Over-optimism About Graduation and College Financial Aid*

Emily G. Moschini[†] Gajendran Raveendranathan[‡] Ming Xu [§]

First version: January 5, 2022
This version: May 20, 2022

Abstract

Student loans are a common form of consumer credit in the United States, yet many student debtors have not completed a college degree. We establish empirically that this outcome is not foreseen by most recent college enrollees; such over-optimism about the probability of college graduation may lead to over-borrowing. We incorporate such over-optimism into an overlapping generations model, which also includes family transfers and a private student loan market. These model attributes are disciplined with individual-level panel microdata from the Bureau of Labor Statistics and the US Department of Education. We examine the effects of eliminating over-optismism and relaxing federal borrowing limits in this framework, both across steady states and in the transition. We find that eliminating over-optimism reduces welfare in the long run due to equilibrium adjustments of labor taxes and family transfers, despite providing a boost for low-skill students in the short run. By contrast, expanding access to federal student loans yields welfare gains for high-skill students, while reducing welfare for those with low skill—especially low-skill students from poor families.

JEL classification numbers: I22, I26, D83
Keywords: Post-secondary education, Student loans, Over-optimism, Student loan limits.

*We are grateful to Sergio Salgado for his help with the Survey of Consumer Finances, as well as Mons Chan, Eugenia Gonzalez, Dirk Krueger, Hyunju Lee, Igor Livshits, Sergio Ocampo, Radek Paluzinski, and Thomas Phelan for their comments.

[†]College of William & Mary. E-mail: egmoschini@wm.edu. Website: https://sites.google.com/view/emilymoschini/.
[‡]McMaster University. E-mail: raveeg1@mcmaster.ca. Website: https://sites.google.com/site/gajendranraveendranathan/home.
[§]Queen's University. E-mail: ming.xu@queensu.ca. Website: https://www.mingecon.com/.

1 Introduction

Student loans are now the largest category of consumer credit after mortgages, with outstanding balances amounting to 7.6 percent of GDP.[1] Many student debtors have not completed a college degree: in recent cohorts, around 30 percent of students who begin a 4-year bachelor's degree do not complete a their degree by age 30.[2] This outcome is not foreseen by most recent college enrollees: students' expected probability of graduating is systematically higher than the realized rate for their skill level, especially for those with low skill.[3] Consequently, young adults may be over-enrolling in college (and subsequently borrowing too much), due to overly-optimistic beliefs about their likelihood of enjoying the college wage premium later in life. At the same time, while college students have access to federal student loans with fixed interest rates and generous repayment schemes, borrowing from this source is capped by yearly and cumulative borrowing limits (CRS (2019)).[4] Although many students also having access to private student loans, this alternative form of financing can be costly for students to access. Overall, this indicates that many college students may be borrowing-constrained. Additionally, a college education is an investment in human capital with positive externalities, so that society benefit from boosting college attendance by expanding federal student loan borrowing limits.

In this paper, we study the effect of over-optimism about the probability of college completion: first, its role in generating observed patterns of educational attainment in the United States; and, second, the welfare effects of relaxing federal borrowing limits in the presence of these overly optimistic beliefs. Motivated by our empirical analysis, our study is implemented in an overlapping generations model environment which incorporates over-optimism about college graduation probability among high school graduates and their parents. We allow for private loans as a more expensive alternative to public loans as a means to pay for college, and for family transfers to young adults to adjust in response to changes in in their environment. We also incorporate heterogeneity in skill endowments—which, as in the data, are correlated with family background—and in family income. This heterogeneity allows us to examine heterogeneity in the effects of counterfactual policies along margins very much of interest to policy makers: skill (merit) and income (means). Our policy analysis incorporates the transition path from a baseline equilibrium to the new steady state, so that we provide insight into both short- and long-run policy effects.

Our model is disciplined with novel empirical evidence from two main data sources: the 1997 National Longitudinal Survey of Youth (NLSY97) and the High School Longitudinal Study of 2009 (HSLS:09). Both of these studies are nationally representative, individual-level panel surveys. In the NLSY97, which is a public data set, we can observe student expectations of the probability with which they will earn a college degree by age 30. Because the panel dimension of the NLSY97 is long enough for us to check whether or not students do indeed earn a bachelor's degree by age 30, we then construct the rate of college comple-

[1]Source: Federal Reserve Board G19, Bureau of Economic Analysis, authors' calculations.

[2]Author's calculations, 1997 National Longitudinal Survey of Youth.

[3]Author's calculations, 1997 National Longitudinal Survey of Youth.

[4]Under current US policy, the federal student loan limit for dependent students is sufficient to pay for 1.49 years of college. Here, college cost is average tuition and fees plus room and board, where net tuition and fees refers to tuition and fees after netting out all grants and subsidies (NCES (2019)). Author's calculations.

tion—and compare it to the average expected probability of completion—by skill quantile for the pool of college enrollees.[5] This exercise establishes that over-optimism about college graduation is widespread and especially severe among students in the lowest quantile of skill.[6] In fact, parents of high school graduates exhibit similar patterns of over-optimisml.[7] We use this information to discipline model primitives related to beliefs. In the HSLS:09, which is restricted-use data, we observe uptake of both public and private student loans. We use this information to discipline model primitives related to the student loan market, for both federal and private student loans.

The model is calibrated to match a set of moments related to college enrollment, college graduation, student loan uptake and repayment, and public and private transfers. With the calibrated model in hand, we perform two policy experiments. First, we shut off over-optimism in the economy, and document the effects on college enrollment, output, and welfare. This exercise can be viewed as a policy intervention that provides information about true college graduation probabilities instantaneously to everyone (e.g., high school graduates and parents). Second, we re-institute baseline over-optimism, and then raise the borrowing limit on federal student loans so that they can be used to pay for all four years of college expenses. In both exercises, we examine the welfare gains over the transition path from the baseline steady state equilibrium to the new equilibrium.

In the first policy experiment, we find that shutting off over-optimism actually reduces welfare for the average 18-year-old in the long run. In the initial period of the transition, those with low skill (who had the largest gap between expectations and reality) benefit from the information intervention because over-enrollment goes to zero. However, those with higher skill immediately are worse off as parents reduce transfers due to the the direct effect of the information intervention. In the long run, transfers go down even further, as the composition of the population changes to reflect lower educational attainment. Eventually, the reduction in college enrollment leads to a contraction of the labor income tax base, causing the labor income tax to rise, which negatively affects all consumers. This policy experiment uncovers the complex role of over-optimism in the economy: although over-optimism leads to 18-year-old students to over-enroll in college, it also benefits them in two ways. First, as long as parents are over-optimistic (as our empirical work indicates they are), this raises the inter-vivos transfers that 18-year-olds receive from their parents. Second, over-optimism raises the college education rate and lowers labor income taxes, thus benefiting 18-year-olds later via general-equilibrium effects.

In the second policy experiment, we find that the average 18-year-old gains from an expansion in federal student loan borrowing limits. However, welfare changes are heterogeneous: students in the bottom skill quantile experience welfare losses, while students in the highest skill quantile experience welfare gains. Low-skill students are hurt by limit expansions because the extent of their over-enrollment worsens; these students tend to come from poor families, so that access to more federal loans makes a big difference in their

[5]We measure skill using high school GPA.

[6]This modelling choice is also supported by evidence from the literature on the college dropout decision. For example, using a panel study of students in Berea College, Stinebrickner and Stinebrickner (2012) find that the expected graduation rate was 86 percent, whereas the actual graduation rate was 65 percent. These findings are based on the Berea Panel Study, which is an ongoing survey that collects data on cohorts that entered Berea College in 2001 and 2002.

[7]We also provide supportive evidence for over-optimism among students and their parents from the HSLS:09.

ability to finance college enrollment (but does not change their likelihood of graduating once enrolled). Poor families also raise children with high skill, however, for whom the consequences of over-optimism are less severe. Overall, in the long run an expansion in federal borrowing limits increases the college graduation rate. This results in a larger tax base, and hence, a lower income tax rate, which benefits all consumers.

This study is not the first to examine college enrollment and college financial aid policies. Indeed, previous related studies—which include Lochner and Monge-Naranjo (2011), (Chatterjee and Ionescu, 2012), Krueger and Ludwig (2016), Ionescu and Simpson (2016), (Luo and Mongey, 2019), and Abbott et al. (2019)—also examine dropout risk, the role of private loans, and family transfers.[8] Our study builds on this literature, and additionally includes over-optimism and as well as evidence on the role of private student loans, which we use to discipline that market in our model economy. In particular, we incorporate the general equilibrium life cycle framework of Krueger and Ludwig (2016) by augmenting their model with student loan repayment decisions (Luo and Mongey, 2019), endogenous and exogenous dropouts (Chatterjee and Ionescu, 2012).

A key assumption maintained in previous structural studies of the federal student loan program is that student expectations about academic outcomes are consistent with realized outcomes. We deviate from these papers by introducing over-optimism about college graduation. Our study of expanding the federal student loan limit is directly comparable to policy exercises in these papers that increase college financial aid either by expanding the federal student loan limit or by increasing public grants. In those frameworks, more financial aid raises welfare. With our empirically motivated model innovation, we demonstrate that welfare may actually fall for some students due to such a policy. Our empirical evidence on over-optimism complements previous work demonstrating overconfidence among college enrollees about graduation, in particular Stinebrickner and Stinebrickner (2012). This study uses a small panel at a single college to find that students with low high school GPA are especially overly optimistic about their likelihood of college graduation. Using the NLSY97, we provide new evidence that overconfidence about educational attainment is common and is related to student skill. We then use the same panel data to discipline the model's positive relationship between student skill (high school GPA), college enrollment, and college graduation likelihood.

The role of private student loans as an alternative source of college financing has been emphasized in previous work. As argued by Lochner and Monge-Naranjo (2011), incorporating private student loans into a study of college financial aid is important because private student loans can be imperfect substitutes for federal student loans and grants. However, while the current literature has rigorously incorporated key features of the federal student loan program into their model economies, there is no consensus about the modeling of the private student loan market. For example, Lochner and Monge-Naranjo (2011) assume that lenders price repayment risk based on student skill, and therefore, low skill students are less likely to have access to private student loans. Ionescu and Simpson (2016) assume a pooling equilibrium in which lenders price the student loan based on the credit risk type of the borrower. Abbott et al. (2019) assume that students from low income families do not have access to private student loans. We are able to capitalize on information in the HSLS:09 to document stylized facts about the private student loan market which are reflected in our

[8] Additional studies in this literature are Ionescu (2009), Ionescu (2011), and Caucutt and Lochner (2020).

model framework. These attributes include the fact that private loans are rarely the only source of student debt, which indicates that students take out federal loans before turning to private loans, and that private loan uptake rates are above zero for students from poor families as well as students with low skill. In particular, we assume a pooling equilibrium for the private student loan market, where there is no discrimination in access by family income or student skill. To capture the pecking order of loan types, we introduce a psychic cost for private loans, which represents costs resulting from predatory lending and hidden fees. To our knowledge, we are the first in this structural literature to leverage data to guide modeling choices related to the private student loan market.

