

10-15-2019

## Social Dimensions of Student Debt: A Data Mining Analysis

Dirk Witteveen

University of Oxford, Nuffield College, [dirk.witteveen@nuffield.ox.ac.uk](mailto:dirk.witteveen@nuffield.ox.ac.uk)

Paul Attewell

City University of New York, The Graduate Center, [pattewell@gc.cuny.edu](mailto:pattewell@gc.cuny.edu)

Follow this and additional works at: <https://ir.library.louisville.edu/jsfa>



Part of the [Educational Sociology Commons](#), [Education Economics Commons](#), [Higher Education Commons](#), [Higher Education Administration Commons](#), and the [Inequality and Stratification Commons](#)

---

### Recommended Citation

Witteveen, Dirk and Attewell, Paul (2019) "Social Dimensions of Student Debt: A Data Mining Analysis," *Journal of Student Financial Aid*: Vol. 49 : Iss. 1 , Article 1.

Available at: <https://ir.library.louisville.edu/jsfa/vol49/iss1/1>

This Research Article is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in *Journal of Student Financial Aid* by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. For more information, please contact [thinkir@louisville.edu](mailto:thinkir@louisville.edu).

---

## Social Dimensions of Student Debt: A Data Mining Analysis

### Cover Page Footnote

We thank the anonymous reviewers for their insightful comments and suggestions. All research has been conducted at the City University of New York, The Graduate Center.

# Social Dimensions of Student Debt: A Data Mining Analysis.

By Dirk Witteveen and Paul Attewell

*Media commentary on undergraduates' loan debt portrays a crisis in which many students are unable to pay back their loans, having borrowed large sums and lacking sufficient post-college income to repay. Several scholars have questioned the media accounts, noting that indebtedness is highest among students from high income families, while defaults predominate among low debt students. Using a data mining technique known as CART, we analyze national data on the indebtedness of recent baccalaureate graduates, to uncover combinations of social characteristics that are associated with loan pressure: the ratio of indebtedness to post-college earnings. We find that students from lower income families who attend expensive institutions - especially for-profit colleges - accumulate high debt. In contrast to earlier scholarship, after controlling for the net cost of attending a college, we find that lower-income students face much higher loan pressure than students from more affluent families.*

**Keywords:** *higher education, student debt, loan burden, social inequality*

Total student loan debt in the US has risen to over 1.36 trillion dollars, exceeding the nation's total credit card debt (Federal Reserve Bank of NY, 2017). This loan burden has been portrayed as a social problem, with high educational debt linked to delays in marriage and dropping rates of home ownership, or a “disruption of the intergenerational social compact” among other ills (Best & Best, 2014; Brown & Caldwell 2013; Chakrabarti, Gorton & van der Klaauw, 2017; Dworkin, 2012; Elliot & Johnson, 2014; Luong, 2010).

The image of student indebtedness as a problem contrasts with a conventional economic perspective that views student loans – like credit more generally – as playing a very important positive social function (Akers & Chingos, 2016; Baum, 2016). Federal student loans provide financial resources to individuals who could not otherwise afford to go to college, and hence lower the financial barriers to college entry. For students already in college, loans help with educational and living expenses, and (for some) lessening the need to work for pay in term time.

In principle, then, loans should improve undergraduate retention and degree completion.

Ideally, scholars would be able to assess the positive and negative consequences of educational loan indebtedness, quantifying the overall balance, and identifying the characteristics of individuals who gain as well as pinpointing those for whom indebtedness outweighs benefits. In attempting to do this, however, scholars have encountered a puzzle. They expected students with loans would be more likely to stay in

---

*Dirk Witteveen is a Postdoctoral Research Fellow at Nuffield College, University of Oxford. Paul Attewell is a Professor in Sociology and Urban Education PhD Program at The Graduate Center, City University of New York.*

college and complete a degree than otherwise similar students who did not obtain loans, but most studies have failed to find any such positive impact of receiving loans on undergraduate retention and degree completion.

Reviewing numerous studies, Hossler, Dundar, and Shapiro (2013: 158) note: "...student financial assistance in the form of grants has a positive effect on retention, and loans have little or no impact." (See also Alon, 2007; Desjardins, Alberg & McCall 2002; Perna, 1998; Singell, 2004; St. John, 1989; Hossler, Ziskin, Gross, Kin & Cekic, 2009). Alon (2007: 306) using an instrumental-variables approach to address selection issues similarly concludes: "Loan dollars have no impact on graduation, even after controlling for aid eligibility." For an earlier generation of research, Perna (1998) found neither direct nor indirect effects of loans on bachelor's degree completion; she also cited studies prior to her own that either found that loans were unrelated to persistence or in one study associated with lower persistence. In sum, the puzzle of loans and college graduation remains unresolved: loan debt has surged over time, yet a graduation payoff to taking out loans has yet to be convincingly documented.

An alternative question regarding the impact of student loans asks how the loan debt that an undergraduate accumulates is related to that student's post-college earnings. There are several reasons why loan debt and post-college earnings might be positively associated. A student might choose to attend a more expensive college requiring more borrowing but leading to a more valuable credential and better-paid job. Alternatively, students might take out loans to enable them to work less for pay during the school year, freeing them to study more, or to progress towards the degree more rapidly. That choice might also pay off in better earnings post-college compared to students who avoid loan debt.

On the other hand, some students who graduate with large loans might face financial pressure to take a job immediately after graduation, because of their need to start paying off those loans, while their counterparts with less debt might feel freer to delay taking a job until they can find a better-paying position, or even decide to continue their education. If that scenario were common, higher undergraduate loan indebtedness might be associated with *lower* rather than higher post-college earnings. A related possibility is that some undergraduates who have accumulated large loan balances prior to completion might decide to stop out of college in order to work, and thus avoid increasing their debt level. But once stopped out, about half of students do not return and complete the credential (DesJardins, Ahlburg & McCall 2006). Under that scenario high debt early in college might be associated with lower post-college earnings.

In the absence of compelling theory as to the relationship of loan indebtedness and post-college earnings, and with several plausible mechanisms implying opposite effects of loans, in this paper we opt for a data mining strategy, an exploratory approach that searches for combinations of factors that are associated with undergraduates incurring either higher or lower loan debt. In particular, we examine the relationship between the dollar amount of loans (at graduation) and earnings several years after graduating with a baccalaureate. We compute a measure that we call "loan pressure" as the ratio of a student's loan debt to annual earnings, an indicator of the difficulty in paying off loans. We then search for combinations of institutional and student factors that are associated with higher or lower loan pressure. Finally, we unpack the complex relationship between parental background and student loan debt.

