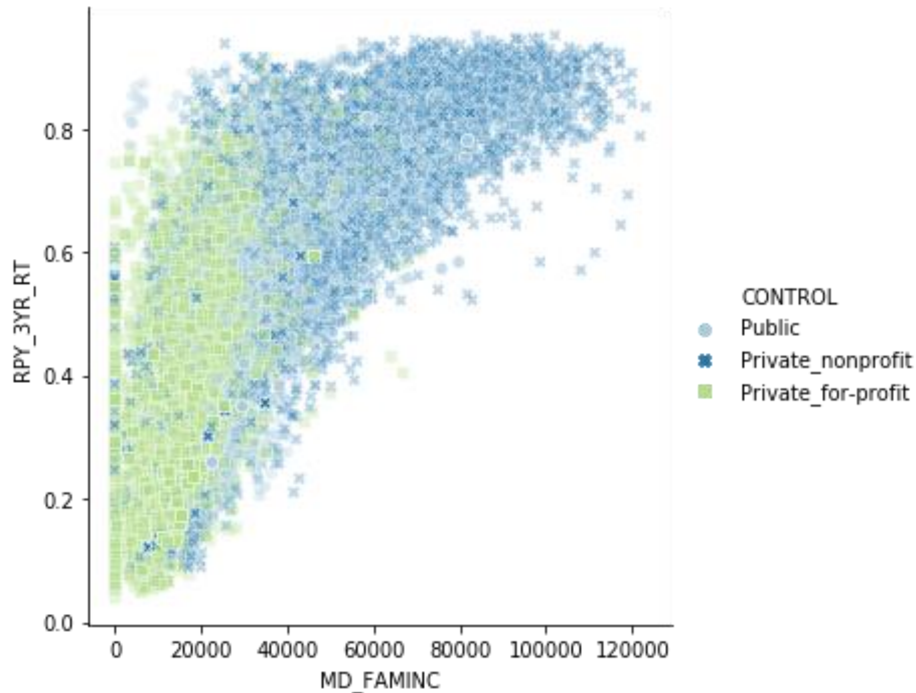

Figure 6: Scatter plot of median family income
and repayment rate

Independently of the median family income for the average student, race also proves to be important in describing the variation in repayment rates for each school group. The proportion of students who were black is negatively associated with repayment rates and is statistically significant for all school types. In other words, schools with larger black student population ratios are found to have lower repayment rates. It was the most negative for private nonprofit schools (-0.31) and the least negative for private for-profit schools (-0.14). Although not as significant, we find similar results for the variable representing the proportion of students who were Hispanic. This variable was negative for each school group, but it was not statistically significant for public schools. It was the most negative for private nonprofit schools (-0.17) and the least negative for private for-profit schools (-0.04). These findings may reflect lower earnings after graduation, although the data used here do not give information about future earnings.

6.3 Institutional Characteristics

Changing the highest degree that an institution can award their students does appear to have a significant effect on the expected repayment rate for that institution according to the RE models. However, not all of the variable levels were significant for each school group. For instance, when using bachelor's degree as the reference value, changing to a certificate degree did not have a clear effect on repayment rates for public schools. The coefficients for the different variable levels for public and private nonprofit schools had equivalent signs. On the other hand, the coefficients for the different variable levels for private for-profit were always positive.

To analyze these relationships further, we created three separate boxplots for each of the school groups with repayment rate as the dependent variable (figure 7). Public schools that offered graduate degrees appeared to have the highest repayment rates on average with less variation than some of the other highest degrees. The average repayment rates seemed to increase as the highest degree increased from non-degree awarding to graduate degree with the exception of associate degrees which fared worse than certificate degrees. The mean repayment rate for public schools with an associate degree as the highest degree awarded was very similar to the mean repayment rates for institutions that do not offer degrees. Perhaps, those schools have similar student populations.

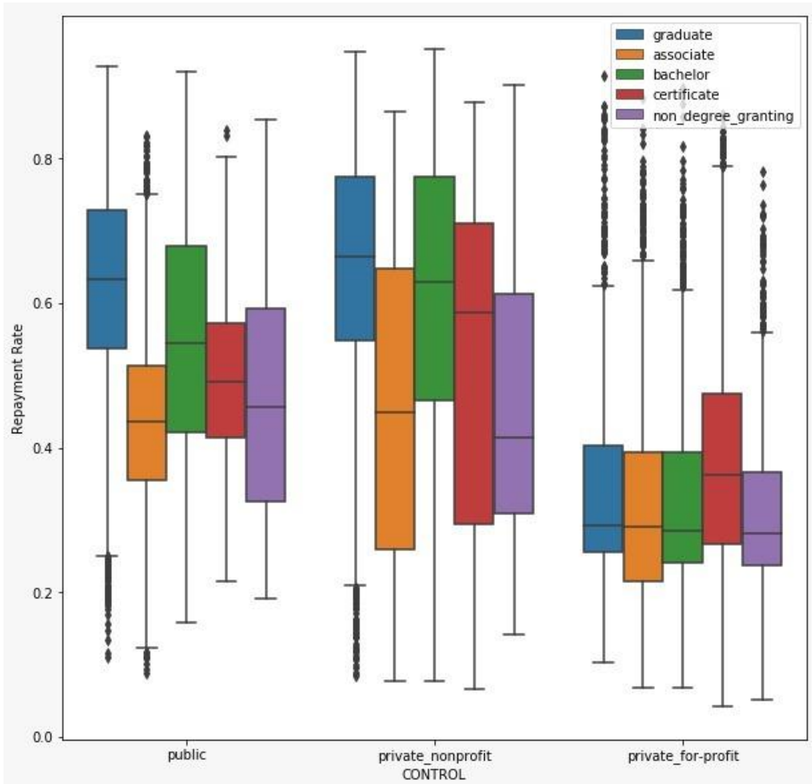

Figure 7: Box plots of highest degree awarded and
repayment rate across school groups