Finally, we contribute to the consumer credit literature on over-borrowing. Recently, several policies have been implemented in the market for consumer credit cards that are aimed at reducing access. For example, the Credit CARD Act of 2009 imposed several restrictions in the United States. In addition, in Canada, Quebec has recently imposed restrictions on borrowing. Some papers have quantitatively analyzed over-borrowing in the market for consumer credit cards. For example, Exler, Livshits, MacGee, and Tertilt (2021) analyze implications of policies aimed correcting for over-borrowing in the unsecured credit market due to over-optimism about earnings, while Nakajima (2012) studies the impact of increased access to unsecured credit for consumers with hyperbolic discounting preferences. Nakajima (2017) studies how bankruptcy policy reforms in the unsecured credit market impact consumers with hyperbolic discounting preferences. One of the key takeaways from these studies is that quantity restrictions even in the presence of over-borrowing lead to overall welfare losses. We focus on student loans rather than credit cards, and find that expanding borrowing limits (reducing quantity restrictions) leads to welfare gains on average, but causes welfare losses for low skill students—especially those from low-income families.

This paper proceeds as follows. Section 2 overviews our main empirical findings. Section 3 lays out the model, section 4 describes the model parameterization, and section 5 analyzes the main proporties of the model equilibrium. Section 6 reports the results of our main policy experiments. Section 7 concludes.

2 Data

The two main data sets we draw on are the National Longitudinal Survey of Youth 1997, the High School Longitudinal Study of 2009. We supplement these findings with additional insight gained from the 2019 Survey of Consumer Finances (SCF). The NLSY97 follows a nationally representative cohort of young adults in the United States, from 1997 until 2019, and provides information on expected probabilities of college graduation as well as realized education outcomes by high school GPA. We use this information to show that, conditional on enrolling in a 4-year college degree, students tend to overestimate their likelihood of earning a BA by age 30. To document how financial aid interacts with skill and persistence outcomes, we turn to the HSLS:09. The HSLS:09 follows a nationally representative sample of 9th-grade high school students in the United States from 2009 until 2016, and provides information on the extensive and intensive margins of both federal and private student loan uptake for a later cohort than the NLSY97, which interacted with more recent iterations of US financial aid policy (in particular, borrowing limits).[9] However, because the panel dimension of the HSLS:09 does not allow us to observe repayment outcomes for student debtors, we supplement the student debt information from the HSLS:09 with the 2019 wave of the 2019 SCF. The SCF allows us to observe how loan repayment delinquency and the distribution of student loan balances are associated with college completion in the US population in recent years. We use the SCF to demonstrate that a sizable fraction of student debtors did not complete their education, and that these "dropouts" are over-represented among student debtors who are delinquent on their loan payments.

Taken together, these empirical findings indicate that low-skill students who enroll in college tend to drop out more than their higher-skilled peers, and that students incorrectly believe that they are unlikely to drop out. In addition, students who do not persist in college have significant extensive and intensive student loan uptake patterns, similar to those who persist with their education, and therefore must finance payments on that debt without enjoying the college wage premium. Indeed, our findings indicate that college dropouts tend to have trouble repaying their student loans later in life. These empirical findings motivate our study of college financial aid programs in a model environment that incorporates unanticipated college dropout.

2.1 Expectations versus educational attainment outcomes in the NLSY97

The NLSY97 asks sample members about their expected probability of earning a BA by age 30 twice, in 1997 (round 1) and in 2001 (round 5).[10] The phrasing of this question varies slightly across questionnaires in 1997 and 2001, but can be paraphrased as: "What is the percent chance that you will have a four-year college degree by the time you turn 30?" The response of the student is a percent value between 0 and 100. To construct student expectations about college completion, we use the most recent student response to this question collected while the student was still in high school, in either 1997 or 2001.[11] For each individual with a valid response to the expectations question while in high school, we then flag those who by age 30 had enrolled in a bachelor's degree, and separately flag those who had earned a bachelor's degree (BA).[12] Next, we assign each individual to a skill quantile using the distribution of high school GPA among those who eventually graduate from high school. We compare outcomes with expectations about educational attainment by skill quantile in Table 1. For those who enroll in a 4-year post-secondary (bachelor's) degree

[9] We also use the HSLS:09 to supplement the NLSY97 overconfidence finding by capitalizing on HSLS:09 questions about educational attainment expectations which asked when students are in the spring of their junior year of high school. These findings are presented in Appendix C. For details on this question in the HSLS:09 survey, see Table 30 and the associated discussion in the Appendix. The phrasing of the question in the HSLS:09 is not as straightforwardly refutable as the similar question in the NLSY97 because in the HSLS:09 there is no age limit condition on the outcome being asked about (i.e., "by age 30"), and because the response in the HSLS:09 has to be categorical while the NLSY97 is required to be a probability of the queried outcome occurring. Thus, our main results on overconfidence are established with the NLSY97.

[10] The survey also asks parents once, in 1997. For a comparison of student responses as defined here with parent responses, see Figure 7 in Appendix I. Compared with their children, parents exhibit a similar overconfidence about the likelihood of college attainment in the NLSY97.

[11] In the 2001 questionnaire, respondents were divided into 4 groups for the beliefs questions. Only groups 1 and 3 were asked about educational attainment expectations.

[12] To construct these flags, we use the created variable "CV_ENROLLSTAT". To identify those who enroll, we flag those who are either enrolled in a BA or a higher level of education than a BA or who are not enrolled in school and have at least a BA. To flag those who have earned a BA, we created a new flag that is set to 0 for observations who are currently enrolled in a BA, and equal to the first flag otherwise. For years 1998-2004 (rounds 2-8) the edited version of this created variable is used.

sometime before turning 30, this Table tabulates the frequency of BA attainment by age 30 for each of these high school GPA quantiles (to measure the outcome), and compares it with the average expected probability of earning a BA by age 30 for that skill quantile. The tabulation is done for the sample of students who enroll in a bachelor's degree because, in order for overconfidence to have an effect on financial aid uptake, the overconfidence must be present among those who enroll in college.[13] Table 1 indicates that the expected probability of earning a 4-year degree by age 30 is much higher than the realized rate of attaining that outcome within a given skill quantile. This is especially true for those with the lowest skill.

Table 1: Expectations versus outcomes: BA attainment by age 30

Skill quantile	Obs(Q)	Outcome	Expected probability
1	227	32.60	80.90
2	400	56.75	87.24
3	565	79.82	93.44
Obs	1192		

Notes: Table 1 shows the graduation rate and expected probability of graduation by skill quantile, among students who enroll in a 4-year degree by age 30. Sample: 4-year college enrollees. Source: NLSY97.

Does over-optimism persist from when data is collected during high school until the point of college enrollment? We are argue that it does, and offer suggestive evidence in support of this claim by restricting attention to a group for whom we can measure overconfidence on either side of the college enrollment decision. To do this, we restricting attention to students who answer the 1997 question while still in high school, and answer the 2001 question while enrolled in a 4-year college degree. Results are shown in Table 2; note that here we do not break down the statistics by skill quantile due to the small sample size.[14] The responses given in high school and while enrolled in college indicate that students maintain beliefs about graduation likelihood after they enroll in college at similar levels to their beliefs in high school. If they were learning, the expected probability after enrolling would be closer to the realized probability of graduating, which is about 70 percent. Instead, the expected probability of graduating actually increases slightly.

Table 2: Learning and expectations versus outcomes: BA attainment by age 30

Outcome	Expected probability	
% BA by age 30	**In high school**	**In college**
69.62	92.07	93.14
Obs	316	

Notes: Table 2 compares the realized graduation rate with the expected probability of graduation reported in high school and while enrolled in college. Sample: respondents enrolled in high school in 1997 and were enrolled in a 4-year bachelor's degree program in 2001. Source: NLSY97.

[13] However, skill quantiles are assigned using the distribution of high school GPA among high school graduates, because the distribution of skill among college enrollees incorporates selection by expectations.

[14] The sample size is small for several reasons: a small proportion of respondents meet the education timing criteria and the design of the round 5 questionnaire only asks the relevant question for 2 of the 4 sub-groups surveyed.

These findings from the NLSY97 indicate that overconfidence about the likelihood of earning a bachelor's degree is widespread among those who enroll in college, especially among those with low skill. Next, we turn to the HSLS, in order to document how uptake of student loans and receipt of grant aid vary by student skill and college persistence status.

2.2 Student aid and private loan uptake in the HSLS:09

The HSLS:09 is a representative panel survey of 9th-grade (high school) students in the United States who attend schools that also contain 11th grades. The panel begins in 2009 and follows focal students until 2016, although some information on post-secondary transcripts and student records is collected for academic years after 2016 (HSLS (2009)).[15] This survey contains information on the focal ninth-grade high school student (e.g., their high school honors-weighted GPA after high school graduation) and on their family (e.g., household income and parental education). We use this information to make two points: first, that there is sizable student loan uptake and grant receipt among students who do not complete college; and, second, that private loans are turned to frequently by many kinds of college students, but it is quite rare for private loans to be the only type of education loan a student has. Note that all tabulations with HSLS:09 data, both here and in the Appendix, make use of survey weights as describedin table captions.

For the analysis presented in this section, we restrict our sample to students who have graduated from high school by the summer of 2013 and enrolled in a 4-year bachelor's degree program in the Fall of 2013. Among this group, we additional restrict attention to observations for whom we observe household income, state of residence, parental educational attainment, and the student's race and ethnicity at least once while they are in high school, and for whom we the student's honors-weighted high school GPA is reported in the 2013 Update. We also require that the student reports their educational attainment expectations in their junior year of high school. In what follows, we consider how persistence in college status is related to several other variables. By "persisting", we mean maintaining enrollment in a 4-year program for each academic year after their first year (the 2013-2014 academic year) until their fourth year (the 2016-2017 academic year). Someone who does not persist leaves college for at least one academic year after enrolling. Unlike the NLSY97, the short panel dimension of the HSLS:09 prevents us from using more long-term measures of college completion.

Table 3 reports persistence rates by skill quantile in the HSLS:09. Skill quantiles are assigned using the distribution of high school GPA for high school graduates that meet the criteria described above, except that skill quantiles are assigned using the sample before conditioning on college enrollment.[16] Table 3 shows that, for college enrollees, persistence is more common for students in higher skill quantiles.[17]

[15] See HSLS (2020) for a description of the most comprehensive version of this data set. Table 28 in Appendix C and the surrounding discussion contain further information on the structure of the HSLS:09.

[16] To compare these two samples, before and after conditioning on college enrollment in the 2013-2014 academic year, see summary statistics reported in sections C.2.1 and C.5 of the Appendix.

[17] To see how outcomes compare with expectations about college persistence in the HSLS:09, see Table 31 in the Appendix and the surrounding discussion. Patterns of overconfidence by skill quantile in the HSLS:09 are very similar to the analogous NLSY97 results in Table 1. However, since the survey question asks about ever completing college (presumably at some point in the student's life), the data counterpart to precisely check the realized outcomes against the student's expectation is impossible to construct in

Table 3: Persistence in a 4-year BA program

Skill quantile	**Obs(Q)**	**% Persisted**
1	190	39
2	800	56
3	1680	79
Obs	2670	

Notes: Table 3 reports the percent of college enrollees in a given skill quantile who have not left as of the 2016-2017 academic year (their fourth year of college). Sample: students who enrolled in a 4-year program in the Fall of 2013. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Weights are PETS-SR student records longitudinal weights.