## Previous Literature

Two recent books on student debt provide reviews of research and discuss media commentary on student loans. Akers and Chingos (2016) are scholars at the Brookings Institution and Baum (2016) is a fellow at the Urban Institute and a senior author of the College Board's annual publication, *Trends in Student Aid*. It is striking that the subtitles of both books contrast the media's rhetoric to the reality of student aid and debt, a framing that highlights their shared conviction that much media discussion of student loans is profoundly

misinformed and has misled the public. For example, Akers and Chingos (2016: 4) assert: “There is no evidence of a widespread, systematic student loan crisis, in which the typical borrower is buried in student debt for a college education that did not pay off. The crisis that permeates public discussion is a manufactured narrative based largely on anecdotes, speculation, shoddy research, and inappropriate framing of the issue.”

Educational debt has grown considerably in recent decades, but undergraduate loan debt remains manageable for the large proportion of bachelor’s graduates – generally defined as 10% to 20% of one’s annual income (Baum & Schwartz 2006). One in three baccalaureates complete with no debt, and among those who do borrow, the average cumulative debt for the baccalaureate is \$29,600 (Velez, Woo & Simone, 2017: Tables A-1 and A-2). Government regulations effectively limit the maximum amount that can be borrowed by undergraduates from federal programs (Akers & Chingos, 2016). At the same time, the ratio of financial aid (merit- or needs-based) to tuition costs has remained substantial – covering between 20 and 60 percent depending on type of institution. This ratio has also slightly decreased in the last decade (Snyder, de Brey & Dillow, 2017).

The debt situation is quite different for students who continue into masters, doctoral and professional degree programs, in part because federal regulations allow much higher borrowing limits for graduate students (Akers & Chingos, 2016). About 79 percent of master’s students incur loans, averaging \$55,400, and 90 percent of professional doctorates take out loans, averaging \$131,000 (Velez et al., 2017).

Among undergraduates, levels of indebtedness also vary according to the type of institution attended. Government figures for 2015-2016 (Radwin et al., 2018) refer to students at all levels from freshmen to seniors (i.e., not just graduates) and report that students enrolled in four-year for-profit institutions are most likely to take out loans (68 percent) and the amount they borrow averages \$8,800. Students in four-year private non-profit colleges have the next highest proportion (58 percent with loans, averaging \$9,300). Students at public four-year colleges (47 percent with debt, averaging \$7,500) and students at community or two-year colleges (14 percent with loans, averaging \$4,700) have markedly lower levels of student loan debt.

Total accumulation or loan amounts are higher among undergraduates from affluent families than from lower-income undergraduates (Baum 2016, Baum 2018). This is the case for the bivariate relationship between family income and student debt. However, in multivariate analyses presented below we will show that the picture changes substantially after one controls for the cost of education, at which point lower income students are more indebted than their higher-income classmates.

Rather than a ‘loan crisis,’ Dynarsky and Kreisman (2016) argue that the current problem is better characterized as a repayment crisis. This follows from the fact that default is **not** most common among those with high loan balances. Instead, default rates are high among students with relatively low balances, notably among those who attended for-profit or community colleges and whose post-college earnings are low (Dynarsky, 2016). Thus, undergraduate default appears partly to be a reaction among those for whom college did not lead to enhanced earnings, rather than a consequence of heavy levels of debt or heavy monthly repayment schedules.

Both Akers and Chingos (2016) and Baum (2016) suggest that the debt problem is greater among students who do not complete their undergraduate degrees. Our reading of government data (Wei, Horn & Soldner, 2013) suggests a different interpretation: non-completers at public two-year colleges accumulate an average loan debt of \$5700, compared to \$12,500 for 2-year degree completers. Similarly, non-completing public four-year students borrow less than their counterparts who graduate (\$9,300 compared to \$15,100) (Wei et al., 2013). In part this pattern occurs because non-completers don’t stay in college as long as completers. But even when comparing borrowing per credit completed, non-completers borrow less than

completers (Wei et al., 2013), raising the possibility that non-completion may result in some cases from students borrowing too little (loan aversion) rather than too much (cf. Britt, Ammerman, Barrett & Jones, 2017).

Although non-completers have lower debt on average than completers, non-completers' *default rates* are much higher than those of completers. In Scott-Clayton's (2018: Table 2) recent analyses, nearly 40 percent of non-completers default, compared to 14 percent of associate degree graduates and 5.6 percent of bachelor's graduates. It is also striking that 52 percent of non-completers who attended for-profits defaulted (Scott-Clayton, 2018: Table 4).

In sum, prior research indicates that there is a serious and intensifying default problem, frequently originating among non-completers with relatively low levels of debt. Default rates reach their highest levels among two groups: the first are non-completers who attended for-profit colleges. The other group consists of African-American students, both non-completers and graduates, who also have markedly higher debt levels than other racial and ethnic groups, net of controls (Grinstein-Weiss, Perantie, Taylor, Guo & Raghavan, 2016; Scott-Clayton, 2018).

### Analytical Strategy

Prior scholarship demonstrates that the loan landscape is quite complex, with very different patterns of borrowing and repayment evident for non-completers, graduates, and graduate students, as well as contrasts across institutional types and demographic groups. In order to focus our inquiry, and make the analysis more tractable, we limit this paper to just one of these populations -- recent graduates with baccalaureate degrees. We then look at variability of loan amounts and loan pressure for that group, and relate this to dimensions including type of institution, student demographic characteristics, and student academic performance. In particular we search for combinations of factors that when they co-occur are associated with especially high and especially low loan pressure or loan amount, since these combinations of characteristics identify subpopulations whose loans are the most and the least problematic.

The data are drawn from the federal Baccalaureate and Beyond Study (B&B); a nationally-representative sample of persons who graduated with a bachelor's degree in 2007/2008 and then re-surveyed the respondents four years later, in 2012. The initial interview, conducted in the last year of college, contains questions about demographics, course-taking, and achievement. The follow-up in 2012 adds several questions regarding employment and earnings. The B&B data were obtained from the National Center for Education Statistics under a restricted license that requires us to round sample sizes reported below to the nearest ten, to ensure data confidentiality.<sup>1</sup>

We selected all 14,110 students (a rounded figure) who completed the survey in 2008 for our study of college loan-taking. For our analyses of labor market outcomes, we selected only those individuals who obtained their bachelor's degree in 2008, were not enrolled in education when interviewed in 2012, and were employed in 2012. This leads to a second study sample of 8,120 graduates. Appendix A documents the descriptive statistics regarding loan-taking of these excluded groups. We did not observe noteworthy discrepancies between the excluded groups and our study sample. Results are only generalizable to the population of the selected study sample. However, it should be noted that the labor market sample includes a small portion of graduates who (already) finished an advanced degree. This may result in an *underestimation* of the reported undergraduate loan to earnings ratio because advanced degree holders are likely to earn more.