When looking at private nonprofit schools, we did not have diverging conclusions. However, one thing to note was the variation in repayment rates for schools with the certificate degree as the highest degree awarded was higher than the variation of all other degrees. Finally, when looking at private for-profit schools, we had much more diverging conclusions than we had for public and private for-profit schools. When changing the highest degree awarded (our reference layer was bachelor's degree for all institutions), the repayment rates did not seem to change. The only change occurred when switching to the certificate degree, which had an average higher repayment rate than the other degrees. Interestingly, private for-profit schools that offered bachelor's or graduate degrees did not have higher repayment rates than those that did not offer those programs. This is the first time that we are observing this phenomenon since this was not consistent with public and private nonprofit schools.

Next, we wanted to observe if the highest degree awarded by an institution would help us understand why the distribution for private nonprofit schools appeared to be bimodal and why students at for-profit schools might be at a disadvantage when it comes to paying back their student loans (figure 5). We suspected that the cause was the same for both of these observations, particularly that the highest degree awarded for these institutions might play a large role in determining repayment rates. Looking at the frequency plot of the highest degree awarded for each school group in figure 8, it is clear that private for-profit schools consist primarily of schools offering up to associate and certificate degrees, whereas private nonprofit schools consist primarily of schools offering up to bachelor's and graduate degrees. We suspected that in nonprofit school groups, the presence of institutions that offer only up to certificate and associate degree programs might be bringing the average repayment rate lower and causing that bimodal appearance. Similarly, the lack of bachelor's and graduate degree offerings for private for-profit schools could be the reason why this school group had the smallest average repayment rate among the three groups.

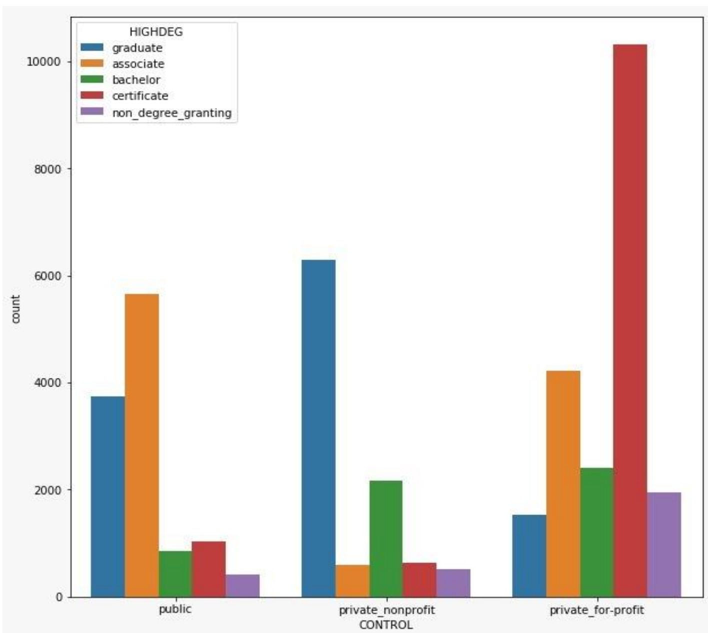

Figure 8: Bar charts of highest degree awarded across school groups

Building off of these hypotheses, we filtered our observations for the highest degree awarded to be associate degree, certificate degree, and non-degree awarding. As we had done previously, we split the institutions into their school groups and visualized the repayment rate distributions, as pictured in figure 9. Doing so, further exaggerated the bimodal appearance of the distribution of repayment rates for private nonprofit schools, which explains why we noticed the bimodal appearance of the original distribution in the first place. Additionally, as expected, the mean of the repayment rates for private for-profit schools does not increase with respect to the mean of the original distribution, but it does not decrease much either which is surprising but could be validated by what we observed previously via the boxplots. Something important to note is that the distribution for public schools seems to more closely resemble the distribution for private for-profit schools. Although public schools on average have higher