Next, in Table 4, we examine the relationship between student loan uptake and persistence status. Over 30 percent of the enrolled population fail to persist towards college completion. Compared to students who do persist, students who do not persist are also more likely to have student debt and owe almost one in four dollars of outstanding student debt in this sample (either federal or private). The share of dollars owed by 2013 enrollees who do not persist in college is lower than the population share of that group, because their average balance is lower than the average balance of students who persist in college. This is likely because students who did not persist paid for fewer years of tuition. Nevertheless, the student loan balance of the average debtor is sizable three years after enrollment, regardless of persistence status.

Table 4: Student loan incidence by persistence status

	% of 2013 enrollees	**% with SL**	**% of SL $**	**Average balance**
Did not persist	32	74	24	17,051
Persisted	68	65	76	24,175
Obs	2670			

Notes: Table 4 divides the pool of 4-year program enrollees into students who persisted in college and those who did not persist. Persistence status is assigned based on their student record indicates that they were enrolled for each academic year between the 2013-14 academic year and the 2016-17 academic year. Within each persistence outcome group, the table reports the group's share of the total population, the percent of the group with a positive student debt balance,the share of aggregated student debt balances that dollars owed by the group represent, and the average student loan balance owed by debtors in the group in by the end of their third year in 2016 dollars. Percentages are rounded to the nearest percentage point. Sample: students who enrolled in a 4-year program in the Fall of 2013. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to the nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

Grant aid, which is another avenue by which society can subsidize post secondary education, is received to a similar extent by students who persist in college and those who do not. This is illustrated in Table 5, which reports grant aid incidence and share of aggregate grant transfers by persistence status. Students who do not persist and those who do receive grant aid at the same rate. However, the share of aggregate grant transfers made to students who do not persist in college is almost one in four. Like student loan balances, the average grant aid received is slightly lower among non-persisters who got this type of aid, because they received it

the HSLS:09. One would need to observe the survey respondent until the end of their life.

for fewer years that students who persisted in college.[18]

Table 5: Grant aid incidence by persistence status

	% received grants	% of grant $	Average $ amount
Did not persist	76	23	9,573
Persisted	75	77	12,630
Obs	2670		

Notes: Table 5 reports the rate of grant aid receipt, share of aggregate grant transfers, and average grant aid for grant recipients, by persistence status. Percentages are rounded to the nearest percentage point and dollars are 2016 dollars. Sample: students who enrolled in a 4-year program in the Fall of 2013. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to the nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

Now that we have given an overview of student loan uptake and grants across the two persistence categories, we focus on private loans in order to better understand their role. Table 6 reports the sources of student loans in the sample of 2013 enrollees by source of loan. This table has two main takeaways, which hold for both students who persist and those who do not persist: first, that more than 1 in 5 students takes out a private student loan during college; and, second, that students tend to take out a federal loan and then sometimes turn to private loans (i.e., there is a "pecking order" for loan types). For intuition about the second claim, note that the share (in percentage points) of all students with only private loans is quite low, compared to the share with only federal student loans or with both kinds of student loans. If students often took out private loans without also using federal loans, the share of student debtors with only private loans would be higher instead of being almost zero.

Table 6: Student loan portfolio composition

	Either	Federal only	Private only	Both
Did not persist	74	48	2	23
Persisted	65	44	1	19
Obs	2670			

Notes: Table 6 reports, by persistence status, the percent of all college enrollees who have either, only federal, only private, or both types of student loans. Percentages are rounded to the nearest percentage point. Sample: students who enrolled in a 4-year program in the Fall of 2013. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to the nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

Access to private loans is possible for the very poor or the very low skilled among enrollees in the 4-year program. This is demonstrated by the extensive margin uptake rates for private loans reported in Table 7, which the fraction of each family income and skill quantile cell that has taken out a private loan three years after they began college. In this table, family income and skill quantiles are assigned using the distribution of high school graduates.

[18]For a breakdown by type of aid, see Tables 52 and 51 in the Appendix.

Table 7: Extensive margin of private loan uptake

		High school GPA		
	Q	**1**	**2**	**3**
	1	23	19	14
Household income	**2**	27	31	26
	3	28	21	18
Obs 2670				

Notes: Table 7 reports the percent of each cell that has a positive private student loan balance by the summer of 2016. Percentages are rounded to the nearest percentage point. Rows are student family income quantiles using parents' income during high school; columns are high school GPA quantiles assigned using the distribution of high school GPA among high school graduates. Sample: students who enrolled in a 4-year program in the Fall of 2013. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Weights are PETS-SR student records longitudinal weights.

To summarize, in the HSLS:09 more than 30 percent of students who enroll in a 4-year degree right after high school fail to persist in college. Those who leave 4-year programs without graduating tend to hold a sizable share of student debt balances owed by their cohort, and to have received substantial grant aid, despite usually using that money to finance fewer years of school than students who persist toward degree completion. Private loans act as a supplement to federal loans, and are accessed by students at non-negligible rates regardless of family income or student skill. It seems likely that students who do not complete college will face poor labor market prospects and potentially have more difficulty repaying their student loan debt compared to college graduates. For a better understanding of repayment patterns among students who do not complete college, we now turn to the Survey of Consumer Finances.

2.3 Student loan balances and repayment delinquency in the SCF

The SCF is a repeated cross-sectional survey of US households which collects information on household balance sheets every three years. We focus on the 2019 wave of the SCF for two reasons: first, because our policy analyses will focus on the pre COVID-19 US economy; and, second, because the 2019 survey contains more detailed information on education loans than previous waves of the SCF.[19]

The SCF records information on up to 6 education loans owned by a household; for each loan, there is a separate set of variables which record loan-specific responses to various questions about that loan. A feature of this structure is that it serves to make the SCF very detailed at the loan level, while a drawback is that it also makes the definition of a "dropout" or "delinquent" household somewhat subjective. In particular, given an education loan portfolio for the household, one must decide how to map from a portfolio where none, one, or all loans may be late in repayment to an indicator for the household being delinquent. Similarly, one must decide how to map from a portfolio where none, one, or all loans may be associated with a household that

[19]"Education loans" is the term used in the SCF codebook. We use this term interchangeably with "student loans". In the SCF, the term "education loans" is more general than the federal student loan program and includes loans from private lenders. In recent waves of the SCF, it has become possible to further distinguish whether the education loan is a federal loan or not via the introduction of new variables: the majority of education loan balances in the SCF are owed to the federal government, but there are still sizable private student debt balances reported.

dropped out of the program the loan was used to pay for to an indicator for the household having dropped out (i.e., not persisted in college).

In Table 8, we report the share of households and the share of education loan dollars by educational attainment status in the 2019 SCF.[20] Panel A reports the shares for all debtors; panel B reports the same statistics for delinquent debtors. Here, delinquent debtors are those households who are not currently making payments for at least one of the loans that they owe payments on.[21]

Panel A of Table 8 shows that, among households with positive student loan balances, a sizable fraction (about 1 in 5) did not complete their education. About one tenth of the dollars owed are owed by dropouts. This is not surprising: as with student debt balances in the HSLS, the quantity of debt owed by dropouts and those currently enrolled or in the grace period is lower than their share of the population of debtors because these two groups have not, on average, used their loans to pay for as many years of education as graduates have. Nevertheless, dropouts are clearly a significant portion of student debtors, and they do hold a sizable fraction of the outstanding federal student debt in the United States.[22]

Panel B of Table 8, meanwhile, shows that dropouts are overrepresented among deliquent households.[23] In part, this reflects the difficulty of making payments on a loan without simultaneously enjoying the wage premium from a post-secondary degree, but it may also reflect selection into college non-completion. As in the population of all households with student debt in Panel A, the share of dollars owed by dropouts is lower than their share of the population, because they paid for fewer years of schooling. Overall, the SCF indicates that not completing college is associated with delinquency in student loan repayment.

Table 8: Education loans in the 2019 Survey of Consumer Finances

Population	**Statistic**	**Group**	**Percent of Group**
All debtors	Share of HH	Dropouts	19.53
		Graduates	65.35
		Enrolled or in grace period	15.74
	Share of $	Dropouts	11.63
		Graduates	74.45
		Enrolled or in grace period	14.81
Delinquent debtors	Share of HH	Dropouts	50.54-57.31
		Graduates	42.69-49.46
	Share of $	Dropouts	29.18-37.13
		Graduates	62.87-70.82

Notes: Table 8 presents statistics from the 2019 SCF on educational loans. The first group of statistics presents information on the composition of all households with a positive balance of debt, and how the aggregate balance of student debt is distributed across different types of households. The second group of statistics presents information for households that are not making payments on their educational loans because they cannot afford to do so (delinquent households).

Having outlined overconfidence in the NLSY97 and significant loan uptake and grant aid receipt among

[20] In all of our analysis of households in the SCF, only loans owed by the respondent or their spouse are considered. See Table 57 in the Appendix D for details on variables used.

[21] Qualitative results for our variables of interest have proven robust to the various choices of definitions for "delinquent" that we considered.

[22] Although not shown here, this is true whether one focuses on federal or private loans. For these figures we have pooled both kinds of loans.

[23] Note that such households are by construction, neither enrolled nor in the grace period (i.e., who are either dropouts or graduates).

dropouts (non-persisters) in the HSLS:09, we next turn to the model environment we build for policy analysis, which incorporates these features of the data. After parameterizing the model, we compare the model's implications with several moments, including these untargeted repayment statistics from the SCF.

3 Model

Our model economy builds on Krueger and Ludwig (2016), Chatterjee and Ionescu (2012), and Luo and Mongey (2019). We enrich Krueger and Ludwig (2016)'s general equilibrium life cycle model with college choice by incorporating over-optimism about graduation likelihood, motivated by our findings in section 2. We also incorporate endogenous and exogenous college dropouts , as in Chatterjee and Ionescu (2012), and key features of the US market for student loans into our model economy. The features of the federal student loan system are largely based on Luo and Mongey (2019); the features of the private student loan market are based on patterns which we use micro data to provide novel evidence for.

3.1 Overview of the model environment

Time is discrete and runs forever; a period is one year. For ease of exposition, we omit time subscripts. Let j denote the age of consumers; consumers start making decisions when they turn 18 ($j = 1$ in the model). At the beginning of $j = 1$, with an exogenous probability q 18-year-olds consumers may choose whether to go to college; with a probability 1-q, they do not have this option. [24] The 18-year-old's college entrance decision will be based on their skill (high school GPA), idiosyncratic productivity, and initial net assets. Skill, e, is an endowment that is drawn once from a distribution which depends on parental education. The skill endowment determines several factors: the consumer's over-optimistic probabilities about college graduation likelihood at the time of enrollment, the true annual probabilities of continuing college education, deterministic earnings productivity, and grants for college from the government and private sources. The idiosyncratic component of earnings, η, is an AR(1) process. The net assets, a, are determined at the start of adulthood by a one-time familial inter-vivos transfer from parents. Thereafter assets are affected by borrowing and savings decisions of the individual, as well as their own transfers to their future children.

When making the college entrance decision, consumers are overly optimistic about their probability of college graduation. At the time of enrollment, consumers with skill e believe they will continue their education each year of college with probability $\hat{p}(e)$. The true annual probability of continuing a college education be given by the function $p_g(j,e)$. The consumers are overly optimistic as long as $\hat{p}(e) > p_g(j,e)$ for all e and j, as is the case in our baseline calibration. Furthermore, a higher $\hat{p}(e) - p_g(j,e)$ implies higher over-optimism. If $\hat{p}(e) = p_g(j,e)$, consumers have the correct beliefs about graduation likelihood.

