When Do Students and Taxpayers See a Return? Optimal Accountability Thresholds in Higher Education

DOUGLAS WEBBER, TEMPLE UNIVERSITY, 12-1-2021

INTRODUCTION

Higher education has come under increased scrutiny in recent years due, in part, to rising costs, a ballooning student debt portfolio at the national level (currently at $1.6 trillion), and countless examples of students struggling with their individual debt burdens. A number of high-profile policy solutions have been proposed to address the oft-mentioned "student debt crisis", including debt forgiveness,[1] free college,[2] and doubling federal need-based funding such as the Pell grant.[3]

Running parallel to these student-centric proposals are reforms aimed at the institutional side of the market. These range from informational improvements, which aim to provide better data on student outcomes at each school/program,[4] to formally sanctioning institutions whose students appear to receive little value-added from their postsecondary investment. This paper seeks to provide guidance to policymakers regarding this latter set of policy proposals.

There are two main rationales for the government to provide oversight and maintain a system of accountability in the market for higher education: safeguarding taxpayer funds, and protecting the time and monetary investment of students. Given that student loan debt cannot be forgiven through bankruptcy (except in limited circumstances) this second rationale is especially critical, as poor outcomes can financially cripple an individual for years or decades.

Policymakers and researchers generally agree that the current accountability system which governs the higher education landscape is inadequate. For example, 709 schools have loan portfolios where less than 25% of students have made any progress paying down the principal balance on their loans three years after leaving school and entering repayment. However, there is a disconnect between many in the policy community over what the actual goals of accountability policies should be. One side tends to focus their rhetoric on the protection of taxpayers' investment in higher education, while others view the chief goal of accountability as a means of consumer (e.g. student) protection.

While both goals are certainly laudable, the disconnect can be problematic. Any policy which maximizes student outcomes likely also includes some degree of government subsidization. Similarly, a policy which maximizes taxpayers' return could harm some share of students through inadequate access to financial capital. Making matters worse, there is no work to my knowledge which attempts to quantify these gains/losses, making it all the more difficult to come to a consensus. This paper attempts to fill that void.

I examine a key decision the federal government faces, how to decide which institutions and programs are granted access to the student loan system. I first discuss the conceptual tradeoffs that exist and what an "optimal" policy looks like, depending on the objective(s) a policymaker is seeking to maximize. I then use data on educational outcomes and lifecycle earnings to provide concrete estimates of optimal policy thresholds under a variety of different assumptions regarding the costs and benefits (and the weights we place on them) of governmental involvement in higher education.

I specifically look at two possible types of accountability thresholds: 1) debt-to-earnings ratios, and 2) an institution's student loan "repayment rate" (using the College Scorecard definition). Both metrics have been proposed as key accountability measures, with debt-to-earnings ratios coming the closest to actual implementation through the Gainful Employment (GE) system which implemented the statutory distinction between most programs offered at public and non-profit institutions from most offered by for-profit institutions (and non-degree programs elsewhere).

I find that the optimal debt-to-earnings threshold to be in the 11-12% range and the optimal repayment rate threshold to be in the 15%-20% range. This is the case regardless of whether the objective is to safeguard taxpayer dollars or to maximize students' financial outcomes. However, the threshold which maximizes government revenues is much more sensitive to individual modeling decisions, and is considerably higher under even small changes in the assumed discount rate or fraction of tax revenue generated from college education. These results can be interpreted as 1) schools with debt-to-earnings or repayment rates on the wrong side of these benchmarks cost the government more money than they generate in loan repayments and taxes, and 2) these schools leave the average student financially worse off for having attended the institution.