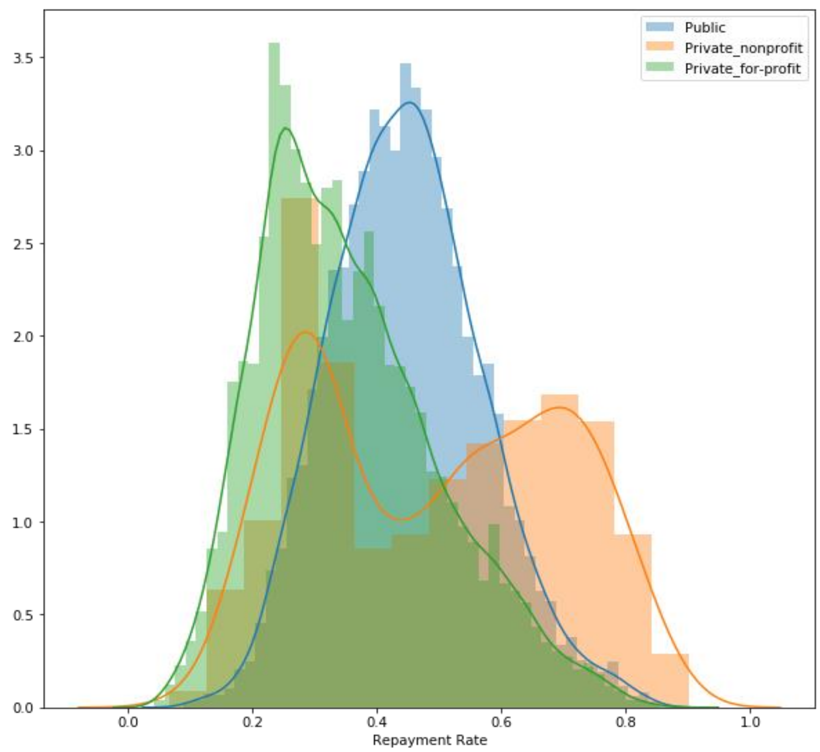

Figure 9: Histograms of repayment rates across school groups where highest degree awarded is associate degree or below

repayment rates, the repayment rates do not appear that different after excluding bachelor's and graduate degree offerings, which seem to place public schools at an advantage when it comes to students paying back their loans.

We proceeded to filter the data again, but this time for institutions that offer bachelor's and graduate programs (figure 10). We found that the distribution of repayment rates for public schools more closely resembles the distribution of repayment rates for private nonprofit schools in this case. In fact, they are nearly overlapping. This could suggest that when removing public schools that do not offer bachelor's and graduate degrees, public school students have an equivalent likelihood of paying back their loans as private nonprofit school students. Additionally, filtering in this way removes the bimodal appearance of the distribution for private nonprofit schools, which we hypothesized would happen.

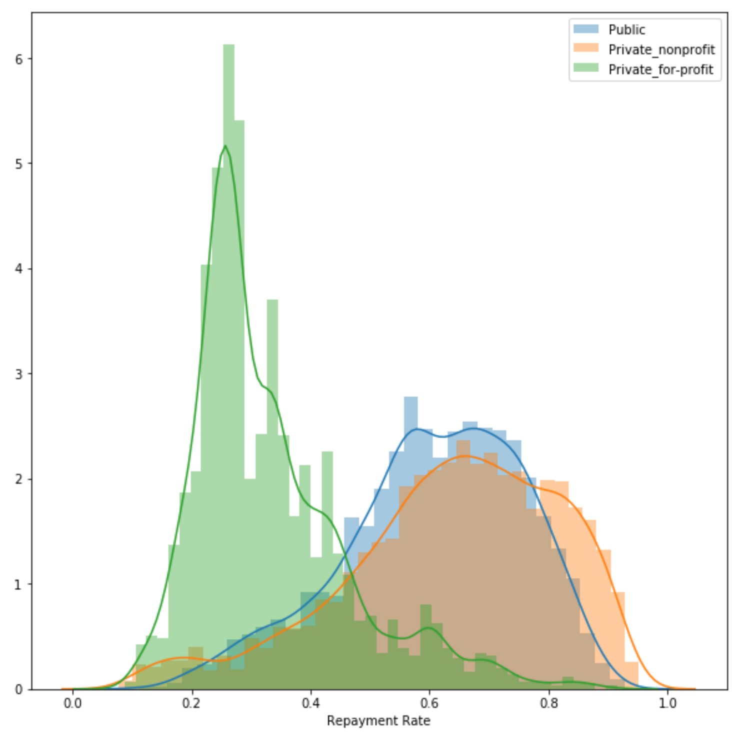

Figure 10: Histograms of repayment rates across school groups where highest degree awarded is bachelor's or graduate

An additional insight that can be taken away from this visualization is that the distribution of repayment rates for private for-profit schools does not change that significantly. The only observable change is that the frequencies of lower repayment rates is higher with respect to the frequencies of higher repayment rates. This is a surprising result, since we suspected that private for-profit schools have worse repayment rates on average than the other school groups because they did not offer a similar proportion of bachelor's and graduate degrees. Moreover, it is reasonable to conclude then that the highest degree awarded is a much more significant variable when it comes to describing the variation in repayment rates for public schools than the repayment rates for the other two school groups.