We allow consumers to learn their true probability of graduation after enrollment, although such learning does not seem to happen quickly if at all in the data (see Table 2). We do this in the interest of keeping our

[24] The q shock process captures reasons for which consumers may not go to college that are not incorporated into our model. An alternative approach is to assume stochastic psychic costs to go to college as in Abbott, Gallipoli, Meghir, and Violante (2019). Our approach can be interpreted as a a nested version of stochastic psychic costs, where with probability $1 - q$, the psychic cost of college is large enough, so that those consumers will not go to college.

policy experiment results as lower bounds for welfare changes. Specifically, after the first year of college, consumers may choose to dropout of college (endogenous dropout) or may dropout of college with an annual probability of $p_d(j, e) = 1 - p_g(j, e)$ (exogenous dropout). The exogenous dropout captures dropouts due to differences in academic ability.

In our model economy, college lasts 4 years. The annual pecuniary cost of college (tuition and fees) is denoted by κ. College students work part time ℓ_{pt} and therefore forego earnings from full time work. Additionally, they incur the effort (or the psychic cost) of college λ. The benefit of graduating from college is higher labor earnings; consumers must graduate from college to enjoy the college wage premium. The pecuniary cost of college may be financed with any of the following sources: familial inter-vivos transfers from parents, grants from public and private sources, earnings from part time work, student loans borrowed from the federal student loan program, or student loans borrowed from the private loan market. The federal student loan program is characterized by a cumulative student loan limit $\bar{A}$ and a constant student loan interest rate r_{SL}, which is assessed starting from the year after the age of college graduation ($j > 4$). The assumptions about accruing interest implies that there is an interest-free grace period for student loans for the duration of college. Student loan payments are required to begin after the age of college graduation, and the loan balance is expected to be paid off in T_{SL} years.

Consumers can also borrow from the private student loan market. We model the private loan market using findings from our empirical analysis in section 2, as well as additional findings presented in Table 58 of section D.2 in the Appendix. In particular, we find that students have access to private loans in each cell of the household income and skill distribution, indicating that they do not need a co-signer (Table 7), and that there is a clear pecking order in which students borrow first from the federal student loan program and then go to the private market (Table 6). Additionally, using data from the SCF, we do not find evidence that private student loan interest rates vary significantly depending on ex-post outcomes such as graduation status or income or even delinquency status (Table 58). These features of the data motivate our assumption of a pooling equilibrium for the private student loan market, in which there is no discrimination in access by family income or student skill (or any other characteristics). To capture the pecking order finding of Table 6, we introduce a psychic cost for acquiring private student loans. This captures costs resulting from predatory lending and hidden fees.

After the age of college graduation, consumers with a positive student loan balance (federal, private, or both) may be either college graduates or dropouts. At this point in the life cycle, student debtors must make begin to make decisions about whether to make their required loan payments: they may choose to repay only federal loans, only private loans, both types of loan, or neither type loan. Upon paying off their student loans, consumers may save and solve a standard consumption-savings problem.[25] Consumers who refuse to make a payment on any of their student loan debt are considered delinquent, and their disposable income

[25] We assume that student loans must be paid off for consumers to save because that makes the model extremely tractable by allowing us to use two state variables for assets, federal student loans, and private student loans rather than three state variables. We believe this is not a costly assumption to make for the following reason. In an environment in which consumers cannot be delinquent, the optimal strategy would be to pay off all loans and not save as long as the interest rates on loans are higher than the savings interest rate. In our framework, this incentive might be somewhat offset because of the delinquency option, but it is not a quantitatively significant concern.

above $\bar{y}$ is garnished at the rate τ_g. Delinquent debtors also incur a penalty cost equal to a fraction ϕ_D of the missed payment for the particular year in which they are delinquent. The penalty fee and the missed payment are added to the outstanding balance for the next period.[26] Besides these pecuniary costs delinquent consumers also incur a stigma cost indexed to the type of loan, where ξ_D and ξ_D^{pr} denote the stigma costs for federal and private loans, respectively. It is important to note that it is not possible to default on either a federal or a private student loan and have the outstanding debt written off. Instead, missing a payment leads to punitive confiscation of income. This is consistent with the US federal student loan system, as well as with private student loan policies (the latter requires an official court order to garnish wages). In both cases, in the data, student loans may eventually be classified as defaulted loans, but are almost never discharged.[27]

All consumers have a child when they are aged j_f. This child will grow up and leave the household j_a years after birth. At the beginning of the period when the child leaves the household, parents altruistic choose how much to transfer inter-vivos to their child after observing the child's skill, e.[28] This transfer is motivated by parental altruism, and the parent is also overly optimistic about the likelihood of their child graduating from college once enrolled.[29]

Consumer's retire at age j_r; at this stage, they stop working and receive Social Security. Consumers survive each period with survival probability ψ_j, and live for a maximum of J periods.

The government—in addition to running the federal student loan program, providing grants for college education, and providing Social Security—incurs an exogenous government consumption cost (a fixed fraction of GDP). Government expenditures are financed through consumption taxes with a flat rate as well as progressive income taxes. Only the income tax rate adjusts, when necessary, to balance the government's budget constraint.

Lastly, output is produced by a final goods firm, which operates a Cobb-Douglas production technology in which the inputs are capital and efficiency units of labor.

3.2 Elements of the consumer life cycle problem

This section describes the consumer's life cycle problem in more detail. Several relevant primitives of the model environment are outlined first; we then provide and explain the value functions consumer's solve at each stage of their life, which incorporate these primitives.

Persistence in college enrollment Dropping out of college can occur due to an exogenous shock or by choice. As mentioned above, $p_d(j, e)$ denotes the true annual exogenous probability of dropping out of

[26]This reflects the current US system. See Yannelis (2020) for more institutional details on student loan delinquency and penalties.
[27]Specifically, for federal student loans in the United States, 270 days of missed payments is necessary before the loan is classified as being defaulted on and the associated penalties are applied.
[28]The structure of this transfer is similar to Krueger and Ludwig (2016) and Abbott, Gallipoli, Meghir, and Violante (2019).
[29]See section A.1 of the Appendix for empirical motivation of this model ingredient.

college.[30] The exogenous dropout probability function given by

$$p_d(j, e) = (1 - p(e))\rho_d(e)^{j-1}, \tag{1}$$

where $p(e)$ determines the probability of continuing in college and $\rho_d(e)$ determines the persistence of the probability of dropping out of college conditional on college year. To illustrate, if $p(e)$ is low, the student is more likely to drop out of college. If $\rho_d(e)$ is low, the college dropout probability is less persistent with each year of college, and hence, the student is less likely to drop out of college the longer they continue their education. The probability of continuing in each year of college $p_g(j, e)$, is equal to $1 - p_d(j, e)$.

Student loan payments amounts As mentioned above, consumers are expected to make payments starting at age $j = 5$ (after college graduation).[31] Both federal and private loans are expected to be paid off in T_{SL} years. Equation 2 specifies the full repayment function $\rho_R(a, j)$ for federal student loans.

$$\rho_R(j, a) = \begin{cases} -\dfrac{r_{SL}}{1 - (1 + r_{SL})^{-(T_{SL}+5-j)}} a & \text{if } a < 0 \text{ and } 4 < j \leq T_{SL} + 4 \\ -(1 + r_{SL})a & \text{if } a < 0 \text{ and } j > T_{SL} + 4 \\ 0 & \text{otherwise } (a > 0). \end{cases} \tag{2}$$

If there is an outstanding balance $(a < 0)$ and if j is still within the standard repayment period $(4 < j \leq T_{SL}+4)$, the loan is amortized with an interest rate of r_{SL} (first case in equation 2). If there is an outstanding loan and if the standard repayment period has expired $(j > T_{SL} + 4)$, then the outstanding principal plus interest is due (second case in equation 2). If $a > 0$, then there is no student loan debt, and hence, no payment.

Instead of repayment, consumers who owe payments may choose delinquency. Loans in delinquency are not discharged; instead, the agent's disposable income income above $\bar{y}$ is garnished at the rate τ_g. This leads to a partial repayment function in delinquency given by

$$\rho_D(j, a, y) = \min[\tau_g \max[y - T(y) - \bar{y}, 0], \rho_R(j, a)], \tag{3}$$

where the garnishment amount is bounded above at the full repayment amount $\rho_R(j, a)$.

College students can also borrow from the private student loan market. We use x to denote the outstanding private student loan balance. When consumers borrow on the private market $(x > 0)$, they are expected to pay off private loan debt in T_{SL} years after the age of college graduation. The loan is amortized with an interest rate r_{SL}^{pr}. Unlike the federal student loan interest rate, which is set by the government, the interest

[30] See section 2.1 for empirical evidence motivating this model specification. This model attribute is also supported by the findings of Stinebrickner and Stinebrickner (2012); in that paper, the authors argue that it is heterogeneity in ability, rather than heterogeneity in effort, that drives the college dropout decision. For example, even for students in the same major who put in the same hours of study, they find significant differences in academic performance.

[31] We impose the same structure for dropouts.

rate for private student loans is determined by market forces in a pooling equilibrium.[32] If private loans are not fully paid off within T_{SL} years, all loans and interest become due every year until the balance is fully repaid. The repayment function for private student loans is given by

$$\rho_R^{pr}(j,x) = \begin{cases} \dfrac{r_{SL}^{pr}}{1-(1+r_{SL}^{pr})^{-(T_{SL}+5-j)}}x & \text{if } x>0 \text{ and } 4<j\leq T_{SL}+4 \\ (1+r_{SL}^{pr})x & \text{if } x>0 \text{ and } j>T_{SL}+4 \\ 0 & \text{otherwise } (x=0). \end{cases} \tag{4}$$

If consumers are delinquent on private loans, their disposable income above $\bar{y}$ is garnished at the rate τ_g[33] Similar to the federal loan partial repayment function, the partial repayment for private loans is capped at the full repayment amount ρ_R^{pr}, and can be written as

$$\rho_D^{pr}(j,x,y) = \min[\tau_g^{pr}\max[y-T(y)-\bar{y},0], \rho_R^{pr}(j,x)]. \tag{5}$$

Preferences An agent's utility depends on total household consumption c, the agent's age j (which also determines whether or not they have a child), and their education status $s \in \{h, \ell\}$. It is given by

$$U(c,j,s) = \frac{\left(\frac{c}{1+\zeta\,1_{j\in\{j_f,..,j_f+j_a-1\}}}\right)^{1-\sigma}}{1-\sigma} - \lambda 1_{s=h \text{ and } j\in\{1,2,3,4\}}. \tag{6}$$

where h refers to an high-education agent who is either enrolled in college or is a college graduate, and ℓ refers to a low-education agent who did not go to college or who dropped out of college. Together with j, s indicates whether or not an agent is in college. Utility is exhibits constant relative risk aversion over per-capita household consumption, with a relative risk aversion given by σ. When the child lives with the parent, $j \in \{j_f, .., j_f + j_a - 1\}$, the child will be included in total household consumption with an adult equivalence parameter ζ. Note that college students, for whom $s = h$ and $j \in \{1, 2, 3, 4\}$), are subject to an effort/psychic cost λ.

Income Income depends on age, education, skill, AR(1) earnings productivity, and net assets, summarized by the tuple (j, s, e, η, a). Positive net assets are indicated by $a > 0$; these savings earn an interest rate r. Federal student loan debt is indicated by $a < 0$. The remaining elements of te tuple determine the income

[32] See section D.2 in the appendix for empirical evidence motivating this modeling choice.