It is important to note that every critical modeling decision is "biased" in favor of colleges in these analyses. For example, from the government's perspective there are many other subsidies provided to higher education; the most notable for schools/ programs at risk of falling below these thresholds would be Pell Grants. From the perspective of students, I assume that students enter college at age 18, meaning that their incomes would be measured for accountability purposes during their mid-20's. This produces a conservative threshold because for any given level of earnings, debt would be predicted to be more sustainable for a younger borrower who is more likely to see natural wage growth. Put another way, a 26-year-old making $30,000 is likely to make more over the next twenty years than a 41-year-old also making $30,000. I purposefully choose to make the most conservative (e.g.in favor of schools with very poor outcomes) assumptions possible given the dramatic implications for schools and students if they were to lose access to Title IV funding.

A BRIEF HISTORY OF FEDERAL ACCOUNTABILITY REGIMES FOR HIGHER EDUCATION

The federal government began collecting and publishing institution-level measures of student loan performance in 1989 in response to concerns about persistently high student loan default rates. After a brief phasing-in period, Congress tied an institution's ability to receive student loans to a set of thresholds based on two-year Cohort Default Rates (CDR). This type of CDR reflects the proportion of student who default on their student loans within two years of officially leaving school (either graduating or dropping out). The Higher Education Opportunity Act of 2008 slightly altered the regulations to instead depend on three-year CDR thresholds amid concerns that following students for two years was not long enough to adequately assess the financial health of a loan portfolio. The current definition of "default" within a student loan context requires no payments to be made for 270 consecutive days.

In order to lose Title IV eligibility (the ability to participate in the student loan system), an institution must have a CDR of greater than 30% for three consecutive years, or a CDR of greater than 40% for a single year. Several criticisms have emerged over the use of CDR thresholds in this context. First, default is a worst-case scenario which does not accurately reflect the financial health of a students' ability to repay their loans. A former student could pay $1 toward their loan every six months and stay out of default, but would clearly be in severe financial trouble based upon nearly any other metric.

Second, very few schools have CDR's in the vicinity of the thresholds despite having poor outcomes along many other earnings and loan repayment metrics. Only a handful of (typically small) schools per year are even at risk of CDR-based penalties.[5] Thus, despite many institutions having poor loan-based outcomes, virtually no schools actually feel any pressure from CDR thresholds.

This is due in part to the advent and growth in popularity of Income Driven Repayment (IDR) plans. Under IDR plans, students who earn below certain income levels (typically 150% of the poverty line for a given family size) are not responsible for making any student loan payments, and thus will not be considered as defaulting even if they make no payments for years at a time. Moreover, due to the specific definition of default, it is possible for schools to manipulate their CDR by encouraging

students to enter programs such as forbearance until they are outside of the three-year monitoring window. Indeed, at least 1,300 institutions employ consultants to help "manage" their default rates.[6]

A final concern with CDR-based thresholds is their blunt nature, denying federal financial aid to all students at an institution which barely fails the threshold test while doing nothing to schools which barely pass. While this is a limitation of any threshold-based accountability system, it may be particularly problematic in the case of higher education because of heterogeneity in outcomes across programs within an institution. For example, should nursing students be denied student loans because they attend a school over-populated with degrees which underperform in the labor market? Conversely, programs with extremely poor outcomes can be allowed to continue if they exist at an institution with stronger outcomes in other departments.

One solution proposed in congressional legislation[7] is to replace the CDR with a different measure of a student loan portfolio's financial health, the repayment rate. Among student loans, the College Scorecard defines a repayment rate as the proportion of students who have made at least $1 of progress paying down their principal loan balance. Such a change solves many of the issues with CDR's listed above. Repayment rates do a much better job of illustrating borrowers who are actually struggling with repaying their loans, are significantly less manipulable (unless institutions are willing to pay off students' accrued interest), and are not diluted by the rise of IDR-style programs. Moreover, if thresholds are applied at the program level, rather than at the institution level, the bluntness of the penalty will be far less crucial.