There were two other interesting observations that we wanted to analyze further. Firstly, returning to figure 9, we wanted to understand the bimodal appearance of the distribution of repayment rates for private nonprofit schools. We decided to split the distribution into the two separate populations that the distribution suggested. Furthermore, any observations with a repayment rate below 45% we put in one dataset and any observations with a repayment rate above or equal to 45% we put in another dataset. We then ran t-tests to compare the continuous variables for each of the datasets in order to see how they differed. Although this analysis would not tell us which variables were more significant, it would give us a more comprehensive understanding of how the sub-populations of private nonprofit schools differed. At a 95% confidence level, we observed that the average Black and Hispanic populations were higher in the lower repayment group: 29.7% versus 9.94% and 24.9% versus 9.64%. The average admission rate was higher for the lower repayment group: 75.5% versus 65.6%. The average median family income was also higher for the higher repayment group: $35,518 versus $17,464. These factors suggest that the private nonprofit schools with lower repayment rates may be screening for students with more marginalized identities/ backgrounds and lower abilities. Interestingly, the average cost of attendance is higher for the higher repayment group: $27,108 versus $21,564, suggesting that the cost of attendance is important in determining the value of a private nonprofit associate degree, certificate degree or non-degree granting education. This highlights the interplay between signaling and screening.

The other observation that we wanted to further explore was the distinct difference in repayment rates between private for-profit and private nonprofit schools. Similarly, we ran t-tests to compare the continuous variables for each of the datasets. At a 95% confidence level, we observed that the average Black and Hispanic populations were higher in private for-profit schools: 24.2% versus 14.4% and 17.5% versus 11.8%. The average admission rate was higher for the private for-profit schools: 75.6% versus 65.1%. The average median family income was higher for the private nonprofit schools: $44,938 versus $16,766. The cost of attendance is also higher for the private nonprofit schools: $34,666 versus $21,705. The same variables that were significant when analyzing the sub-groups of private nonprofit schools continue to be significant for understanding the differences between private for-profit and private nonprofit schools.

Returning to the RE models, there is an interesting relationship between an institution's admissions rate and its repayment rate. The coefficient for the admissions rate variable was negative and significant across all school groups. This implies that more selective (smaller admissions rate) institutions had higher repayment rates. Perhaps, more selective schools offer more value to their students, and this value-add helps students pay back their loans once they receive their degrees. An alternate explanation is that a selective school does not actually increase a student's ability to pay off their loan, but the students entering selective schools are more likely to not default than the students entering non-selective schools.

As expected, the coefficient for the withdrawal rate variable was negative and significant across all school groups as the value that they would receive upon completion of the degree, namely the diploma and skills, would have helped them secure job prospects that would aid them in paying off their student debt. Nonetheless, the magnitude of the coefficients did differ significantly, and the coefficient of the variable for private for-profit schools was more than double the coefficients for both public and private nonprofit schools (26%, 13% and 12% respectively). This could suggest that private for-profit schools have lower ability students that cannot finish their degrees and because of this, they cannot pay their student debt. Similarly, the coefficient for the cost of attendance variable was positive and significant for all groups. The coefficient of the variable for private for-profit schools was smaller than the coefficients for both public and private nonprofit schools by a nearly a factor of five ($4.16 * 10^{-7}$, $2.17 * 10^{-6}$and $1.87 * 10^{-6}$ respectively).

6.3 Regional Attributes

None of the regional attributes, such as the city population estimates, median city income estimates, and state unemployment rates, were significant for all school groups. The city population estimate variable was positive and significant for all private schools. The median city income estimate variable was positive and significant for public and private for-profit schools. Both of these schools largely attract local students, who probably have similar family incomes to the median city income estimate. Therefore, it makes intuitive sense that the repayment rates for these schools are dependent on the median city income in the same way that they are dependent on the median family income. Similarly, the state unemployment rate variable was negative and significant for public and private for-profit schools.

6.4 Time

Since we are utilizing panel regressions, we can observe how repayment is expected to change over time with respect to an institution. From the results, it appears as the year increases, the coefficients become more negative. This means that as time has increased, repayment rates on average have decreased. However, this relationship is only clear across all student groups from 2011 to 2013. The decrease in repayment rates could be attributed to the global economic recession that occurred in 2008. According to Long (2014), the 2008 recession is distinctive from other recessions as, despite the fact that college costs and student debt levels were at historic highs, college attendance actually increased during the recession. This was especially true for colleges in the states most affected by the recession. Long (2014) also states that due to easy educational financing available through loans, the gains in attendance were concentrated among unconventional students (such as part-time students) and students of color. Students enrolling around the time of the recession would be reflected in the cohort data up to 4 years after their enrollment (depending on their degree type). This decrease in repayment rates could be attributed to a higher number of underprepared and/ or disadvantaged students taking on high levels of debt for schools that might not provide them with a degree that adds enough substantial value. For both public and private for-profit schools, the coefficient values plateau or even slightly decrease in magnitude after 2013. On the other hand, for private nonprofit schools, the coefficient values increase in magnitude. In other words, the repayment rate decreases over time. This could be attributed to the increase in the cost of attendance that we have observed over time. Even though the cost of attendance has increased for all school groups, the rate at which it has increased for private nonprofit schools is greater.

7. Conclusion

Our findings confirm that the average student in private for-profit schools has the worst repayment rate among school groups whereas the average student in private nonprofit schools has the highest repayment rate. For an average student in public schools, the repayment rate oscillates between the two extremes, conforming more closely to the repayment rate of private for-profit schools or private nonprofit schools depending on the highest degree that the institution awarded. If a public school offers a bachelor's or graduate degree as their highest degree, the repayment rate of the average student increases resembling the upper bound; for all other degrees, their repayment rate decreases resembling the lower bound. Regional attributes (state unemployment rate, city population, and city income) were significant in describing the variation in repayment rates for both public and private for-profit schools, but regional attributes were largely not significant for private nonprofit schools. This may be driven by the fact that more of the student body from both public and private for-profit schools lives locally. Therefore, the student body is more likely to pursue local work opportunities after graduation.