[33] In US, private lenders are also allowed to garnish student loans as long as they acquire a court order.

function y, which is given by:

$$y_{j,s,e,\eta,a} = \begin{cases} s = h & \begin{cases} w\epsilon_{j,\ell,e}\eta\ell_{pt} & \text{if } j = 1 \\ w\epsilon_{j,\ell,e}\eta\ell_{pt} + r\max(a,0) & \text{if } 1 < j \leq 4 \\ w\epsilon_{j,h,e}\eta + r\left[\max(a,0) + Tr_j\right] & \text{if } 4 < j < j_r \\ ss_{h,e} + r\max(a,0) & \text{if } j \geq j_r \end{cases} \\ s = l & \begin{cases} w\epsilon_{j,\ell,e}\eta & \text{if } j = 1 \\ w\epsilon_{j,\ell,e}\eta + r\left[\max(a,0) + Tr_j\right] & \text{if } 1 < j < j_r \\ ss_{\ell,e} + r\max(a,0) & \text{if } j \geq j_r, \end{cases} \end{cases} \tag{7}$$

where w is the wage rate and $\epsilon_{j,s,e}$ is a deterministic life cycle productivity component that depends on age, education level, and skill. Tr_j is accidental bequests, which depends on age; accidental bequests are a consequence of the assets of the deceased being greater than zero.

When consumers first enter the labor market at age 18 ($j = 1$), their only source of income is labor earnings. If they chooses to go to college ($s = h$), they work part time and receive $w\epsilon_{j,\ell,e}\eta$ in labor earnings per unit of labor supply. Note that, because this agent is still in college, they work part time (ℓ_{pt}) and their life cycle productivity component and their AR(1) productivity component are drawn from the distribution for $s = \ell$. If a working-age consumer does not go to college or drops out of college ($s = \ell$), they work full time and receive $w\epsilon_{j,\ell,e}\eta$ in labor earnings (labor supply is inelastic and equal to one). If a working-age consumer is a college graduate, they also work full time and their labor earnings is given by $w\epsilon_{j,h,e}\eta$. In this case, the deterministic life cycle component and the AR(1) productivity component are drawn from the distribution for $s = h$. When consumers retire at age j_r, they receive Social Security $ss_{s,e}$. The level of Social Security transfers an agent receives depends on both their education and skill. [34]

After the first year ($j > 1$), consumer income includes any interest from positive net assets and accidental bequests, $r[\max[a, 0] + Tr_j]$, in addition to labor earnings or Social Security. When consumers are 18 ($j = 1$), they receive an inter-vivos transfer from their parents. Interest income on inter-vivos transfer accrues to the parents, not the transfer recipient.

Consumer problems before college graduation age (j $\leq$ 4) Given their type, (e, η, a), which records the their skill, e, idiosyncratic AR (1) productivity, η, and net assets, a, an 18-year-old ($j = 1$) has a value function represented by

$$\begin{aligned} \hat{W}(e,\eta,a) =& q\Big[\max_{\hat{d}_s\in\{0,1\}}(1-\hat{d}_s)V(1,\ell,e,\eta,a,x=0) + \hat{d}_s\hat{V}(1,h,e,\eta,a,x=0)\Big] \\ &+ (1-q)V(1,\ell,e,\eta,a,x=0) \end{aligned} \tag{8}$$

[34] See equation 30, Appendix II.

With probability q, the consumer may make a discrete college entrance decision represented by $\hat{d}_s \in \{0,1\}$, where $V(1,\ell,e,\eta,b,x=0)$ is the value of not going to college, and $\hat{V}(1,h,e,\eta,a,x=0)$ is the over-optimistic value of going to college. The first element in value function V and $\hat{V}$ denotes age (everyone starts from age 1 when making college entrance decisions.), and the second element represents school choices (ℓ denotes high school or some college while h denotes a college student or a college graduate). The last element, x, represents the balance of private student loans and is set to 0 to reflect that no one has private loans at age 18. With exogenous probability $1-q$, the consumer does not have the option to enroll, an dproceeds through life as a low-education worker ($s=\ell$).[35]

The value of not going to college (as well as the value of dropping out) for $j \leq 4$ is given by

$$
\begin{aligned}
&V(j,\ell,e,\eta,a,x) = \max_{c,a'} U(c,j,\ell) + \beta\psi_j E_{\eta'|\ell,\eta} V(j+1,\ell,e,\eta',a',x) \qquad (9)\\
&s.t.\\
&(1+\tau_c)c + a' = y_{j,e,\ell,\eta,a} + a + Tr_j - T(y_{j,e,\ell,\eta,a})\\
&a' \begin{cases} a' = a & \text{if } a < 0\\ \geq 0 & \text{otherwise}\end{cases}\\
&c \geq 0,
\end{aligned}
$$

where β is the discount factor, τ_c is the consumption tax rate, and $T(y_{j,\ell,e,\eta,a})$ is the income tax function.[36] As in Krueger and Ludwig (2016), the AR(1) productivity process depends on educational attainment. Therefore, this agent draws their next period AR(1) productivity from the expectation operator that depends on ℓ, in addition to the current shock, η. For consumers who do not go to college, net assets are always weakly positive ($a >= 0$); student loans are the only form of borrowing in our model environment, and are available only while enrolled in college. For consumers who drop out of college, who also solve 9, the stock of debt is frozen at their level of debt a until j>4, at which point they begin repaying their loan.

Equation 9 is also the value function for consumers that drop out of college, who may have outstanding student loans ($a < 0$). In that case, by construction, the consumer is not assessed interest until $j > 4$, and rolls over their debt until they start making payments after the college graduation age.

The overly optimistic value of college is given by $\hat{V}(j,h,e,\eta,a,x)$; when computing the value of college at the enrollment decision stage, consumers mistakenly believe that they will continue in college with probability $\hat{p}(e)$, whereas the true probability of continuing is actually $p_g(j,e)$. The ex-ante value of college

[35] This can also be viewed as a preference shock, rather than as a lottery for college enrollment option.

[36] See Section 3.4.

enrollment for $j = 1, 2, 3$ is therefore

$$\hat{V}(j,h,e,\eta,a,x) = \max_{\hat{c},\hat{a}',\hat{x}'} U(c,j,h) - \xi 1_{a\geq 0 \text{ and } x=0 \text{ and } (\hat{a}'<0 \text{ or } \hat{x}'>0)} - \xi^{pr} 1_{x=0 \text{ and } \hat{x}'>0} \tag{10}$$

$$+ \beta\psi_j E_{\eta'|\ell,\eta}\Big[\hat{p}(e)\max[\hat{V}(j+1,h,e,\eta',\hat{a}',\hat{x}'), V(j+1,\ell,e,\eta',\hat{a}',\hat{x}')] + (1-\hat{p}(e))V(j+1,\ell,e,\eta',\hat{a}',\hat{x}')\Big]$$

$s.t.$

$$(1+\tau_c)\hat{c} + \hat{a}' + (1-\theta(e)-\theta^{pr}(e))\kappa = y_{j,h,e,\eta,a} + a + Tr_j - T(y_{j,h,e,\eta,a}) + (\hat{x}' - x)$$

$$\hat{c} \geq 0$$

$$\hat{a}' \geq -\bar{A}\left(\frac{j}{4}\right)[(1-\theta(e)-\theta^{pr}(e))\kappa + \bar{c}]$$

$$\hat{a}' \leq a \ \text{ if } \ a \leq 0$$

$$\hat{x}' - x \in \left[0, [(1-\theta(e)-\theta^{pr}(e))\kappa + \bar{c}] - \left(\bar{A}\left(\frac{j}{4}\right)[(1-\theta(e)-\theta^{pr}(e))\kappa + \bar{c}] - \big(\max[-\hat{a}',0] - \max[-a,0]\big)\right)\right],$$

where ξ is the psychic cost of acquiring any student loan and ξ^{pr} is the additional costs associated with acquiring any private student loans.[37] $\theta(e)$ and $\theta^{pr}(e)$ are the share of tuition and fees that are paid for by public and private grants, and are a function of skill. $\bar{c}$ is the amount that can be borrowed for room and board expenses while in college. These consumers m ay also choose to drop out after the first year of college captured by the expression $\max[\hat{V}(j+1,h,e,\eta',\hat{a}',\hat{x}'), V(j+1,\ell,e,\eta',\hat{a}',\hat{x}')]$. College students can borrow from federal student loans up to a limit equal to $\bar{A}\left(\frac{j}{4}\right)[(1-\theta(e)-\theta^{pr}(e))\kappa + \bar{c}]$, where $\bar{A}$ is the number of years worth of net tuition and fees plus room and board expenses that the student loan limit can cover.[38] Students can also borrow additional funds from the private student loan market. The combined amount of borrowing from both the private and federal loans cannot exceed total net tuition plus room and board costs, $(1-\theta(e)-\theta^{pr}(e))\kappa + \bar{c}$. Payments on federal and private student loans do not begin until the age after college graduation.

When constructing an overly-optimistic value for college, the post-college continuation value conditional on graduation is based on $E_{\eta'|h,\eta}$ rather than $E_{\eta'|\ell,\eta}$. Furthermore, there will be no endogenous dropout decision in the continuation value at this age, because in the next period the agent will have graduated from college. The rest of the value function for the final year of college remains unchanged.[39]

Note that when consumers make the college entrance decision in (8), they are overly optimistic and will use the inflated value of college from (10) and (23) to compute their expected value. However, we assume that consumers learn their true dropout probabilities in the first year of college so that, while enrolled,

[37] Psychic costs of getting a loan can be thought of as mental anguish from completing paperwork, which may be excessively complicated as noted by Dynarski and Scott-Clayton (2008) and Dynarski, Libassi, Michelmore, and Owen (2021). Costs particular to private loans capture predatory lending and hidden fees; this parameter allows for a pecking order for borrowing first from federal student loans and then private student loans as noted in Table 6.

[38] For example, if $\bar{A}$ is equal to four, then the student loan limit is equal to four years of net tuition and fees plus room and board. The multiplier $\frac{j}{4}$ is an adjustment for the fact that the total student loan limit increases with each year of college.

[39] See (23) in Appendix II.

the consumer's consumption-savings and dropout decisions are based on the following value function for $j = 1, 2, 3$

$$V(j,h,e,\eta,a,x) = \max_{c,a',x'} U(c,j,h) - \xi 1_{a\geq 0 \text{ and } x=0 \text{ and } (a'<0 \text{ or } x'>0)} - \xi^{pr} 1_{x=0 \text{ and } x'>0} \tag{11}$$

$$+ \beta\psi_j E_{\eta'|\ell,\eta}\Big[p_g(j,e)\max[V(j+1,h,e,\eta',a',x'),V(j+1,\ell,e,\eta',a',x')]$$

$$+ (1-p_g(j,e))V(j+1,\ell,e,\eta',a',x')\Big]$$

$$s.t.$$

$$(1+\tau_c)c + a' + (1-\theta(e)-\theta^{pr}(e))\kappa = y_{j,h,e,\eta,a} + a + Tr_j - T(y_{j,h,e,\eta,a}) + (x'-x)$$

$$c \geq 0$$

$$a' \geq -\bar{A}\left(\frac{j}{4}\right)[(1-\theta(e)-\theta^{pr}(e))\kappa + \bar{c}]$$

$$a' \leq a \ \text{ if } \ a \leq 0$$

$$x' - x \in \left[0, [(1-\theta(e)-\theta^{pr}(e))\kappa+\bar{c}] - \left(\bar{A}\left(\frac{j}{4}\right)[(1-\theta(e)-\theta^{pr}(e))\kappa+\bar{c}] - \big(\max[-a',0]-\max[-a,0]\big)\right)\right].$$

The only difference between this value function and the over-optimistic value function given by (10) is that, in (11), consumers use the true probabilities of continuing in college in computing their expected value from college. Again, in the final year of college ($j = 4$), if they are graduating from college their continuation of value will be based on $E_{\eta'|h,\eta}$ rather than $E_{\eta'|\ell,\eta}$. The rest of the consumer's value function for the final year of college remains unchanged.[40]

3.3 Consumer problems after college graduation age (j > 4)

Consumers begin student loan repayment the year after college graduation age, regardless of whether they complete college or drop out.[41]

In every period where $j > 4$, consumers with student loan debt choose between repayment of both loans, delinquency only on federal student loans, delinquency only on private student loans, and delinquency on both federal and private student loans. The idiosyncratic state of a consumer while $j > 4$ and $j \neq j_f + j_a$

[40] See (24) in Appendix II.