---

<sup>1</sup> More information about the 2008/2012 B&B can be found at: <http://nces.ed.gov/surveys/b&b/about.asp>

We initially focus on three important dependent variables: whether or not a student has taken out loans ('loan probability'); the loan amount; and loan pressure. Both 'loan probability' and loan amount refer to the amount borrowed up until students' BA graduation. Any post-graduate loans are excluded from this analysis. We construct a loan pressure variable by calculating the ratio of the loan debt at college graduation (in 2008) to current (2012) earnings among employed individuals. After exploring the complex associations between individual- and college-related factors and loan outcomes, we further concentrate on the intergenerational component of student loan risks. Appendix B presents the descriptive statistics of the parental background groups (organized by family income deciles) as observed during college.

The modeling method we use is a Classification and Regression Tree (CART), also known as "decision trees," "classification trees," or "data partitioning" (Breiman, Friedman, Olshen, Stone, Steinberg & Colla, 1983). This is a non-parametric and computationally-intensive method that searches for combinations of values across several predictors which together define contrasting values on a dependent variable. In our context, this means finding distinctive combinations of student attributes that are associated with different levels of debt, or different levels of debt pressure. We include the following predictors in each model: time to degree, time between high school and college entry, race/ethnicity, gender, parental education, parental income, independent student, type of institution, college selectivity, SAT score, number of children, GPA, Carnegie category, major, age at entry, and age at graduation. In this exploratory analysis (using CART) we do not formulate specific hypotheses for the effects of each individual variable on loan-taking and loan burden. Instead, variables are selected based on their importance for various college attainment outcomes as found in the literature. One important advantage of CART in comparison with traditional predictive modeling techniques is that the model is not sensitive to multicollinearity.

There are a huge number of combinations of values on these predictors that might potentially be associated with debt level or debt pressure. What CART does is systematically work through the possible combinations, in an iterative process, finally sorting students into different groups which maximally contrast on the dependent variable (e.g., on loan amount or loan pressure). These final groups (known as 'terminal nodes' or 'leaves') are mutually exclusive: each student appears in one and only one leaf. The groups are constructed statistically to be as internally homogenous as possible on the dependent variable, so students within a group have relatively similar levels of loan debt, for example. But at the same time the groups are constructed to be as different from one another as possible: the inter-group variance on the dependent variable is maximized. The end result is a spectrum of contrasting groups, going from the most indebted to the least (or the highest loan pressure to the lowest).

In addition to their contrasting levels of loans or loan pressure, what is of particular interest to the analyst is how those groups/leaves are defined: what combinations of values on predictors define each of the groups? Some of the predictors will not be chosen to define groups because they are not strong predictors. In this aspect CART is analogous to stepwise regression. Other predictors may define certain groups/leaves but not others. Where CART differs from conventional regression is that its aim is not to estimate the coefficient or contribution of each single predictor, but rather to identify the relationship of specific combinations of student characteristics to an outcome. Therefore, CART is one of the best suitable techniques to study an outcome variable for which its associations are assumed to be highly complex (and/or undertheorized).

## Findings

### Loan Probability

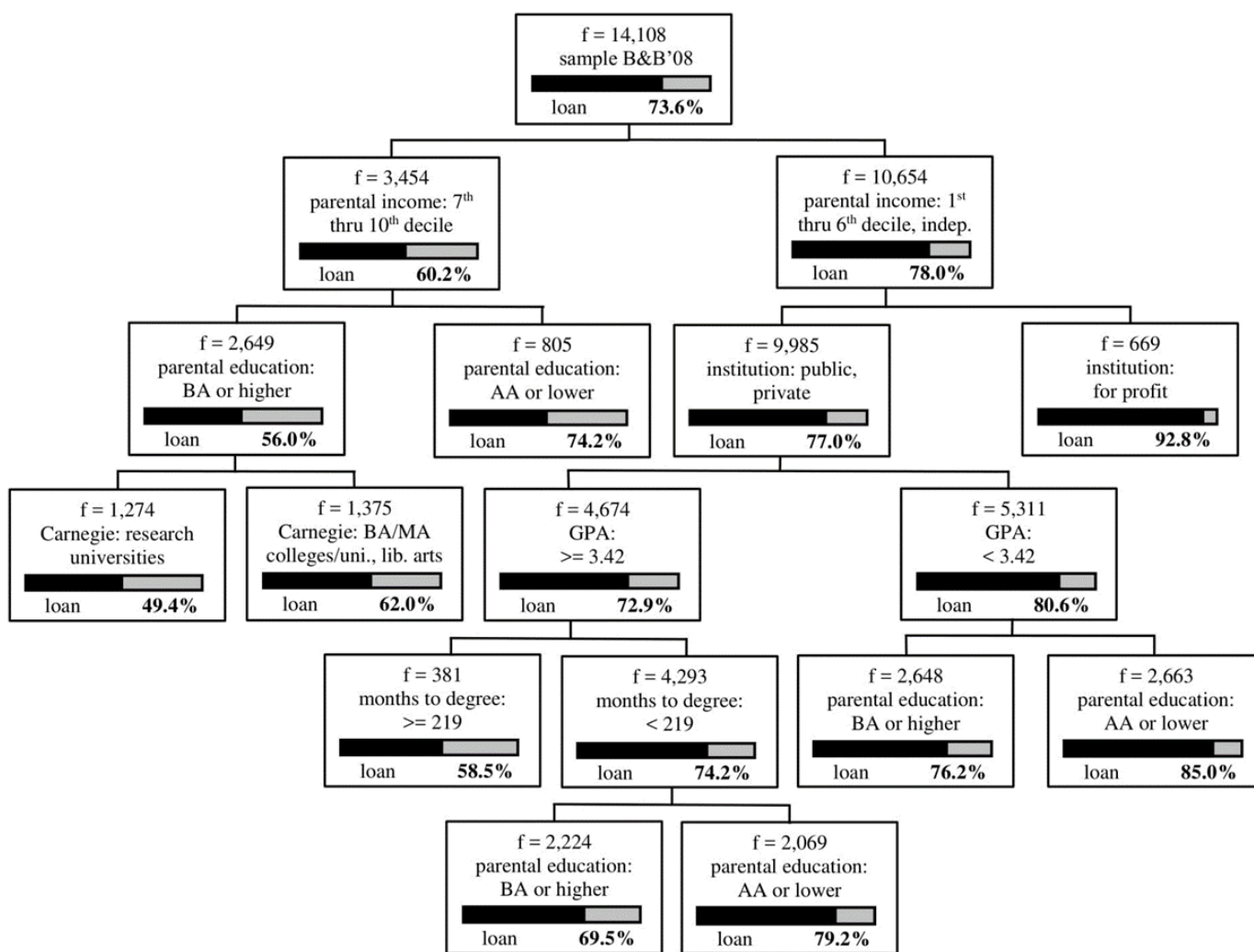
Table 1 presents the results of a Classification and Regression Tree (CART) analysis of who had any student loan debt at the time of their baccalaureate graduation in 2008. CART identifies a subset of values on

multiple predictors that in combination define certain groups of students who have similar levels of loan debt. One can think of the combination of particular values on predictors as a sociological profile. Within CART, which uses a tree metaphor, the groups defined by distinct combinations of predictors are termed “leaves.”