Student attributes regarding race were particularly significant in influencing the repayment rates for private nonprofit institutions. Schools with a larger percentage of Black students are seen to have lower repayment rates. It was the most negative for private nonprofit schools (-0.31) and the least negative for private for-profit schools (-0.14). In accordance with the findings from Lochner and Monge-Naranjo (2014), this is a phenomenon not wholly explained by differences in post-school earnings, choice of major, type of institution. This may suggest that there is racial discrimination in the workplace when employers are hiring. Black students may find it harder to have the means of repaying their loans.

Institutional attributes such as acceptance rates and withdrawal rates were significant. More selective (smaller admissions rate) institutions had higher repayment rates. As expected, the coefficient for the withdrawal rate variable was negative and significant across all school groups. However, the coefficient of the variable for private for-profit schools was more than double the coefficients for both public and private nonprofit schools (26%, 13% and 12% respectively). This could suggest that private for-profit schools have lower ability students that cannot finish their degrees and because of this, they are not able to repay their student debt. Students from more selective schools could also signal to employers that they have been screened for having high abilities, on top of signaling that their degree has given them valuable skills.

Overall, our findings will be valuable for students, allowing them to make more informed decisions about the types of schools that they apply to or choose to attend if they are to take out a loan. Although the highest R-squared value of a school group from our RE models was approximately 60%, we were still able to understand a significant portion of the variation in repayment rates. Our analysis has shown that loan repayments rates for for-profit institutions with large percentages of minority students can be particularly low. We also observe that students from private for-profit schools are more dependent on family income when it comes to paying back their student debt. Making an informed decision about higher education can be difficult, especially given that for-profit institutions convince prospective students to attend through predatory tactics such as using sales representatives (Daly, 2020). For students from minority groups and/or disadvantaged backgrounds, this research can be a helpful resource for understanding which schools will help them signal greater skills to employers and ensure that they repay their loans. Additionally, education policymakers who desire to reduce student debt default rates can utilize this research to develop policies that aid particular institutions where there are student populations that are particularly vulnerable.

Furthermore, our findings show that selectivity impacts repayment rate at for-profit schools significantly more than it does private non-profit and public schools in the regression. This implies that there may be subgroups within for-profit schools that could be categorized by selectivity. Given more institutional and individual-level data, one could explore how more successful for-profit schools (those with higher repayment rates) operate differently from less selective for-profit schools with lower repayment rates. This could help us examine what characteristics makes for-profit schools successful at attracting students of a higher ability and signaling the value of their education to potential employers.

Appendix

Appendix A: Pooled OLS's coefficients and p-values (in parentheses). P-values that are not significant at a 95% confidence level are colored in red. R-squared values for public, private nonprofit and private for-profit schools respectively: 74.41%, 82.32%, and 61.40%.

	Public	Private_nonprofit	Private_for-profit
Constant	0.61 (0.0)	0.57 (0.0)	0.53 (0.0)
Admission_rate	-0.12 (0.0)	-0.07 (0.0)	-0.25 (0.0)
Cost_of_attendance	0.0 (0.0)	-0.0 (0.4)	-0.0 (0.0)
Non-degree-granting	-0.04 (0.0)	-0.05 (0.0)	0.0 (0.59)
Certificate_degree	-0.03 (0.0)	0.01 (0.01)	0.04 (0.0)
Associate_degree	-0.03 (0.0)	-0.02 (0.0)	0.0 (0.28)
Graduate_degree	0.01 (0.02)	0.0 (0.13)	0.02 (0.0)
Withdrawn_by_4yrs	-0.22 (0.0)	-0.29 (0.0)	-0.35 (0.0)
Median_family_income	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Demographics.race_ethnicity.black	-0.28 (0.0)	-0.36 (0.0)	-0.18 (0.0)
Demographics.race_ethnicity.hispanic	-0.06 (0.0)	-0.18 (0.0)	-0.05 (0.0)
Demographics.race_ethnicity.asian	0.23 (0.0)	0.23 (0.0)	0.17 (0.0)
Demographics.race_ethnicity.aian	-0.08 (0.0)	-0.32 (0.0)	-0.19 (0.0)
Demographics.race_ethnicity.nhpi	-0.34 (0.0)	-0.52 (0.0)	-0.01 (0.65)
Share_women	0.05 (0.0)	0.04 (0.0)	-0.03 (0.0)
Percentage_STEM_degrees	0.03 (0.0)	0.06 (0.0)	-0.0 (0.97)
Share_pell-grant	-0.08 (0.0)	-0.14 (0.0)	-0.06 (0.0)
Share_part-time	-0.07 (0.0)	0.03 (0.0)	0.06 (0.0)
City_population_estimates	-0.0 (0.42)	0.0 (0.0)	0.0 (0.0)
Median_city_income_estimates	0.0 (0.06)	-0.0 (0.11)	-0.0 (0.0)
State_unemployment_rates	-0.01 (0.0)	-0.0 (0.0)	-0.0 (0.0)
Year.2011	-0.05 (0.0)	-0.04 (0.0)	-0.05 (0.0)
year.2012	-0.08 (0.0)	-0.06 (0.0)	-0.09 (0.0)
year.2013	-0.11 (0.0)	-0.08 (0.0)	-0.1 (0.0)
year.2014	-0.13 (0.0)	-0.09 (0.0)	-0.09 (0.0)
year.2015	-0.15 (0.0)	-0.1 (0.0)	-0.09 (0.0)
year.2016	-0.14 (0.0)	-0.11 (0.0)	-0.09 (0.0)