The GE regulatory structure was simultaneously more and less ambitious than CDR-based accountability. In reliance on the statutory definitions, GE was designed to be applied only for programs which are vocational in nature, which covered nearly all for-profit programs in addition to certificates at non-profit and public schools. Political discussions surrounding accountability often focus on the for-profit sector due to the sector having worse student outcomes (e.g. loan default rates, earnings) on average (Cellini, 2012; Cellini and Chaudhary, 2014; Cellini and Turner, 2019); but in reality, the GE regime was more nuanced in its coverage.

The regulations sought to tie Title IV eligibility to a much strong indicator of student/borrower success, the ratio of annual student loan debt payments to annual earnings. Had GE penalties been fully implemented, any students at programs which had mean/median annual debt payments (based on a 10-year repayment schedule) which exceeded 12% of annual earnings would be deemed to fail the GE threshold. For programs with debt-to-earnings ratios in the 8-12% range there would have been no immediate penalties, but they would be subject to longer-term monitoring. The penalty portion of the GE regulations were never implemented due to the cancellation of the rule in by then Secretary of Education Betsy DeVos.

ACADEMIC LITERATURE ON HIGHER EDUCATION ACCOUNTABILITY

There are two chief justifications for having a strong accountability system within higher education: protection of the taxpayers' investment and protection of students from spending their own money and time on low-quality programs. If all programs yielded the same outcomes, or at least the same outcomes conditional on student ability, then there would be little to no need for such safeguards. But the literature on returns to college quality generally finds large disparities across schools (Hoekstra, 2009; Cohodes and Goodman, 2014; Andrews, Li, and Lovenheim, 2016; Scott-Clayton, 2017; Canaan and Mouganie, 2018). A notable exception to this finding comes from Dale and Krueger (2014) which finds no differences in returns on average across a set of relatively selective schools. An excellent reconciliation of this literature can be found in Hoxby (2018) which examines value-added and productivity (return on spending) across most of the higher education landscape. She finds practically no differences among highly selective schools, but significant dispersion in quality among non/less selective schools.

At the state level, however, there has been a surge in the incidence of accountability programs and research evaluating the impacts. The modal program, known as Performance Based Funding (PBF) links state funding to outcomes such as graduation rates and enrollment of students from disadvantaged backgrounds. For an overview of PBF and evaluations of various programs, see recent work by Shin (2010), Hillman, Tandberg, and Gross (2014), and Kelchen and Stedrak (2016). For a complete accounting of all accountability systems faced by U.S. higher education, see the recent book Kelchen (2018).

Given that this report seeks to provide context for policymakers on where to set accountability thresholds, an especially important literature to be aware of is that which examines the consequences for students and institutions of failing the CDR thresholds. Darolia (2013) documents significant declines in enrollment at sanctioned institutions using a regression discontinuity design. Cellini, Darolia, and Turner (2020) further show spillover effects for sanctioned for-profit institutions (positive for local community colleges, negative for local non-sanctioned for-profits) and an overall two percent decline in total local market enrollment in college. The latter effect is most pronounced among localities with few alternative postsecondary options. This highlights a key challenge of any accountability system. The consequences of student debt from low-performing programs can be very damaging to future financial health, with the most pronounced struggles concentrated among minority student populations who often do not have the safety net of parental wealth to protect them from the worst outcomes such as default (Addo, Houle, and Simon, 2016). But at the same time the negative tradeoffs of reduced access, particularly among disadvantaged populations, must also be considered.
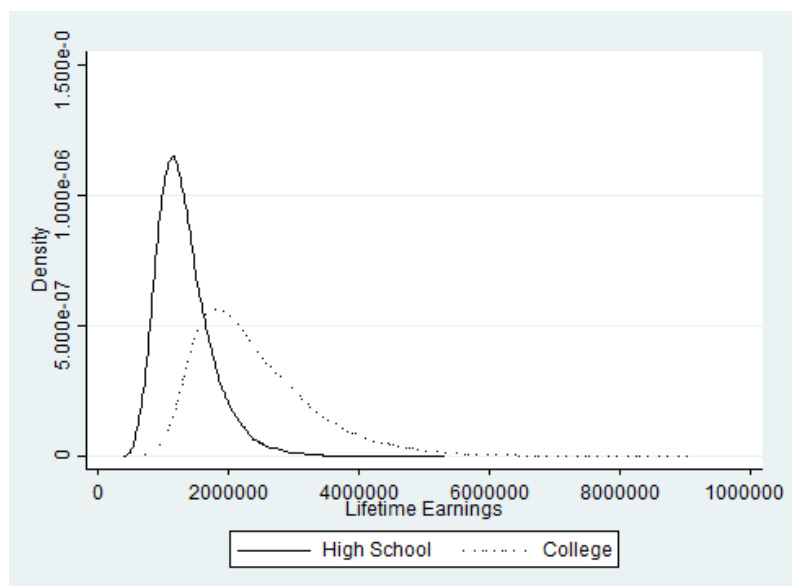
This latter concern, of accountability efforts unintentionally creating education deserts, is an argument in favor of program-based accountability rather than institution-based regulations. Some evidence on this tradeoff is presented in the excellent Brookings report Matsudaira and Turner (2020) which examines virtually all accountability options available to policymakers.