In Table 1, each group (called ‘leaf’) contrasts maximally with each other leaf in terms of proportion of baccalaureate graduates with student loan debt. This particular CART analysis identified nine contrasting leaves representing a spectrum from groups where almost all students took loans to groups where most did not. Figure 1 presents all the splits of the CART algorithm, whereby the nine exclusive leaves can be found at the end of each branch.

Figure 1

*Decision Tree Predicting Having a Loan at College Graduation.*



Source. Baccalaureate and Beyond 2008 – 2012. N = 14,110 (a rounded figure).

Notes. Predictors in model: time to degree, time between high school and college entry, race/ethnicity, gender, parental education, parental income, independent student, type of institution, college selectivity, SAT score, number of children, GPA, Carnegie category, major, age at entry, and age at graduation. Number of splits: 8. Validation: 30%.



Leaf 1 represents the group with the highest proportion of graduates with student loans. This group is relatively small in size (under 5 percent of all graduates) but is distinctive in that almost 93 percent of the students with this profile had borrowed. All the students in this group came from lower- or lower-middle income families (parents in the 1<sup>st</sup> thru 6<sup>th</sup> decile) or were independent when they entered higher education. In addition, all these students had graduated from a for-profit institution.

Second-highest in order of loan propensity is 'Leaf 2'. This group contains similarly lower- to middle class students, with the additional specification of having parents who held an AA or less. They attended either private or public colleges (but not for profit) and had a college GPA of less than 3.42. Nearly 19 percent of the sample are clustered in this group. Leaf 3 (14.7% of the sample) is almost identical to Leaf 2, differing on only one characteristic: they took less than 219 months to graduate. Students with these profiles had an 85% chance and a 79.2% chance to have taken out loans before graduation, respectively.

As Leaf 4 is relatively large and in the middle of the decision tree leaves ranked by loan percentage, we consider it the most typical for four-year college students. It consists of students who graduated from a public or private college and came from families with *higher* educated parents and incomes in the lower thru lower-middle class.

When contrasting these students with the other end of the spectrum (Leaf 9), we find a smaller group of students – 9 percent of all graduates – whose parents had BA degrees (or higher) and incomes in the upper-middle class (7<sup>th</sup> decile and higher).<sup>1</sup> Moreover, they had attended research universities. Although this relatively affluent group is least likely to have taken out a loan prior to graduation, it is still surprising that almost half of them (49.4 percent) nevertheless did so.

The average loan probability is substantially higher in Leaves 7 (62%) and 8 (58%). The former group shares most characteristics with Leaf 7, with exception of having attended slightly different types of colleges, such as liberal arts colleges and universities I & II. Leaf 8, however, shows a narrow path (only 2.7% of the sample) to a lower loan probability for lower- and lower-middle class students: by high achievement (a GPA of 3.42 or higher) and not taking a very long time to graduate.

To summarize, this first CART analysis identified four factors that in combination predict the likelihood that a student will graduate with loan debt: parental education and income in combination with the type of college attended, and the academic skills (college GPA) of the student. The combination of attending a for-profit college and coming from a family with lower parental income makes it almost certain that an undergraduate will graduate with loan debt. Conversely, the combination of a student's high academic skills and affluent parents is associated with the lowest likelihood of student loan indebtedness.

Table 1

*CART Leaves Predicting Having a Loan Debt at College Graduation (2008)*

<b>Leaf</b>	<b>group size</b>	<b>percent loan</b>
1 parental income (1 <sup>st</sup> thru 6 <sup>th</sup> decile, or independent), institution (for profit)	4.7%	92.8%
2 parental income (1 <sup>st</sup> thru 6 <sup>th</sup> decile, or independent), parental education (AA or lower), institution (public, private), GPA (<3.42)	18.9%	85.0%
3 parental income (1 <sup>st</sup> thru 6 <sup>th</sup> decile, or independent), parental education (AA or lower), institution (public, private), GPA (>=3.42), months to degree (<219)	14.7%	79.2%
4 parental income (1 <sup>st</sup> thru 6 <sup>th</sup> decile, or independent), parental education (BA or higher), institution (public, private), GPA (<3.42)	18.8%	76.2%
5 parental income (7 <sup>th</sup> thru 10 <sup>th</sup> decile), parental education (AA or lower)	5.7%	74.2%
6 parental income (1 <sup>st</sup> thru 6 <sup>th</sup> decile, or independent), parental education (BA or higher), institution (public, private), GPA (>=3.42), months to degree (<219)	15.8%	69.5%
7 parental income (7 <sup>th</sup> thru 10 <sup>th</sup> decile), parental education (BA or higher), Carnegie (lib. arts, general, universities I & II, special)	9.8%	62.0%
8 parental income (1 <sup>st</sup> thru 6 <sup>th</sup> decile, or independent), institution (public, private), GPA (>=3.42), time to degree (>=219 months)	2.7%	58.5%
9 parental income (7 <sup>th</sup> thru 10 <sup>th</sup> decile), parental education (BA or higher), Carnegie (research universities)	9.0%	49.4%
<b>Overall</b>	<b>100.0%</b>	<b>73.6%</b>

*Source.* Baccalaureate and Beyond 2008 – 2012. N = 14,110 (a rounded figure)

## Loan Amount

Table 2 reports a different CART analysis, now predicting dollar loan amount at graduation. The leftmost columns for group size and loan debt are our main focus. Additional columns report mean annual out-of-pocket expenses and mean hours worked during college. These two variables were not part of the CART analysis itself but are provided to help us interpret or contextualize the CART findings.

Leaf 1 contains 15 percent of graduates who form the group with the lowest average loan amount. These are all students who attended public universities, who graduated by age 24 and within four years plus a semester (less than 52 months) of entering college. They also have college GPAs of B+ or better. Their average debt at graduation was \$10,139.

These lowest-debt graduates are academically strong students who went to public colleges that are relatively inexpensive, averaging \$3,214 after subtracting aid and scholarships from the cost of attendance. These students worked on average 15 hours a week, which is at the low end for this cohort of undergraduates, and they completed their baccalaureates roughly on time (within four years and one semester). So academically strong students who completed their degrees quickly at low-cost public colleges accumulate the least amount of loans.

Leaf 2 in Table 2 contains a similar group of students but who attended private non-profit colleges and took longer to graduate ( $\geq 52$  months) but were nevertheless under 24 years old at graduation. They also had GPAs of around B+ or better. These students had higher debt than Leaf 1, averaging \$13,511, but this was not due to a higher cost of education – their annual out of pocket expenses were slightly lower than their counterparts in the first leaf. Their higher loan debt reflects their longer time to degree.