Appendix B: Fixed Effects Panel Regression's coefficients and p-values (in parentheses). P-values that are not significant at a 95% confidence level are colored in red. R-squared values for public, private nonprofit and private for-profit schools respectively: 57.14%, 43.86%, and 53.40%.

	Public	Private_nonprofit	Private_for-profit
Constant	0.58 (0.0)	0.66 (0.0)	0.47 (0.0)
Admission_rate	-0.03 (0.0)	-0.01 (0.06)	-0.09 (0.0)
Cost_of_attendance	0.0 (0.04)	0.0 (0.0)	0.0 (0.0)
Non-degree-granting	-0.01 (0.21)	-0.03 (0.02)	0.0 (0.46)
Certificate_degree	0.02 (0.28)	0.0 (0.68)	-0.01 (0.0)
Associate_degree	0.01 (0.01)	-0.01 (0.31)	0.01 (0.06)
Graduate_degree	0.01 (0.2)	-0.01 (0.02)	0.0 (0.33)
Withdrawn_by_4yrs	-0.11 (0.0)	-0.07 (0.0)	-0.2 (0.0)
Median_family_income	-0.0 (0.98)	-0.0 (0.63)	0.0 (0.45)
Demographics.race_ethnicity.black	-0.0 (0.81)	-0.07 (0.0)	-0.05 (0.0)
Demographics.race_ethnicity.hispanic	0.04 (0.02)	-0.05 (0.0)	-0.0 (0.96)
Demographics.race_ethnicity.asian	-0.13 (0.0)	-0.06 (0.01)	-0.0 (0.85)
Demographics.race_ethnicity.aian	-0.07 (0.27)	-0.06 (0.44)	-0.02 (0.59)
Demographics.race_ethnicity.nhpi	-0.11 (0.15)	-0.05 (0.35)	-0.07 (0.02)
Share_women	-0.0 (0.68)	0.02 (0.04)	-0.0 (0.63)
Percentage_STEM_degrees	0.02 (0.16)	-0.03 (0.01)	0.0 (0.53)
Share_pell-grant	-0.02 (0.0)	-0.03 (0.0)	-0.01 (0.02)
Share_part-time	-0.02 (0.01)	0.01 (0.4)	0.03 (0.0)
City_population_estimates	0.0 (0.0)	-0.0 (0.0)	0.0 (0.79)
Median_city_income_estimates	-0.0 (0.24)	0.0 (0.86)	-0.0 (0.3)
State_unemployment_rates	-0.0 (0.18)	0.0 (0.01)	-0.0 (0.0)
Year.2011	-0.04 (0.0)	-0.04 (0.0)	-0.06 (0.0)
year.2012	-0.07 (0.0)	-0.06 (0.0)	-0.11 (0.0)
year.2013	-0.09 (0.0)	-0.07 (0.0)	-0.13 (0.0)
year.2014	-0.1 (0.0)	-0.07 (0.0)	-0.13 (0.0)
year.2015	-0.1 (0.0)	-0.08 (0.0)	-0.13 (0.0)
year.2016	-0.09 (0.0)	-0.08 (0.0)	-0.12 (0.0)

Appendix C: Continuous variables for public schools

	Number of observations	Mean	Standard Deviation	Minimum	Maximum
3-year repayment rate	11,682	0.510	0.155	0.088	0.928
Cost of attendance	11,682	16,043	5,605	3,368	62,551
Withdrawal rate	11,682	0.248	0.111	0.013	0.722
Percentage of first-generation students	11,682	0.443	0.105	0.085	0.787
Percentage of women	11,682	0.587	0.117	0.000	1.000
Percentage of part-time students	11,682	0.343	0.227	0.000	1.000
Percentage of STEM majors	11,682	0.141	0.131	0.000	1.259
Percentage of students with Pell grants	11,682	0.406	0.153	0.000	1.000
Admission rate	11,682	0.730	0.129	0.071	1.000
Median family income	11,682	30,098	15,975	3,256	108,367
City population estimates	11,682	342,357	650,372	142	8,475,976
Percentage of Black students	11,682	0.14	0.17	0.00	0.97
Percentage of Hispanic students	11,682	0.13	0.16	0.00	1.00
Percentage of Asian students	11,682	0.04	0.06	0.00	0.51
Percentage of AIAN students	11,682	0.01	0.04	0.00	0.88
Percentage of NHPI students	11,682	0.00	0.02	0.00	0.51
Median city income estimates	11,682	79,133	47,565	9,618	300,000
State unemployment rates	11,682	7.066	2.089	2.600	17.000