METHODOLOGY

In an ideal (from the researcher's perspective) world, a dataset would exist which follows all workers throughout their lives. If such as dataset were also linked with educational records and student loan repayment outcomes, you could simply summarize lifetime outcomes based on early career debt-to-earnings (or any other metric), and whenever those outcomes are equivalent to your comparison group (e.g. workers who never attended college), you have identified the optimal accountability threshold.

The most challenging aspect of any policy simulation is that such a dataset does not exist, so we must create a synthetic dataset which approximates what the ideal data would look like. I follow the approach from Webber (2014)[8] to construct projections of annual earnings for workers between the ages of 18 and 65. I use five waves (2015-2019) of the American Community Survey to obtain precise estimates of the parameters (e.g. means, variances) of cross-sectional earnings distribution for every age, and I use longitudinal data from the National Longitudinal Survey of Youth (1979 and 1997) to ensure that projections match actual earnings patterns of individuals as closely as possible. Figure 1 shows the projected lifetime earnings distributions for college and high school graduates.

FIGURE 1

Once annual earnings are projected for each year of workers' lives, the task of this report is simplified to an accounting exercise. Individuals are assumed to spend five years in school, graduating at age 23. Despite roughly a third of current undergraduates being over age 25, data on outcomes for these so-called "non-traditional" students is relatively sparse. I thus focus my policy simulations on traditional college students for two reasons. First, this avoids having to make myriad ad hoc assumptions about how age/experience interact with the college wage premium. Second, this assumption produces conservative estimates of optimal repayment rate thresholds due to individuals receiving the college premium over the longest possible time horizon.[9] A similarly conservative estimate is achieved by assuming that all students graduate.

Students are assumed to pay $10,000 per year in out-of-pocket costs to attend college, and to finance the balance of their education with student loan debt. This equates roughly to the situation faced by the typical student attending an in-state public school.[10]

Each simulated student in the analysis accrues $30,000 in loans at the time they leave school, the current national average, and will repay their loans according to the most recent Income Driven Repayment (IDR) policies: 10 percent of discretionary income (the difference between annual income and the federal poverty line for an individual person). Any remaining student loan balance after twenty years is written off by the government. All results below are presented based on this twenty-year time horizon of repayment.

I model student loan repayment under IDR (rather than the standard/fixed repayment plan) because it bases the resulting estimates solely on student labor market outcomes. Interest accrues at the current federal student loan interest rate of 3.73%, and all incomes increase over time at the current long-run projected inflation rate of 2.44%.Taxes and after-tax incomes are calculated based on the tax laws in effect during 2021, and are based on single-filer tax rates.