Moving to the other end of the spectrum, the group of graduates with highest loan debt was constituted of graduates from for-profit colleges (Leaf 9). They had accumulated by far the highest loan debt (\$35,597 on average). We note that these graduates of for-profits accumulated high debt despite working the most hours of all (32 hours per week on average). Leaf 8 includes students who attended a non-profit private college and who came from families with an AA degree or less and whose GPA was not strong ( $< 3.41$ ). They had a high loan debt of \$28,608 by graduation.

Leaf 7 includes students at private non-profit colleges with lower GPAs and whose parents had a BA or higher. This group had the highest out of pocket expenses, presumably because they attended relatively expensive private colleges, but they were not academically strong enough to obtain scholarship support. This group owed \$22,854 on average at graduation.

One large group at the middle of the spectrum is worth noting. Graduates at public universities who graduated aged 24 or older – so they either delayed entry to college or took considerable time to graduate – had a mean loan debt of \$19,153. That is roughly double the debt of the public university graduates in leaf one, who completed college within 4.5 years, but it is a lot less loan debt than the graduates from for-profit colleges.

To summarize Table 2, particular combinations of type of institution attended, parental education, and student's GPA are associated with amount of loan debt, with time to degree having some additional relationship.

Table 2

*Means by CART Leaves Predicting Loan Amount at College Graduation (2008).*

<b>Leaf</b>		<b>group size</b>	<b>loan debt</b> (at college graduation)	<b>annual out of pocket costs</b>	<b>weekly hours worked in college</b>
1	institution (public), age at graduation (<24), GPA (=>3.24), months to degree (<52)	15.1%	\$10,139	\$3,214	15.0
2	institution (private), age at graduation (<24), GPA >=3.24, months to degree (>=52)	7.6%	\$13,511	\$2,821	18.9
3	institution (private), GPA (>=3.41), parental education (BA or higher)	11.1%	\$15,224	\$10,202	15.9
4	institution (public), age at graduation (<24), GPA (<3.24)	14.6%	\$15,701	\$3,804	18.8
5	institution (public), age at graduation (>=24)	20.3%	\$19,153	\$1,957	23.9
6	institution (private), GPA (>=3.41), parental education (AA or lower)	8.4%	\$21,876	\$7,244	23.3
7	institution (private), GPA (<3.41), parental education (BA or higher)	9.6%	\$22,854	\$12,750	17.4
8	institution (private), GPA (<3.41), parental education (AA or lower)	8.2%	\$28,608	\$8,870	23.7
9	institution (for profit)	5.1%	\$35,597	\$9,732	32.0
	<b>overall</b>	<b>100.0%</b>	<b>\$18,618</b>	<b>\$5,807</b>	<b>20.2</b>

*Source.* Baccalaureate and Beyond 2008 – 2012. N = 14,110 (a rounded figure).*Notes.* Annual out of pocket = all tuition and fees minus all grants and scholarships. Weekly hours worked = average work and work-study hours in the last year in college (2007 – 2008)

## Loan Pressure

Graduating with a degree and with sizeable loan debt is not in itself a good or a bad thing. If one's degree subsequently allows one to earn a good wage, then the loan may prove to be well worthwhile. Conversely, if one's diploma does not open the way to a job, then even small amounts of debt may prove economically problematic.

We analyzed this issue using the construct of "loan pressure" – the ratio of undergraduate loan debt at graduation divided by earnings, where the latter were measured four years after baccalaureate graduation. Using earnings four years after graduation captures the earnings power of each individual after they have had some time to establish themselves in the labor market. A loan pressure of .5 means that loan debt at graduation is equivalent to half a year's salary in 2012, and .25 a quarter of a year's salary, and so on.

Two caveats are in order before presenting the results from Table 3. First, we excluded from this analysis all individuals who were still enrolled in college or university in 2012. These were baccalaureates enrolled in graduate school taking masters or higher degrees. Loan debt is certainly important for such individuals, but, in our judgement, it would be misleading to treat graduate students' earnings in 2012 as if those were a measure of long-term earning power. Their inclusion in the CART would seriously obscure the analysis for baccalaureates.

Second, we omitted some individuals because they were far outliers: their loan pressure ratio was more than three standard deviations above the mean for the sample as a whole. These outlier individuals had very high loan pressure in nearly all cases because they were earning little or nothing in 2012. Although these extreme situations are worth exploring more for our understanding of loan pressure, we chose to present analyses that correctly reflect the financial status of a larger group of graduates. If the outliers were included the loan pressure averages of the high loan pressure leaf would be unrealistically reflecting the (highest) burden from student loans on currently employed graduates.

Table 3 presents the leaves from a third CART analysis, where the outcome variable is the ratio of loan debt at graduation to earnings four years later. It identified nine groups or leaves. The group with the lowest loan pressure accumulated debt equivalent to a quarter (25.5 percent) of annual earnings, while the group with the highest pressure had loans equivalent to about 83 percent of earnings.

Table 3

*Means by CART Leaves Predicting Loan Pressure Four Years after Graduation (2012), among Employed & Non Enrolled.*

Leaf		group size	loan pressure	loan debt (at college graduation)	earnings in 2012	annual out of pocket costs
1	institution (public), months to degree (<49), GPA (>=3.23)	15.1%	.255	\$10,420	\$50,173	\$3,366
2	institution (private, for profit), parental education (BA or higher), SAT (4 <sup>th</sup> quartile, no record)	7.2%	.344	\$15,603	\$56,563	\$11,595
3	institution (public), months to degree (>=49), major (engineering, computer / information, business, health)	13.9%	.347	\$16,377	\$59,253	\$2,366
4	institution (public), months to degree (<49), GPA (<3.23)	6.8%	.369	\$14,113	\$46,965	\$4,594
5	institution (public), months to degree (>=49), major (bio/phys science, tech., math, agriculture, other applied, soc. science, general, education, humanities)	23.8%	.485	\$16,911	\$42,559	\$2,522
6	institution (private, for profit), parental education (BA or higher), SAT (bottom 3 quartiles)	13.9%	.498	\$20,516	\$50,630	\$11,113
7	institution (private, for profit), parental education (AA or lower), GPA (>=3.18), major (computer / information, engineering, business, bio/phys science, tech., math, agriculture, health, soc. science)	9.9%	.512	\$23,562	\$57,164	\$7,071
8	institution (private, for profit), parental education (AA or less), GPA (>=3.18), major (education, other applied, humanities, general, other)	4.4%	.737	\$27,255	\$43,093	\$8,901
9	institution (private, for profit), parental education (AA or less), GPA (<3.18)	5.1%	.835	\$31,582	\$45,927	\$9,654
<b>Overall</b>		<b>100.0%</b>	<b>.446</b>	<b>\$17,931</b>	<b>\$50,090</b>	<b>\$5,681</b>

*Source.* Baccalaureate and Beyond 2008 – 2012. Authors' calculations of a subsample of college graduates employed and not enrolled in education in 2012. N = 8,120 (a rounded figure).