Appendix D: Continuous variables for private nonprofit schools

	Number of Observations	Mean	Standard Deviation	Minimum	Maximum
3-year repayment rate	10,199	0.611	0.193	0.067	0.952
Cost of attendance	10,199	34,666	12,608	2,200	83,703
Withdrawal rate	10,199	0.192	0.106	0.001	0.651
Percentage of first-generation students	10,199	0.355	0.129	0.006	0.818
Percentage of women	10,199	0.599	0.165	0.000	1.000
Percentage of part-time students	10,199	0.155	0.195	0.000	1.000
Percentage of STEM majors	10,199	0.163	0.164	0.000	1.554
Percentage of students with Pell grants	10,199	0.415	0.196	0.000	1.000
Admission rate	10,199	0.651	0.180	0.027	1.000
Median family income	10,199	44,939	22,955	0	123,136
City population estimates	10,199	797,333	1,198,967	61	8,475,976
Percentage of Black students	10,199	0.145	0.191	0.000	1.000
Percentage of Hispanic students	10,199	0.118	0.181	0.000	1.000
Percentage of Asian students	10,199	0.036	0.056	0.000	0.889
Percentage of AIAN students	10,199	0.007	0.031	0.000	0.897
Percentage of NHPI students	10,199	0.003	0.013	0.000	0.456
Median city income estimates	10,199	85,782	46,638	9,618	300,000
State unemployment rates	10,199	7.228	2.322	2.600	17.000

Appendix E: Continuous variables for private for-profit schools

	Number of Observations	Mean	Standard Deviation	Minimum	Maximum
3-year repayment rate	20,419	0.350	0.142	0.042	0.915
Cost of attendance	20,419	21,705	5,723	6,157	91,532
Withdrawal rate	20,419	0.251	0.111	0.005	0.735
Percentage of first-generation students	20,419	0.531	0.078	0.144	0.970
Percentage of women	20,419	0.724	0.254	0.000	1.000
Percentage of part-time students	20,419	0.194	0.230	0.000	1.000
Percentage of STEM majors	20,419	0.075	0.183	0.000	1.051
Percentage of students with Pell grants	20,419	0.648	0.172	0.000	1.000
Admission rate	20,419	0.757	0.104	0.087	1.000
Median family income	20,419	16,766	6,880	0.000	70,452
City population estimates	20,419	602,091	911,220	1,314	8,475,976
Percentage of Black students	20,419	0.242	0.222	0.000	1.000
Percentage of Hispanic students	20,419	0.175	0.209	0.000	1.000
Percentage of Asian students	20,419	0.025	0.051	0.000	1.000
Percentage of AIAN students	20,419	0.008	0.020	0.000	0.694
Percentage of NHPI students	20,419	0.004	0.020	0.000	0.692
Median city income estimates	20,419	88,323	43,992	9,618	300,000
State unemployment rates	20,419	7.295	2.159	2.600	17.000

References

141932 Executive Summary Booklet. (2020). *United States Senate*. http://help.senate.gov/imo/media/for_profit_report/ExecutiveSummary.pdf.

Brown, M., Haughwot, A., & Lee, D. (2014). *Measuring Student Debt and Its Performance*. Federal Reserve Bank of New York Staff Reports. Retrieved from https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr668.pdf

Buuren, S. van. (n.d.). Mice Function | R Documentation. Retrieved from https://www.rdocumentation.org/packages/mice/versions/3.8.0/topics/mice.

Cellini, S. R., & Turner, N. (2016). Gainfully Employed? Assessing the Employment and Earnings of For-Profit College Students Using Administrative Data. doi: 10.3386/w22287

Chen, R., & Wiederspan, M. (2014). Understanding the determinants of debt burden among college graduates. *The Journal of Higher Education*, *85*(4), 565-598.

City And Town Population Totals. *United States Census Bureau*. *2010-2018*. 2019.

Cochrane, D., & Szabo-Kubitz, L. (2014). At What Cost? How Community Colleges That Do Not Offer Federal Loans Put Students at Risk. Issue Brief. *Institute for College Access & Success*.

CollegeBoard. (2019). *Federal Loans: Repayment Rates*. Retrieved from https://research.collegeboard.org/pdf/2019-trendsinsa-figs-13a-13b.pdf\

Daly, L. (2019, April 14). Getting a Student Loan to Attend a For-Profit College? Here's Why You Should Think Twice. Retrieved from https://www.fool.com/student-loans/getting-student-loan-to-attend-for-profit-college-heres-why-you-should-think-twice/.

Deeper In Debt: Women & Student Loans – AAUW : Empowering Women Since 1881. [online] Available at: <https://www.aauw.org/resources/research/deeper-in-debt/> [Accessed 17 April 2020].

Dvorkin, E., Bowles, J., & Shaviro, C. (2018). Deeper in Debt: For-Profit Schools Driving Student Loan Default in New York State. Data Brief. *Center for an Urban Future*.

Dynarski, S. (2020). An Economist's Perspective On Student Loans In The United States. *Brookings Institute*, https://www.brookings.edu/wp-content/uploads/2016/06/economist_perspective_student_loans_dynarski.pdf.