RESULTS

Tables 1 presents the estimated[11] optimal thresholds from the perspective of the government's student loan portfolio. The second, third, and fourth rows individually add several adjustments to the most basic simulation: administrative costs, taxes, and "selection" (the notion that increased earnings among college graduates may not be entirely due to attending college in a causal sense). Administrative costs have been estimated to be roughly 1.7% of loan volume.[12] Taking these costs into account has only a small impact on the breakeven level from the government's perspective.

Tax revenue has the potential to drastically alter the calculus of the government's student loan investment, assuming that the college degree increased earnings in a causal sense. In other words, even if a borrower doesn't repay their student loan, if they earn more (and thus pay higher taxes) as a result of their college education then taxpayers are still seeing a positive return.

In the results labeled with "selection" I assume that one-third of the college wage premium is due to selection, and adjust the counterfactual scenario accordingly. This is the magnitude of selection estimated in prior work (Webber, 2014) and other similar work. The final row incorporates all of the above adjustments at the same time.

While the results based on nominal dollars are a useful starting point to establish a baseline, any policy must take into account the discounted nature of future payments. This is particularly important given the prevalence of Income Driven Repayment plans and the fact that the wage gains from education accrue throughout the lifecycle (rather than all up front). At the time the government receives the bulk of student loan repayments under IDR, the real value of those payments has been diluted by years of inflation. Moreover, every dollar that the government lends today is one that could be used on some other form of spending, even if it is simply paying down the federal debt.

There is some debate over how significantly to discount future repayments for the purposes of scoring the federal student loan portfolio. This paper discounts future repayments by the interest rate on 30-year Treasury Bonds. This rate supposedly reflects the long-term costs of borrowing that the government faces. In practice, this approach simply looks at repayment in real dollars, as the 30-year Treasury Bond interest rate and long-run inflation expectations are naturally quite close.

A compelling argument can also be made for the "Fair Value Accounting" methodology, which takes into account market risk/

return when calculating a discount factor. For example, there is uncertainty about how much students will repay, and there are other investments the government could make which may have a greater return. As Fair Value Accounting necessarily discounts future repayments to a much more significant degree, my method of discounting will again lead to a lower bound on the estimated thresholds. See Delisle and Richwine (2014) for a detailed discussion of Fair Value Accounting in a student loan context.

The large impact that discounting has on whether a given school contributes a net loss to the federal student loan portfolio can easily be seen in Table 1. The baseline optimal threshold, which compares dollars repaid by students to dollars lent by the government, is 43.6%. This implies that nearly half of institutions (though comprising only about a third of students), and effectively the entire for-profit sector, act as a net drain on the federal budget. This finding is in line with recent Congressional Budget Office projections of the undergraduate portion of the student loan portfolio.

The protection of taxpayers' investment is certainly an important consideration for policymakers, but it ignores much of the rationale often used to justify a significant government presence in the market for higher education. Table 2 thus examines the college investment decision purely from the point of view of the individual borrower, and whether their explicit and implicit (opportunity cost of time) investment is likely to pay off.

Table 2 presents optimal thresholds under a variety of specifications, including accounting for taxes, selection, and an increased emphasis on lower income individuals (e.g. the specifications which use the natural log transformation).

The thresholds in this table should be interpreted as the point at which college graduates are equally well-off when compared to the average high school graduate. The same assumptions from the above analyses, namely those which produce conservative estimates, are maintained here. Note that in addition to my attempt to make assumptions which estimate a conservative bound on the optimal threshold, this comparison to the average high school graduate is itself a fairly low bar to judge the success of college investment.