*Notes.* About 200 outliers (>3 times the standard deviation on loan pressure) were eliminated before running the decision tree analysis. Annual out of pocket = all tuition and fees minus all grants and scholarships. Weekly hours worked = average work and work-study hours in the last year in college (2007 – 2008). Loan pressure = loan debt at graduation / earnings in 2012. In this table, the ratio of group means of loan debt and earnings does not equal loan pressure (Simpson's Paradox)

The group with the lowest loan pressure (Leaf 1) graduated “on time” (less than 49 months) from a public college and came from families with a parent with a BA degree or higher. Their earnings were not especially high, nor were their annual out of pocket expenses during college particularly low. What caused their loan pressure to be low (.255) was that they graduated with the lowest debt of any of the leaf groups, averaging about \$10,000 (substantially lower than the grand mean of \$17,931).

The next lowest group in terms of loan pressure (.344, Leaf 2) are graduates from either private non-profit or for-profit colleges, who came from highly-educated families, and reported high SAT scores prior to college entry. Their colleges cost over three times as much as the public college graduates of leaf one. Nevertheless, their accumulated loans were modest: \$15,603 compared to \$10,420 in Leaf 1. This is perhaps because of financial assistance from their affluent parents. (The survey does not provide data on this issue.)

At the other extreme, the group of undergraduates with the highest loan pressure (Leaf 9) were those who came from lower-educated families and attended either private non-profit or for-profit institutions. Their annual out-of-pocket educational expenses were above average (\$9,654), yet certainly not the highest. Most striking is that they had borrowed \$31,582 on average by graduation. Their earnings were at the low end for their cohort (\$45,927). Their high level of loan pressure was principally driven by their high level of loan debt leading to a loan pressure equivalent to 83.5% of their annual income.

The two next highest loan pressure groups (Leaves 7 and 8) reflect the same family background; they also came from lower-educated families and graduated from for-profit and non-profit private colleges. Juxtaposing their college pathways reveals a mechanism that explains some of the loan pressure variation. Students clustered in Leaf 8 display a GPA of 3.18 or higher and majored in education, humanities, and other applied fields. Interestingly, compared to Leaf 9, they had borrowed less, and even earned a bit less, leading to a substantially lower loan pressure ratio of .737. However, the only difference between Leaf 8 and Leaf 7 is that the latter consists of usually better-paying majors: computer science, engineering, business, health, and social sciences. Their loan pressure is much lower (.512) and is to a large extent the result of much higher earnings (more than \$57,164).

Summarizing this CART analysis of loan pressure, we see that combinations of several factors result in different loan pressures: type of institution, parental background, type of major, and length of time in college. No single factor is determinative because these factors can offset one another in some circumstances or exacerbate one another in others. For example, graduates who had affluent and well-educated parents and graduated from expensive colleges (Leaf 2) had relatively low loan pressure. Conversely, graduates who attended private and for-profit colleges and who came from less-educated families (presumably with fewer resources) and who majored in less lucrative subjects constituted the group with highest loan pressure (Leaf 9).

It is the particular combination of family background, type of institution, time-to-degree, and major that is associated with the amount of loan pressure. The most problematic combination – the highest loan pressure – is associated with undergraduates from less affluent and less educated families who attend expensive colleges. They accumulate high levels of loan debt relative to modest post-graduation earnings.

### **Parental Background and Loan Debt**

How can the positive bivariate relationship between parental background and loan amount be explained? In order to unpack this effect, we examined the costs of higher education by parental background group.

Figure 2 plots the parental background group against the cost-adjusted predicted loan amount (among dependent students). Focusing on the percentages above each parental income bar, these represent the share

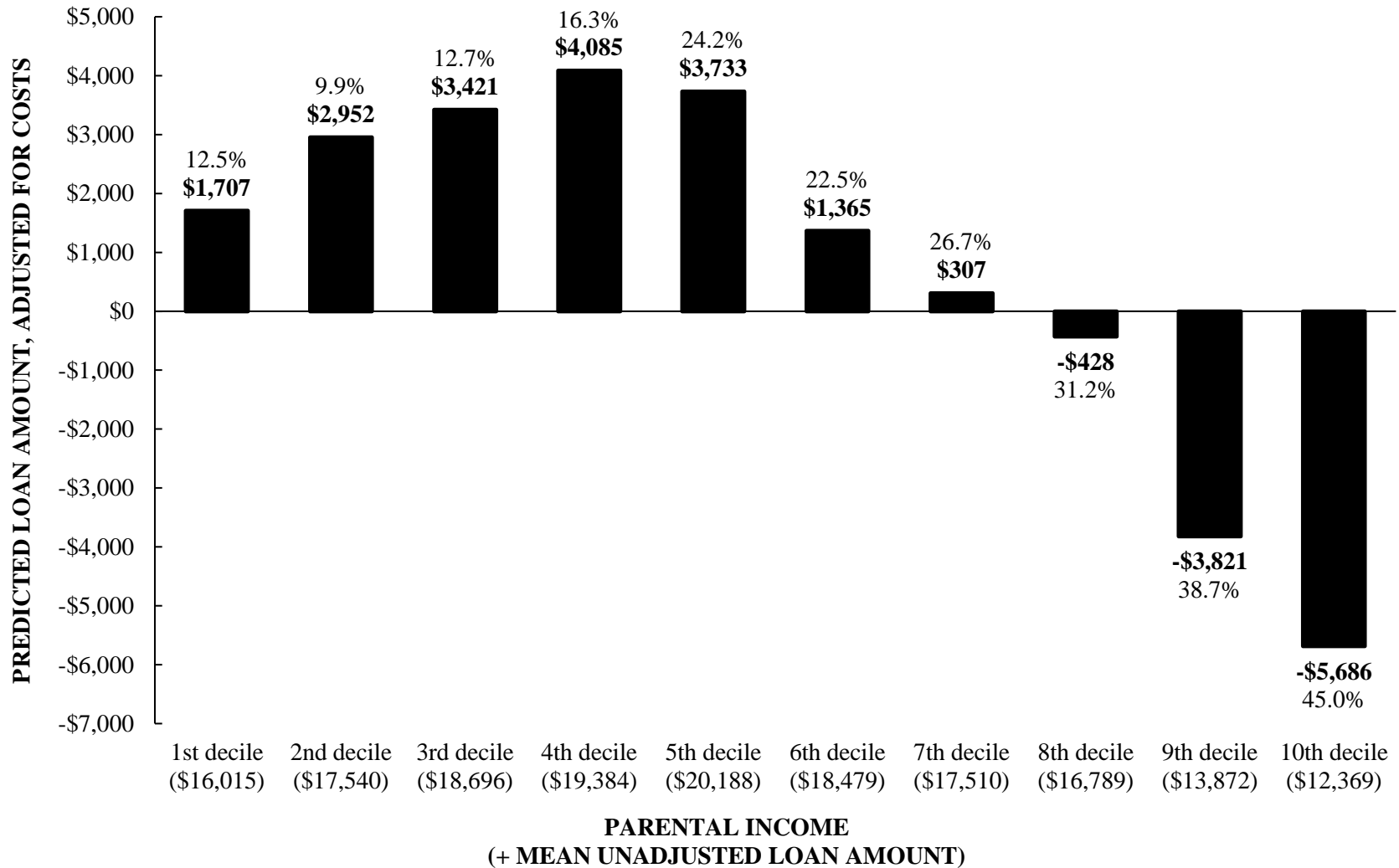
of students that attended 'high out-of-pocket' colleges (defined as the 75<sup>th</sup> percentile or higher; \$10,660). The annual out-of-pocket cost (tuition minus aid and scholarships) represents the amount that students and their families require to attend college. As parental income increases, the percent of students experiencing high out-of-pocket costs goes up. For instance, only 12.5 percent of lowest-income students experience high out-of-pocket costs of their college attendance, while about 45 percent of students from the most affluent families are faced with a relatively high level of college expenses. In other words, this part of the association between background and loan debt is both linear *and* positive; higher-SES students do in fact attend colleges that lead to higher net expenses.