[41] In the United States, federal student loans typically have a six month grace period after graduation, in which repayment does not need to be made. Since our model period is one year, the fact that repayment starts one period after graduation is consistent with the current policy for college graduates. For dropouts, it is a slight approximation.

is given by the tuple (j, s, e, η, a, x). Their value function is given by

$$V(j,s,e,\eta,a,x) = \max_{d_f\in\{0,1\},d_x\in\{0,1\}} (1-d_f)(1-d_x)V^R(j,s,e,\eta,a,x)+ \tag{12}$$
$$d_f(1-d_x)V^{D_f}(j,s,e,\eta,a,x) + (1-d_f)d_xV^{D_x}(j,s,e,\eta,a,x) + d_fd_xV^D(j,s,e,\eta,a,x),$$

where $d_f \in \{0,1\}$ and $d_x \in \{0,1\}$ denote the federal and private student loan repayment decisions, respectively. $V^R(j,s,e,\eta,a,x)$ denotes the value of repayment on both loans, $V^{D_f}(j,s,e,\eta,a,x)$ denotes the value of delinquency on only federal loans, $V^{D_x}(j,s,e,\eta,a,x)$ denotes the value of delinquency on only private loans, and $V^D(j,s,e,\eta,a,x)$ denotes the value of delinquency on both types loans.

The value of repayment for $j > 4$ and $j \neq j_f + j_a$ is given by

$$V^R(j,s,e,\eta,a,x) = \max_{c,a'} U(c,j,s) + \beta\psi_j E_{\eta'|s,\eta} V(j+1,s,e,\eta',a',x') \tag{13}$$
$$s.t.$$
$$(1+\tau_c)c + a' = y_{j,s,e,\eta,a} + a + 1_{\{a<0\}} r_{SL} a + Tr_j - T(y_{j,s,e,\eta,a}) - \rho_R^{pr}(j,x)$$
$$a' \begin{cases} = (1+r_{SL})a + \rho_R(j,a) & \text{if } \ a<0 \\ \geq 0 & \text{if } \ a \geq 0 \ \text{ and } \ x=0 \\ = 0 & \text{otherwise } (a \geq 0 \ \text{ and } \ x>0) \end{cases}$$
$$x' = x(1+r_{SL}^{pr}) - \rho_R^{pr}(j,x)$$
$$c \geq 0.$$

The consumer who chooses repayment must make a payment equal to $\rho_R(j,a)$ on their federal student loans. If this consumer has outstanding federal loans ($i.e.$, $a < 0$), then $a' = a(1+r_{SL}) + \rho_R(j,a)$. Consumers with private student loans must pay $\rho_R^{pr}(j,v,x)$. As in Ionescu and Simpson (2016), we assume debtors cannot choose to pay down their federal or private loans faster than the required payment amount.[42] Consumers can alternatively choose delinquency on either type of loan or on both loans; this delinquency decision indexes their value function when in default.

If a consumer is delinquent on federal loans only, their value function for $j > 4$ and $j \neq j_f + j_a$ is given by

$$V^{D_f}(j,s,e,\eta,a,x) = U(c,j,s) - \xi_D + \beta\psi_j E_{\eta'|s,\eta} V(j+1,s,e,\eta',a',x') \tag{14}$$
$$s.t.$$
$$(1+\tau_c)c = y_{j,s,e,\eta,a} + Tr_j - T(y_{j,s,e,\eta,a}) - \rho_D(j,a,y_{j,s,e,\eta,a}) - \rho_R^{pr}(j,x)$$
$$a' = (1+r_{SL})a + \rho_D(j,a,y_{j,s,e,\eta,a}) - \phi_D[\rho_R(j,a) - \rho_D(j,a,y_{j,s,e,\eta,a})]$$
$$x' = x(1+r_{SL}^{pr}) - \rho_R^{pr}(j,,x),$$

[42] See Section 3.2. This assumption treats federal loans and private loans equally in repayment, and it greatly reduces the computational burden.

where ξ_D is the stigma cost of choosing delinquency on federal loans. In the case of non-repayment of federal loans, consumers do not make a consumption-savings decision. Instead, they have their income garnished to make a partial payment of $\rho_D(a, y_{j,s,e,\eta,a})$. Therefore, they consume whatever remains from their disposable income plus accidental bequests after making the partial payment on federal loans and full payment on private loans. As mentioned above, ϕ_D is the fraction of missed payment (difference between full payment and partial payment) that is charged as a collection fee. The outstanding principal plus interest is then augmented by the missed payment plus the collection fee (net of any partial payment). Similarly, if a consumer chooses delinquency only on private loans only, then their value function for $j > 4$ and $j \neq j_f + j_a$ is given by

$$
\begin{aligned}
&V^{D_x}(j,s,e,\eta,a,x) = U(c,j,s) - \xi_D^{pr} + \beta\psi_j E_{\eta'|s,\eta} V(j+1,s,e,\eta',a',x') \qquad (15)\\
&s.t.\\
&(1+\tau_c)c + a' = y_{j,s,e,\eta,a} + a + 1_{\{a<0\}} r_{SL} a + Tr_j - T(y_{j,s,e,\eta,a}) - \rho_D^{pr}(j,x,y_{j,s,e,\eta,a})\\
&a' = 1_{a<0}(1+r_{SL})a + \rho_R(j,a)\\
&x' = (1+r_{SL}^{pr})x - \rho_D^{pr}(j,x,y_{j,s,e,\eta,a}) + \phi_D[\rho_R^{pr}(j,x) - \rho_D^{pr}(j,x,y_{j,s,e,\eta,a})],
\end{aligned}
$$

where ξ_D is the stigma cost of choosing delinquency on private loans. As in the case of delinquency on only federal loans, here the consumer does not make a consumption-savings decision. Instead, they pay the fixed amount of federal student loans repayment $\rho_R(j,a)$, and they are subject to wage garnishment due to defaulting in private loans. The garnishment amount is denoted by $\rho_D^{pr}(j,x,y_{j,s,e,\eta,a})$ as described in Section 3.2. Similar to the case of delinquency on federal loans, here the consumer is subjected to a collection fee which is equal to a fraction ϕ_D multiplied by the difference between full payment and partial repayment in private loans.

Lastly, the valueof choosing delinquency on both types of loans is given by

$$
\begin{aligned}
&V^{D}(j,s,e,\eta,a,x) = U(c,j,s) - \xi_D - \xi_D^{pr} + \beta\psi_j E_{\eta'|s,\eta} V(j+1,s,e,\eta',a',x') \qquad (16)\\
&s.t.\\
&(1+\tau_c)c = y_{j,s,e,\eta,a} + Tr_j - T(y_{j,s,e,\eta,a}) - \rho_D(j,a,y_{j,s,e,\eta,a}) - \rho_D^{pr}(j,x,y_{j,s,e,\eta,a})\\
&a' = (1+r_{SL})a + \rho_D(j,a,y_{j,s,e,\eta,a}) - \phi_D[\rho_R(j,a) - \rho_D(j,a,y_{j,s,e,\eta,a})]\\
&x' = (1+r_{SL}^{pr})x - \rho_D^{pr}(j,x,y_{j,s,e,\eta,a}) + \phi_D[\rho_R^{pr}(j,x) - \rho_D^{pr}(j,e,x,y_{j,s,e,\eta,a})].
\end{aligned}
$$

A consumer who chooses this outcome is subject to wage garnishment and a collection fee (same as the previous two cases) both from federal student loan program as well as the private student loan lenders and their consumption for the current period and outstanding loan balances for the next period follow from the same set of delinquency rules described above.

As in Krueger and Ludwig (2016) and Abbott, Gallipoli, Meghir, and Violante (2019), we assume that parents are altruistic towards their children and make a one-time familial inter-vivos transfer when the child leaves the household, when parents are at age $j = j_f + j_a$. In this period, and after receiving this transfer, the child becomes an independent decision maker. Parents make their decisions after observing their child's skill, e_c. The parent's value function is given by

$$V(j,s,e,\eta,a,x,e_c) = \sum_e \pi(e_c|s)\big[\max_{d_f\in\{0,1\},d_x\in\{0,1\}}(1-d_f)(1-d_x)V^R(j,s,e,\eta,a,x,e_c)+ \tag{17}$$
$$d_f(1-d_x)V^{D_f}(j,s,e,\eta,a,x,e_c)+(1-d_f)d_xV^{D_x}(j,s,e,\eta,a,x,e_c)+d_fd_xV^D(j,s,e,\eta,a,x,e_c)\big],$$

where $\pi(e_c|s)$ is the child's high school GPA probability function that depends on the parent's education. The value of repayment for $j = j_f + j_a$ is given by

$$V^R(j,s,e,\eta,a,x,e_c) = \max_{c,a',b} U(c,j,s) + \beta\psi_j E_{\eta'|s,\eta}V(j+1,s,e,\eta',a',x') + \beta_c E_{\eta'|\ell}\hat{W}(e_c,\eta',b) \tag{18}$$

$$s.t.$$
$$(1+\tau_c)c + a' + b = y_{j,s,e,\eta,a} + a + 1_{\{a<0\}}r_{SL}a + Tr_j - T(y_{j,s,e,\eta,a}) - \rho_R^{pr}(j,x)$$
$$a'\begin{cases} = (1+r_{SL})a + \rho_R(j,a) & \text{if } a<0 \\ \geq 0 & \text{if } a\geq 0 \text{ and } x=0 \\ = 0 & \text{otherwise } (a\geq 0 \text{ and } x>0)\end{cases}$$
$$x' = x(1+r_{SL}^{pr}) - \rho_R^{pr}(j,x)$$
$$b\begin{cases} = 0 & \text{if } a<0 \text{ or } x>0 \\ \geq 0 & \text{otherwise } (a\geq 0 \text{ and } x=0)\end{cases}$$
$$c\geq 0,$$

where b is the familial inter-vivos transfer, $\hat{W}(e_c,\eta',b)$ is the child's value function, and β_c disciplines the intensity of parental altruism towards the child. The child's AR(1) productivity η' is drawn from the stationary distribution for an agent without a college degree. If the parent has any federal or private student loans, they will not make a familial inter-vivos transfer to their child ($b = 0$), which is consistent with the assumption that consumers cannot save until they have paid off their student loans (a rare event in equilibrium). This consumer's value functions for delinquency are given in Appendix D.4.