TABLE 1: Optimal Accountability Thresholds – Taxpayers

	DEBT-TO-EARNINGS		REPAYMENT RATE	
	No Discounting	Discounting	No Discounting	Discounting
Baseline	12.7%	6.1%	14.3%	43.6%
With Admin costs	12.3%	5.6%	15.1%	48.7%
With Taxes	14.1%	12%	10.9%	14.8%
With Selection	12.7%	6.1%	14.3%	43.6%
All	13.5%	11.3%	11.4%	17.0%

TABLE 2: Optimal Accountability Thresholds – Students

	DEBT-TO-EARNINGS		REPAYMENT RATE	
	No Discounting	Discounting	No Discounting	Discounting
Linear Baseline	13.9%	12.7%	12.8%	15.1%
With Taxes	13.7%	12.6%	13.3%	16.2%
With Selection	12.9%	11.8%	15.1%	18.6%
All	12.5%	11.2%	16.1%	19.9%
Log Baseline	13.1%	12.1%	13.9%	16.6%
With Taxes	13.3%	12.3%	15.0%	16.8%
With Selection	12.6%	11.3%	14.3%	16.3%
All	12.7%	11.2%	14.5%	17.2%

Taken together, the above results suggest that debt-to-earnings ratios in the 11%-12% range or repayment rates in the 15%-20% range would be good levels to set accountability thresholds. This is the case regardless of whether you are looking to safeguard taxpayer dollars or maximize students' financial well-being. It should be noted though that while the results for student well-being are stable across all models, the optimal threshold purely from a taxpayer perspective is highly sensitive to modeling choices. Modestly different assumptions regarding the rate of discounting or the causal effect of education on earnings would lead to dramatically more punitive thresholds (e.g. penalizing far more programs/institutions).

This analysis supports the conventional wisdom in the policy community that the vast majority of higher education is adding value to students, but that there is a small contingent of schools/programs which are a drain on both students and taxpayers. For instance, based on the most recently available repayment rate data from the College Scorecard, nearly 80 institutions (enrolling roughly 30,000 students) have below a 15% repayment rate. This increases to more than 300 schools (and 225,000 students) for a 20% repayment rate threshold. My results also suggest that the original GE proposed thresholds from the 2014 GE rule were well targeted.

REFERENCES

Andrews, Rodney J and Li, Jing and Lovenheim, Michael F, "Quantile treatment effects of college quality on earnings", *Journal of Human Resources* 51, 1 (2016), pp. 200–238.

Addo, Fenaba R and Houle, Jason N and Simon, Daniel, "Young, black, and (still) in the red: Parental wealth, race, and student loan debt", *Race and Social Problems* 8, 1 (2016), pp. 64–76.

Canaan, Serena and Mouganie, Pierre, "Returns to education quality for low-skilled students: Evidence from a discontinuity", *Journal of Labor Economics* 36, 2 (2018), pp. 395–436.

Cellini, Stephanie Riegg, "For-profit higher education: An assessment of costs and benefits", *National Tax Journal* 65, 1 (2012), pp. 153.

Cellini, Stephanie Riegg and Chaudhary, Latika, "The labor market returns to a for-profit college education", *Economics of Education Review* 43 (2014), pp. 125–140.

Cellini, Stephanie R and Darolia, Rajeev and Turner, Lesley J, "Where Do Students Go when For-Profit Colleges Lose Federal Aid?", *American Economic Journal: Economic Policy* 12.2 (2020): 46-83.

Cellini, Stephanie Riegg and Turner, Nicholas, "Gainfully employed? Assessing the employment and earnings of for-profit college students using administrative data", *Journal of Human Resources* 54, 2 (2019), pp. 342–370.

Cohodes, Sarah R and Goodman, Joshua S, "Merit aid, college quality, and college completion: Massachusetts' Adams scholarship as an in-kind subsidy", *American Economic Journal: Applied Economics* 6, 4 (2014), pp. 251–85.

Dale, Stacy B and Krueger, Alan B, "Estimating the effects of college characteristics over the career using administrative earnings data", *Journal of human resources* 49, 2 (2014), pp. 323–358.

Darolia, Rajeev, "Integrity versus access? The effect of federal financial aid availability on postsecondary enrollment", *Journal of Public Economics* 106 (2013), pp. 101–114.