However, the relationship between parental background is also mediated by the type of college attended. The bars in Figure 2 represent the residuals after regressing the post-college loan debt on the out-of-pocket costs. The graph therefore reflects the deviation from the expected loan amount – i.e. over- or under-borrowing – by family background, as these are adjusted for the net costs of education. For students from lower-income families, the loan debt exceeds the expected amount by several thousands of dollars – up to about \$4,000 more for the 4<sup>th</sup> parental income decile. Conversely, high-family income students, from families in the 9<sup>th</sup> and 10<sup>th</sup> income decile groups, take out \$3,821 to \$5,686 *less* than expected given the higher average costs of the colleges they attend.



Figure 2

*Loan Amount by Parental Income Decile, regression-adjusted for Out-of-Pocket Costs (% indicate share in high cost institutions).*



Source. Baccalaureate and Beyond 2008. N = 7,700 (a rounded figure)

Notes. Calculations include dependent students. “High” out-of-pocket costs defined as the 75<sup>th</sup> percentile (a college loan debt of \$10,660). Sample weights applied

From an intergenerational perspective, college students experience two stratification patterns regarding loan debt that operate in opposite directions. First, higher-income students are much more likely to attend more expensive schools, while on average lower- and middle-income students are attending less expensive schools or slightly more expensive schools with some aid (tuition remission or scholarships). Second, adjusted for the class-rooted pattern of out-of-pocket costs, lower-income students are much *more* likely to take out higher loans than their more affluent classmates and similar institutions. Hence, our analyses indicate that the variation in expenses reverses the association between parental background and loan debt.

## Conclusion and Discussion

Rising undergraduate loan debt undoubtedly reflects, to some extent, increasing opportunities for students who are otherwise not able to attend the college of their choice. However, the current explosion of post-college loan debt poses the question whether a loan burden may also create a social problem.

Our approach is different from many traditional educational and sociological studies. We do not estimate loan-or labor market outcomes based on a series of individual- and institutional predictors. Instead, we focused on the complexities of unequal exposure to loan debt among undergraduates. For the purpose of revealing these complexities, we used CART analysis to find interactions between many demographic factors and college factors, as they relate to our outcomes of interest: loan probability, loan amount, and loan pressure. The benefit of CART is that it maximizes the explanatory variance in the dependent variable(s), which is the preferred analytical strategy in absence of strong directional hypotheses and a research interest in the interactions between a multitude of individual-level risk factors.

We found that the parental education and parental income, often in combination with the type of college attended, generate the strongest predictive model to estimate both loan probability and loan amount at college graduation. In addition to these structural components of selection and sorting in higher education, we found that a high average GPA is associated with clusters that avoid loan-taking and loan amount. Merit-based scholarships may partially explain this as the annual out-of-pocket costs are also much lower for these high achievement groups. We also noted that graduates who attend for-profit institutions carry the highest loan debt, regardless of background or achievement, and despite working 32 hours on average.

Our CART analysis of loan pressure – the ratio of the undergraduate loan amount to current earnings – suggests that attending either private or for-profit colleges is associated with a high loan burden for students from lower-educated families. Within this demographic cluster, the loan pressure for the low-achieving students is about .80, whereas higher-achieving (lower loan debt) and high-paying majors (higher pay) reduce the loan pressure to .50 to .70. In contrast, the lowest loan pressure levels (.25) are found among very highly-performing students at public institutions who get through their program within four years, as well as students from higher-educated families who attended private institutions.

Although our exploratory approach generated some complex sets of interactions, it is remarkable that the strongest predictors are consistent across loan risk outcomes. But perhaps most striking is the fact that many predictors that were part of each model never came up as significant factors. These included gender, race, ethnicity, and age at college entry. (The full list can be reviewed in the methods section).

Since parental background came to the fore as the most consistent predictor of loan risk, we further examined this relationship. We found that there is a linear and positive relationship between parental income and the net cost of higher education. However, after adjusting for these college choices, we found that students from more affluent families borrow much less than their lower- and middle-class counterparts. In other words, lower-income students need to take out more money for several reasons. And if we account for one of these, the net costs of attending college, we still find much higher loan debts among lower-

income students than higher income students. This contrasts with Baum's (2016, 2018) research that suggests that loan amounts are higher among undergraduates from affluent families than from lower-income undergraduates. Her interpretation is correct for the bivariate relationship, but once one controls for cost of institution it is clear that lower income students are most burdened by debt.

### **Implications for Practice**

This research documents that loan debt and loan pressure are higher for low-income students than for high-income students who attend an equivalent institution. Need-based financial aid and merit scholarships typically do not fully cover the costs of college (Snyder, de Brey & Dillow, 2017) and especially exacerbate the loan burden of low-income students. Admissions and financial aid staff face a quandary, especially in colleges with higher tuition and fees, because they seek to increase student diversity, but in doing so they are often putting low-income students in situations that require considerable indebtedness.

At a minimum this places an ethical obligation on colleges to provide accurate and clear information to prospective students as to the likely shortfall of aid, and the likely costs that will have to be borne by the student. Unfortunately, researchers have documented that some colleges considerably misstate the typical costs of attendance, either by understating housing, commuting and educational costs, or by omitting various use fees, health insurance and expenses that students will incur.

Given our findings regarding the social (class) background influence on exposure to loans, loan debts, and loan to earnings ratio, we believe that full and honest disclosure of the costs of attending one's institution is paramount to correctly inform parents and students. Beyond this, every college should realize that its mix of tuition-discounting and need-based aid has uneven consequences for the levels of indebtedness that different kinds of students will face.