3.4 The Government budget constraint and production firm problem

The government collects consumption and income taxes. The income tax function follows Heathcote et al. (2017), and is given by

$$T(y) = y - \gamma y^{1-\tau_p}, \tag{19}$$

where τ_p governs the tax progressivity of income taxes and γ is used to balance the government budget in every period. The government uses its tax revenue to finance government consumption, Social Security transfers, college grants, and the federal student loan program.

The production function is Cobb-Douglas given by

$$K^{\alpha}(ZL)^{1-\alpha}, \tag{20}$$

where K is aggregate capital stock, Z is aggregate labor productivity, L is total efficiency units of labor, and α is the capital share. This production function assumes that an efficiency unit of labor from a college graduate is perfectly substitutable with an efficiency unit of labor from a worker without a college degree. Because the general equilibrium effects of our policy analyses in regard to wage rates are small, this assumption will not affect our results. The capital stock depreciates at rate δ. The representative firm rents capital at an interest rate $r+\delta$ and hires workers at the wage rate w. First order conditions from the firm's profit maximization problem leads to the following standard conditions given by

$$r = \alpha K^{\alpha-1}(ZL)^{1-\alpha} - \delta \tag{21}$$

$$w = (1-\alpha)K^{\alpha}L^{-\alpha}Z^{1-\alpha}. \tag{22}$$

The definition of the equilibrium is given in Appendix D.4.

4 Model Parameterization

The parameters of this model are divided into those set outside the model and those calibrated jointly within it to target moments in the US economy. Tables 9 and 10 present parameters determined outside of the model equilibrium, and Table 11 presents internally calibrated parameters.

The parameters presented in Table 9 relate to education. In Panel A, parameters governing the federal student loan program are reported. First, we set the aggregate federal student limit, $\bar{A}$, to the current cumulative yearly borrowing limit for four years of college, normalized by the average annual net tuition and fees plus room and board.[43] The parameter value for $\bar{A}$ therefore represents the number of years of college tuition and fees, plus room and board, that 4 years of annual limits are able to finance.[44] We set the student loan interest rate, r_{SL}, to the average real interest rate on federal student loans from 1992-2019 (FSA (2022)). We set the maximum number of years that one can be in repayment for a student loan, T_{SL}, to 20 (CRS (2019)).[45] The garnishment rate conditional on delinquency on federal student loans, τ_g, is set it to 15 percent, which is the current rate of garnishment for these loans (Yannelis (2020)). We set the garnishment rate for private loans, τ_g^{pr}, equal to the garnishment rate for federal loans: this is consistent with the US system, where garnishment is allowed for delinquent private loans as long as the loan provider receives a court order. The

[43] This limit has been in place since July 1, 2012. In the federal student loan program, there are yearly limits and lifetime limits on borrowing. Yearly limits depend on one's academic year (freshman, sophomore, etc.).

[44] The yearly limits we use for this parameter are for dependent students, who represent the vast majority of federal student loan borrowers. See Table 50 in the Appendix.

[45] In the US, those with student loans may choose between a standard repayment plan of ten years and an income based repayment plan. Time limits for repayment vary across these plans from 10 to 25 years (CRS (2019)). Note that agents in our model can choose to repay their loan at a faster rate, for e.g., in ten years like the standard repayment plan, or at a slower rate, for e.g., in 20 years like the income based repayment plan.

Table 9: Externally estimated parameters, part 1

Parameter	Description	Data Target	Value
Panel A: Federal student loan program			
$\bar{A}$	Limit	Ave. 2016-2018, CRS (2019) and NCES (2019)	1.493
r_{SL}	Interest rate	Ave. 2016-2018, FSA (2022)	0.025
T_{SL}	Maximum years to repay	CRS (2019)	10
τ_g	Federal SL garnishment rate	Yannelis (2020)	0.150
τ_g^{pr}	Private SL garnishment rate	$\tau_g^{pr} = \tau_g$	0.150
ϕ_D	Student loan penalty rate	Luo and Mongey (2019)	0.185
Panel B: Tuition subsidies and college working hours			
$\theta(e=1)$	Public tuition subsidy | skill $= e$	HSLS, Table 48, and Krueger and Ludwig (2016)	0.286
$\theta(e=2)$			0.332
$\theta(e=3)$			0.360
$\theta^{pr}(e=1)$	Private tuition subsidy | skill $= e$		0.122
$\theta^{pr}(e=2)$			0.142
$\theta^{pr}(e=3)$			0.154
ℓ_{pt}	Part-time working hours while in college	HSLS, Table 44	0.319
Panel C: Child skill given parental education			
$\pi(e_c=1\|s=\ell)$	Fraction skill $= e$ | Parental education $s = \ell$	HSLS, Table 39	0.424
$\pi(e_c=2\|s=\ell)$			0.343
$\pi(e_c=3\|s=\ell)$			0.233
$\pi(e_c=1\|s=h)$	Fraction skill $= e$ | Parental education $s = h$		0.185
$\pi(e_c=2\|s=h)$			0.322
$\pi(e_c=3\|s=h)$			0.493

Notes: Panel A reports the policy parameters of the federal student loan program. Panel B reports the proportional tuition subsidy rates from public and private sources (i.e., grants and scholarships), and time spent working while in college. Panel C reports the conditional probability of drawing high school GPA e given parental education s. Data sources are provided in the second column. Panels B and C use moments from the HSLS; see table notes in the the referenced tables above for survey weights used. HSLS:09 data source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File.

student loan penalty rate, ϕ_D, which is common to both federal and private loans, is set to 0.185 following Luo and Mongey (2019). This penalty rate captures non-repayment fines and collection fees for the loan provider.

Panel B of Table 9 reports the estimated proportional tuition subsidies (grants and scholarships) from public and privates sources, as well as the working time available during college. To calibrate shares of college tuition subsidized by the government $\theta(e)$ and shares of college tuition subsidized private beneficiaries $\theta^{pr}(e)$, we combine estimates from HSLS with estimates from Krueger and Ludwig (2016). First, we compute shares of tuition subsidized by either government or private sources by skill quantile in the HSLS:09.[46] Next, we use estimates from Krueger and Ludwig (2016) showing that government subsidies pay for 38.8 percent of total tuition and private subsidies pay for 16.6 percent of total tuition. This implies the government's share of total tuition subsidies is 70 percent and private beneficiaries' share of total tuition subsidies is 30 percent. To assign values to $\theta(e)$, we multiply the total share of tuition subsidized by the government or private entities by 0.7, and to assign $\theta^{pr}(e)$ we multiply it by 0.3. The generosity of grants from both sources is increasing in the student skill quantile, e. Working hours while in college, ℓ_{pt}, is set using the average time spent working as a fraction of 40 hours a week for students enrolled in a 4-year program as reported in the HSLS.

Panel C of Table 9 reports parameter values for $\pi(e_c|s)$, which is the conditional distribution of child skill

[46] In the HSLS:09, we cannot distinguish if the subsidy was received from the government or a private sources.

given parental education. Note that, in the HSLS, parents with a college education are more likely to have children with higher skill (high school GPA), but nevertheless a significant fraction of children with highly educated parents are in the lowest skill bin.

Next, we turn to Table 10, which presents parameters calibrated outside of the model equilibrium that are not related to education. Panel A reports parameters governing demographics. The first four rows in this panel govern the length of different phases of life: first, child bearing age, j_f, is set to 13, which implies agents have a child when they turn 30; second, the number of years before the child moves out, j_a, is set to 18, which implies agents move out and make the college entrance decision when they turn 18; third, the retirement age, j_r, is set such that agents retire at 65; and, fourth, the maximum life span J is set such that agents live for at most 100 years. As in Krueger and Ludwig (2016), we set conditional survival probabilities ψ_j for $j = 1, ..., j_f + j_a - 1$ to one. This avoids there being children without parents in the model.[47] For $j \geq j_f + j_a$ (i.e., for agents starting at age 48), we use estimates from the Social Security Administration life tables. Values for ψ_j are not reported in Table 10 because there is one value for each age j.

Panel B of Table 10 reports parameters which govern preferences and technologies in the model. The relative risk aversion parameter, σ, to 2, which is standard in the macro literature (Chetty, 2006). The capital share parameter, α, to 0.36, which is implied by a labor share of 0.64 (Elsby et al., 2013).[48] The depreciate rate of capital, δ, is set to 0.076, as in Krueger and Ludwig (2016). To discipline the life cycle productivities $\epsilon_{j,s,e}$, which depend on age, j, education, s, and skill endowment, e, we perform the estimation described in Appendix B with results summarized in Appendix Table 27.

Table 10: Externally estimated parameters, part 2

Parameter	Description	Data Target	Value
Panel A: Demographics			
j_f	Child bearing age	30 years	13
j_a	Years for child to move out	18 years	18
j_r	Retirement age	65 years	48
J	Maximum life span	100 years	83
ψ_j	Survival probability	SSA (2020)	-
Panel B: Preferences & technology			
σ	Risk aversion	Standard value, see Table 1 Panel D of Chetty (2006)	2
α	Capital share	Elsby et al. (2013)	0.360
δ	Depreciation rate	Krueger and Ludwig (2016)	0.076
$\epsilon_{j,s,e}$	Earnings life cycle profile	PSID and NLSY	See Table 27
Panel C: Government			
τ_c	Consumption tax rate	Krueger and Ludwig (2016)	0.050
τ_p	Income tax progressivity	Heathcote et al. (2017)	0.181
g	Government consumption	Ave. share of GDP 2016-2018, BEA (2022a) and BEA (2022b)	0.141

Notes: This table contains parameters set outside of the model which are not related to education. Panel A reports parameters that govern model demographics; Panel B reports those which govern preferences and technologies; Panel C reports government policy parameters that are set exogenously.

Panel C, the last panel of of Table 10, contains values for parameters related to the government. The con-

[47]This is not a significant abstraction; young working-age consumers have high survival probabilities.
[48]There is a lively and ongoing debate about time trends in the labor share (see the discussion in Grossman and Oberfield (2021)). Our results are not sensitive to the value of this parameter.

sumption tax rate τ_c is set to 5 percent (Krueger and Ludwig, 2016); the progressivity of the income tax function, τ_p, to 0.181 as in Heathcote, Storesletten, and Violante (2017); and, lastly, government consumption as a share of GDP, g, is set to 0.17 using figures from the Bureau of Economic Analysis (BEA).

The remaining parameters are calibrated jointly in equilibrium to target moments of the US economy. The parameter symbols and descriptions, as well as the target moments and corresponding model moments from this exercise, are reported in Table 11.[49] The first column of the table gives the parameter symbol; the second column, the parameter description; the third column, the parameter value. Columns 4 through 6 give the target moment's data source, the moment in the data, and the moment in the model, respectively.