Delisle, Jason and Richwine, Jason, "The Case for Fair-Value Accounting", *National Affairs* 21 (2014), pp. 95–111.

Hillman, Nicholas W and Tandberg, David A and Gross, Jacob PK, "Performance funding in higher education: Do financial incentives impact college completions?", *The Journal of Higher Education* 85, 6 (2014), pp. 826–857.

Hoekstra, Mark, "The effect of attending the flagship state university on earnings: A discontinuity-based approach", *The Review of Economics and Statistics* 91, 4 (2009), pp. 717–724.

Hoxby, Caroline M, "The productivity of us postsecondary institutions", in *Productivity in Higher Education* (University of Chicago Press, 2018).

Kelchen, Robert, *Higher education accountability* (JHU Press, 2018).

Kelchen, Robert and Stedrak, Luke J, "Does Performance-Based Funding Affect Colleges' Financial Priorities?", *Journal of Education Finance* 41, 3 (2016), pp. 302–321.

Matsudaira, Jordan D., and Lesley J. Turner. "Towards a framework for accountability for federal financial assistance programs in postsecondary education." (2020).

Scott-Clayton, Judith, "What Accounts for Gaps in Student Loan Default, and What Happens After. Evidence Speaks Reports, Vol 2, # 57.", *Center on Children and Families at Brookings* (2017).

Shin, Jung Cheol, "Impacts of performance-based accountability on institutional performance in the US", *Higher Education* 60, 1 (2010), pp. 47–68.

Webber, Douglas A, "The lifetime earnings premia of different majors: Correcting for selection based on cognitive, noncognitive, and unobserved fac…", *Labour economics* 28 (2014), pp. 14–23.

Webber, Douglas A, "Are college costs worth it? How ability, major, and debt affect the returns to schooling", *Economics of Education Review* 53 (2016), pp. 296–310.

ENDNOTES

1 https://medium.com/@teamwarren/im-calling-for-something-truly-transformational-universal-free-public-college-and-cancellation-of-a246cd0f910f

2 https://berniesanders.com/issues/free-college-cancel-debt/

3 https://www.thirdway.org/memo/why-we-should-double-the-pell-grant

4 https://www.insidehighered.com/news/2019/05/22/trump-administration-releases-new-program-level-loan-data

5 https://www.thirdway.org/memo/hea-replacement-for-cohort-default-rate-leaves-more-questions-than-answers

6 https://www.washingtonpost.com/news/grade-point/wp/2018/04/26/colleges-are-using-consultants-to-manipulate-student-loan-default-rates-gao-says

7 https://www.shaheen.senate.gov/imo/media/doc/Student%20Protection%20and%20Success%20Act%20of%202019.pdf

8 See either Webber (2014) or Webber (2016) for an extensive discussion of the methodology.

9 If I instead assume that students enter college at age 28, keeping all other parameters constant, the optimal threshold estimates are generally more sensitive to modelling assumptions, and typically in the 6%-9% range for the debt-to-earnings ratio.

10 https://trends.collegeboard.org/college-pricing/figures-tables/average-net-price-over-time-full-time-students-public-four-year-institution.

11 I choose not to present standard errors because I believe their inclusion would convey a false sense of precision in the results. As in any policy simulation of this type, I could artificially obtain standard errors arbitrarily small simply by simulating a larger population. For example, with roughly 65,000 simulated lifecycles, the standard errors on the typical threshold range from 0.1% (debt-to-earnings) to 0.5% (repayment rate) when the simulation procedure is bootstrapped. The principal source of uncertainty that readers should be worried about, particularly in Table 1 is the set of assumptions used to generate the thresholds, rather than any sort of sampling variation/prediction error. I would be happy to provide results under any set of alternative assumptions that is of interest to readers.

12 https://www.newamerica.org/education-policy/edcentral/federal-student-loans-administrative-costs