## References

- Akers, B., & Chingos, M. (2016). *The game of loans: The rhetoric and reality of student debt*. Princeton: Princeton University Press.
- Alon, S. (2007). The influence of financial aid in leveling group differences in graduating from elite institutions. *Economics of Education Review*, 26, 296–311.
- Baum, S. (2016). *Student debt: Rhetoric and realities of higher education financing*. New York: Palgrave Macmillan.
- Baum, S., & Schwartz, S. (2006). *How much debt is too much? Defining benchmarks for manageable student debt*. New York: College Board. Retrieved from: <https://files.eric.ed.gov/fulltext/ED562688.pdf>
- Baum, S. (2018). *Affluent households owe the most student debt*. Washington DC: Urban Institute. Retrieved from: <https://www.urban.org/urban-wire/affluent-households-owe-most-student-debt>.
- Best, J., & Best, E. (2014). *The student loan mess*. Berkeley: University of California Press.
- Breiman, L., Friedman, J., Olshen, R., Stone, C., Steinberg, D., & Colla, P. (1983). *CART: Classification and regression trees*. Belmont: Wadsworth.
- Britt, S., Ammerman, D., Barrett, S., Jones, S. (2017). Student loans, financial stress, and college student retention. *Journal of Student Financial Aid*, 47(1), 25–37.
- Brown, M., & Caldwell, S. (2013). *Young student loan borrowers retreat from housing and auto markets* (Liberal Street Economics paper). New York: Federal Reserve Bank of New York. Retrieved from <http://libertystreeteconomics.newyorkfed.org/2013/04/young-student-loan-borrowers-retreat-from-housing-and-auto-markets.html>.
- Chakrabarti, R., Gorton, N., & van der Klaauw, W. (2017). *Diplomas to doorsteps: Education, student debt, and homeownership* (Liberal Street Economics paper). New York: Federal Reserve Bank of New York. Retrieved from <http://libertystreeteconomics.newyorkfed.org/2017/04/diplomas-to-doorsteps-education-student-debt-and-homeownership.html>
- DesJardins, S.L., Ahlburg, D.A., & McCall, B.P. (2006). An integrated model of application, admission, enrollment, and financial aid. *The Journal of Higher Education*, 77(3), 381–429.
- Dworkin, E. (2012, March). Will you marry me (after I pay off my student loans)? *Bloomberg Business*. Retrieved from <https://www.bloomberg.com/news/articles/2012-03-28/will-you-marry-me-after-i-pay-off-my-student-loans>.
- Dynarski, S. (2016). *The trouble with student loans? Low earnings, not high debt*. Washington, DC: Brookings Institution. Retrieved from <https://www.brookings.edu/research/the-trouble-with-student-loans-low-earnings-not-high-debt>.

- Dynarski, S., & Kreisman, D. (2013). *Loans for educational opportunity making borrowing work for today's students* (The Hamilton Project, Discussion Paper 2013-05). Washington DC: Brookings Institution.
- Elliott, W., Lewis, M., & Johnson, P. (2014). *Unequal outcomes: Student loan effects on young adults' net worth accumulation*. Lawrence: Assets and Education Initiative.
- Federal Reserve Bank of New York. (2017). *Quarterly report on household debt and credit; November 2017*. New York.
- Goldrick-Rab, S. (2016). *Paying the price: College costs, financial aid, and the betrayal of the American dream*. Chicago: University of Chicago Press.
- Grinstein-Weiss, M., Perantie, D.C., Taylor, S.H., Guo, S., & Raghavan, R. (2016). Racial disparities in education debt burden among low- and moderate-income households. *Children and Youth Services Review*, 65, 166 – 174.
- Hossler, D., Dundar, A., Shapiro, D.T. (2013). Longitudinal pathways to college persistence and completion: Student, institutional, and public policy perspectives. In L.P. Perna & A. Jones (Eds.), *The state of college access and completion* (pp. 140 – 165). New York: Routledge.
- Hossler, D., Ziskin, M., Gross, J., Kin, J., & Cekic, O. (2009). Student aid and its role in encouraging persistence. In J.C. Smart (Ed.), *Higher education: Handbook of theory and research* (pp. 389 – 425). New York: Springer.
- Luong, M. (2010). The financial impact of student loans. *Perspectives on Labour and Income*, 22(1), 29 – 42.
- Snyder, T.D., de Brey, C., and Dillow, S.A. (2019). *Digest of Education Statistics 2017* (NCES 2018-070). National Center for Education Statistics, Institute of Education Sciences. Washington DC: U.S. Department of Education.
- Perna, L.W. (1998). The contribution of financial aid to undergraduate persistence. *Journal of Student Financial Aid*, 28(3), 25 – 40.
- Radwin, D., Conzelmann, J.G., Nunnery, A., Lacy, T.A., Wu, J., Lew, S., Wine, J., and Siegel, P. (2018). *2015–16 National postsecondary student aid study (NPSAS:16): Student financial aid estimates for 2015–16* (NCES 2018-466). Washington, DC: National Center for Education Statistics.
- Scott-Clayton, J. (2018). The looming student loan default crisis is worse than we thought. *Evidence Speaks Reports*, 2(34).
- Velez, E., Woo, J., & Simone, S. (2017). *The debt burden of bachelor's degree recipients* (Stats in brief series NCES 2017-436). Washington DC: National Center for Education Statistics.
- Wei, C., Horn, L., & Soldner, M. (2013). *Federal student loan debt burden of noncompleters* (Stats in brief series NCES 2013-155). Washington DC: National Center for Education

## Appendix A

*Descriptive Statistics of Excluded Cases from Loan Pressure Analysis.*

	<b>frequency</b>	<b>had loan</b>	<b>loan debt</b> (at college graduation)	<b>annual out of pocket costs</b>
not employed in 2012	840	.697	\$17,493	\$5,842
enrolled in education in 2012	2,520	.719	\$17,670	\$5,787
not interviewed in 2012	2,430	.757	\$19,909	\$5,976
included in study sample	8,120	.740	\$17,931	\$5,681

*Source.* Baccalaureate and Beyond 2008 – 2012.

## Appendix B

*Descriptive Statistics of Parental Income Deciles (among dependent students).*

<b>parental income</b>	<b>mean income</b>	<b>modal highest parental education</b>
1st decile	\$9,303	high school / GED
2nd decile	\$25,143	high school / GED
3rd decile	\$36,334	high school / GED
4th decile	\$47,882	bachelor's degree
5th decile	\$60,519	bachelor's degree
6th decile	\$75,791	bachelor's degree
7th decile	\$92,448	graduate degree
8th decile	\$110,439	graduate degree
9th decile	\$137,888	graduate degree
10th decile	\$219,255	graduate degree

*Source.* Baccalaureate and Beyond 2008 – 2012. N = 7,700 (a rounded figure).