Gross, J. P., Cekic, O., Hossler, D., & Hillman, N. (2009). What Matters in Student Loan Default: A Review of the Research Literature. *Journal of Student Financial Aid*, *39*(1), 19-29.

Hausman Test for Endogeneity (Hausman Specification Test). (2018, May 24). Retrieved from https://www.statisticshowto.com/hausman-test/.

Herr, E., & Burt, L. (2005). Predicting Student Loan Default for the University of Texas at Austin. *Journal of Student Financial Aid*, *35*(2), 27-49.

Hillman, N. W. (2015). Borrowing and repaying student loans. *Journal of Student Financial Aid*, *45*(3), 5.

Historical Income Tables: Households.*United States Census Bureau*. 2020.

Ishitani, T. T., & McKitrick, S. A. (2016). Are Student Loan Default Rates Linked to Institutional Capacity?. *Journal of Student Financial Aid*, *46*(1), 3.

Jackson, B. A., & Reynolds, J. R. (2013). The price of opportunity: Race, student loan debt, and college achievement. *Sociological Inquiry*, *83*(3), 335-368.

Kelchen, R., & Li, A. Y. (2017). Institutional Accountability: A Comparison of the Predictors of Student Loan Repayment and Default Rates. *The ANNALS of the American Academy of Political and Social Science*, 671(1), 202–223.

Kreighbaum, A., 2020. The Link Between Completion And Loan Repayment. [online] Insidehighered.com. Available at: <https://www.insidehighered.com/news/2018/08/08/link between-college-completion-and-student-loan-repayment>

Lee, S. (2020). *The For-Profit Higher Education Industry, By the Numbers — ProPublica*. ProPublica. Retrieved 8 April 2020, from https://www.propublica.org/article/the-for-profit-higher-education-industry-by-the-numbers.

Li, W. (2013). *The Economics of Student Loan Borrowing and Repayment*. Federal Reserve Bank of Philadelphia. Retrieved from https://philadelphiafed.org/-/media/research-and-data/publications/business-review/2013/q3/brq313_economics-of-student-loan-borrowing-and-repayment.pdf

Lochner, L., & Monge-Naranjo, A. (2004). *E*ducation and default incentives with government student loan programs. *National Bureau of Economic Research.*

Lochner, L. J., & Monge-Naranjo, A. (2014). Default and repayment among baccalaureate degree earners (No. w19882). *National Bureau of Economic Research.*

Lochner, L., & Monge-Naranjo, A. (2016). Student loans and repayment: Theory, evidence, and policy. In *Handbook of the Economics of Education* (Vol. 5, pp. 397-478). Elsevier

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 Papers on Economic Activity*, *2015*(2), 1–89. doi: 10.1353/eca.2015.0003

MeasureOne. (2019). *The MeasureOne Private Student Loan Report*. Retrieved from https://docs.wixstatic.com/ugd/0aaff0_5a3f9e397ee247e5b33454f5c014c89d.pdf

Mishel, L., & Schieder, J. (n.d.). CEO pay remains high relative to the pay of typical workers and high-wage earners. Retrieved from https://www.epi.org/publication/ceo-pay-remains-high-relative-to-the-pay-of-typical-workers-and-high-wage-earners/

Moretti, E. (2004). Workers Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions. *American Economic Review*, *94*(3), 656–690. doi: 10.1257/0002828041464623

Nova, A. (2019, October 23). Bill would make for-profit colleges ineligible for federal student loans. Retrieved from https://www.cnbc.com/2019/10/23/bill-would-make-for-profit-schools-inelgible-for-financial-aid.html

United States. Senate. Committee on Health, Education, Labor & Pensions. For-Profit-Report. *Executive Summary*. Web. 2012.

Simkovic, M. (2013). Risk-based student loans. *Wash. & Lee L. Rev.*, *70*, 527.

Simon, S., & Emma, C. (2014, March 14). Obama hampers for-profit colleges. Retrieved from https://www.politico.com/story/2014/03/barack-obama-education-for-profit-colleges-104661

Steiner, M., & Teszler, N. (2005). Multivariate Analysis of Student Loan Defaulters at Texas A&M University. *TG (Texas Guaranteed Student Loan Corporation)*.

Stiglitz, J. E. (1975). The theory of" screening," education, and the distribution of income. *The American economic review*, *65*(3), 283-300.

Volkwein, J. F., & Cabrera, A. F. (1998). Who defaults on student loans? The effects of race, class, and gender on borrower behavior. In R. Fossey & M. Bateman (Eds.), *Condemning students to debt: College loans and public policy* (pp. 105-126). *New York: Teachers College Press*.

Webber, Karen L., and Sharon L. Rogers. "Student loan default: Do characteristics of four-year institutions contribute to the puzzle?." *Journal of Student Financial Aid* 44.2 (2014): 2.

Weiss, A. (1995). Human Capital vs. Signalling Explanations of Wages. *Journal of Economic*

Perspectives, *9*(4), 133–154. doi: 10.1257/jep.9.4.133

Wiswall, M., & Gemici, A. (2011). Evolution of Gender Differences in Post-Secondary Human Capital Investments: College Majors. *Institute for Education and Social Policy (IESP) Working Paper*, (03-11).