Panel A of Table 11 presents parameters governed by moments from the HSLS:09. The first two objects are the graduation probability, $p\left(e\right)$ and the persistence of the dropout shock, $\rho_d\left(e\right)$. Note that both of these objects depend on the student's skill endowment quantile, e. These objects are mostly governed by moments from the data on persistence in college, in particular persistence rates to the fourth academic year (Y4), given enrollment in a 4-year degree (Y1), and persistence to the fourth academic year given persistence to the second academic year (Y2). The last two rows in Panel A contain the credit market frictions, ξ_L and ξ_L^{pr}, which govern loan search or debt aversion and the particular cost of taking our a private loan, respectively. These parameters allow private loans to have uptake costs which differ from those of federal (public) student loans. For example, navigating the private student loan market could be more challenging than taking out a federal loan: private lenders may offer complex contracts with adverse terms hidden in the fine print, while federal loans are more standardized. It may also be more difficult to *find* a private lender, compared to borrowing from the government. Consistent with these examples, the calibrated parameter values for ξ_L and ξ_L^{pr} indicate that to match the private loan uptake rate compared to the overall student loan uptake rate (public or private), private loans must have a higher cost of uptake.

Panel B of Table 11 reports parameters which are mostly governed by moments from the NLSY97. These include college effort costs, λ, which is governed mostly by observed college enrollment rates before age 25; the college attendance probability, q, which represents a frictional application process and is chosen to target the enrollment rate of the top skill quantile; and $\hat{p}\left(e\right)$, which is a vector of the minimum expected graduation probabilities for each skill quantile. These probabilities are chosen to align college student expectations about graduation likelihood in the model with the data.

Panel C of Table 11 contains the parameters which are mostly governed by empirical targets from sources other than the HSLS:09 and NLSY97 and are related to education. First is college room and board costs, $\bar{c}$, which is set using the average annual value from 2016 to 2018 as reported in tabulations prepared by the National Center for Education Statistics (NCES (2019)). Second is the yearly net tuition, κ, which is set using different information from the same NCES tabulation as a target. The third row is the income exempt from garnishment in default, $\bar{y}$, which is set to the current amount that is exempt in the US: 15.1 percent of GDP per capita for the population 18 and over, as documented in Yannelis (2020). The fourth and fifth rows of Panel C report the parameters governing costs of being delinquent in repayment of public loans

[49] Although parameters and moments are grouped in Table 11 using the most significant one-to-one relationship between each parameter and target moment, the parameters are calibrated jointly and each parameter can affect all targets moments.

Table 11: Internally calibrated parameters

Parameters			Moments		
Parameter	Parameter Description	Value	Target Source	Data	Model
Panel A: Moments from the HSLS:09					
$p(e=1)$	Graduation probability parameter | skill $= e$	0.580	Y4 | Y1, Table 3	*0.388*	*0.389*
$p(e=2)$		0.743		*0.563*	*0.564*
$p(e=3)$		0.913		*0.787*	*0.787*
$\rho_d(e=1)$	Dropout probability persistence | skill $= e$	0.551	Y4 | Y2, Table 43	*0.672*	*0.671*
$\rho_d(e=2)$		0.631		*0.757*	*0.755*
$\rho_d(e=3)$		0.863		*0.862*	*0.863*
ξ_L	Any loan search and debt aversion cost	0.027	Uptake of student loans by end of Y3, Table 6	*0.650*	*0.590*
ξ_L^{pr}	Private loan cost	4.182	Uptake of private loans by end of Y3, Table 6	*0.200*	*0.200*
Panel B: Moments from the NLSY97					
λ	College effort	0.397	Enrolled BA by age 25, Table 20	*0.468*	*0.469*
q	College attendance probability	0.765	Enrolled BA by age 25 | skill $e=3$, Table 16	*0.765*	*0.765*
$\hat{p}\,(e=1)$	Min. expected graduation probability | skill $= e$	0.949	Expected graduation rate, Table 1	*0.809*	*0.810*
$\hat{p}\,(e=2)$		0.968		*0.872*	*0.877*
$\hat{p}\,(e=3)$		0.983		*0.934*	*0.935*
Panel C: Moments from other sources related to education					
$\bar{c}$	College room and board consumption	0.147	$\frac{\text{Average room + board}}{\text{GDP pc 18+ 2016-2018}}$, NCES (2019)	*0.147*	*0.147*
κ	Annual tuition	0.171	Yearly Net Tuition and Fees, NCES (2019)	*0.088*	*0.088*
$\bar{y}$	Exempt income in SL default	0.151	$\frac{\text{Exempt earnings}}{\text{GDP pc 18+ 2016-2018}}$, Yannelis (2020)	*0.151*	*0.151*
ξ_D	Federal loan delinquency cost	0.168	Average cohort default rate 2016-2018, FSA (2021)	*0.090*	*0.091*
ξ_D^{pr}	Private loan delinquency cost	1.103	$\frac{\text{90+ days delinquent debt}}{\text{Private debt in repayment}}$, Amir et al. (2021)	*0.074*	*0.073*
Panel D: Moments from other sources not related to education					
Z	Aggregate labor productivity	0.307	GDP pc 18+	*1.000*	*1.000*
β	Discount factor	0.974	Capital to output ratio, Figure 3 Jones (2016)	*3.000*	*3.000*
β_c	Altruism for children	0.210	$\frac{\text{Average transfer}}{\text{GDP pc 18+ 2016-2018}}$, Abbott et al. (2019)	*0.637*	*0.637*
χ	Social Security replacement rate	0.189	$\frac{\text{SS expenditure}}{\text{GDP 2016-2018}}$, BEA (2022a) and BEA (2022c)	*0.048*	*0.048*

Notes: Panel A reports parameters related to data moments from the HSLS: all moments use panel weights and specific weights used are noted in referenced Tables; data source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Panel B is parameters related to moments from the NLSY97; Panel C reports parameters and moments that use a variety of other sources and which are related to education; and Panel D presents parameters with related moments from a variety of other sources which are not related to education.

(ξ_D) or private loans (ξ_D^{pr}). These parameters are set so that the model matches either default or delinquency rate from the relevant pool of debtors. For public loans, this moments is reported by the Office of Federal Student Aid (FSA (2021)); for private loans, we use an industry report (Amir et al., 2021).

Panel D, the last panel of Table 11, contains the remaining jointly calibrated parameters. These parameters are mostly governed by moments which are unrelated to education and which are not from the HSLS or the NLSY97. Aggregate labor productivity, Z, is set so that GDP per capita for the population 18 and over is 1 in the model. The discount factor, β, is calibrated to target a capital to output ratio of 3, consistent with (Jones, 2016). The parameter β_c, which governs the degree of altruism towards children, is set so that the model matches the average parental transfers to their children, as computed using the NLSY97 by Abbott, Gallipoli, Meghir, and Violante (2019). Finally, the Social Security replacement rate, χ, is calibrated such that the model matches the average ratio of total social security expenditure to GDP from 2016 to 2018, as measured by the BEA.

5 Properties of the Model Economy

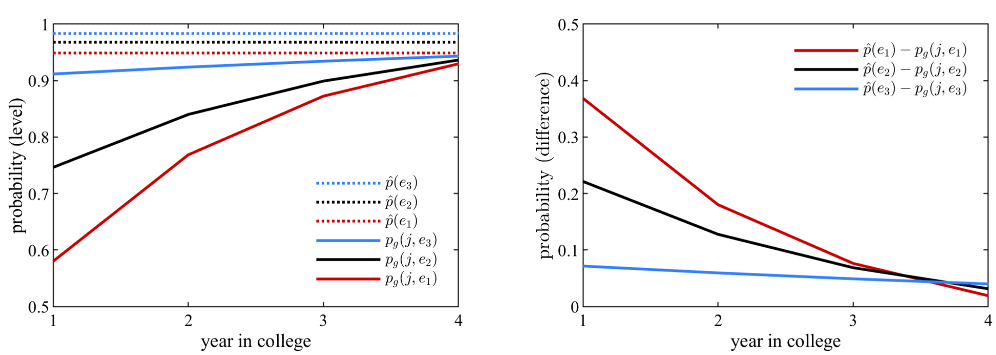
This section presents several properties of the calibrated model economy which are related to our main policy exercises. First, in section 5.1, we examine the role of over-optimism in generating observed enrollment patterns, and quantify the extent of over-enrollment in college by skill quantile. Second, in section 5.2, we examine the role of borrowing constraints on federal student loans and family transfers in generating patterns of enrollment by family income and skill quantile in the baseline equilibrium.

5.1 Illustration of college graduation probabilities and over-optimism

Figure 1 presents a visual illustration of college graduation probabilities and the extent of over-optimism in the model's baseline equilibrium. This information is broken down by student skill quantile and is shown as students progress through four years of college, with college academic year on the x-axis of both subfigures. Subfigure 1a plots the true probabilities of graduation, $p_g(j, e)$, along with the expected probability of college graduation, $\hat{p}(e)$, for agents in each skill quantile e. True probabilities are solid lines, while expected probabilities are graphed as dotted lines. The expected probability of graduating does not change as the student progress through all four years of college.

Subfigure 1b plots the extent of over-optimism in the model, which is computed as the difference between the over-optimistic probability of graduation and the true probability of graduation. Consistent with our findings in the NLSY97, our model predicts that students in lower skill quantiles exhibit greater over-optimism. Furthermore, students are more over-optimistic in their early years of college.

To summarize, Figure 1 indicates that the model captures three important attributes of the data: first, the true probability of graduating increases as students progress through college; second, the true probability of graduating from college is higher among higher skill quantiles; and, third, the expected probability of college graduation, $\hat{p}(e)$, is higher than the true graduation probabilities for all students in all years of college.



(a) Probabilities of college graduation

(b) Extent of over-optimism

Figure 1: Illustration of college graduation probabilities and over-optimism

Note: Figure 1 presents moments from the calibrated baseline equilibrium of the model. Figure 1a plots the true probabilities of graduation $p_g(j, e)$ along with the over-optimistic probabilities of graduation $\hat{p}(e)$. Figure 1b plots the extent of over-optimism, computed by subtracting the true probability of graduation from the over-optimistic probability of graduation.

We next examine the extent of over-enrollment in college that is generated by this over-optimism. In column (1) of Table 12, we report enrollment rates for each skill quantile computed in the NLSY97 (where a student's skill quantile is assigned using their high school GPA). Column (2) contains enrollment rates for high school graduates for each skill quantile, e, in the model's calibrated baseline equilibrium. These rates align closely with the data because we target average college enrollment and enrollment rate in the top skill bin. Column (3) reports enrollment rates in the model when we shut off over-optimism, setting $\hat{p}(e) = p_g(j, e)$. In this column enrollment rates decrease compared to the baseline, especially among low skill students. The difference between enrollment rates in columns (2) and (3), reported in column (4), we define as the extent of over-enrollment in that row's skill quantile. Thus, because the model framework allows us to predict enrollment rates in a counterfactual world without over-optimism, we can use the model to conclude that over-enrollment due to mistaken beliefs is highest among low skill students.

Table 12: College enrollment statistics by skill quantile

	College enrollment rate (% of skill quantile)			**Over-enrollment**
	Data	**Baseline model**	**No over-optimism**	**Percentage point difference**
Skill quantile (e)	**(1)**	**(2)**	**(3)**	**(4) = (3) - (2)**
1	22.81	24.28	4.00	20.28
2	44.09	44.02	31.22	12.80
3	76.49	76.49	69.50	6.99

Note: Table 12 presents enrollment rates in the data and model by skill quantile (assigned with high school GPA in the data). Enrollment rates are computed after high school graduation as percentages of the skill bin's population who enroll in a 4-year degree. Column (1) reports the enrollment rates in the data, column (2) reports the enrollment rate in the baseline model, column (3) reports the enrollment rate in the baseline model equilibrium when consumers are not over-optimistic ($\hat{p}(e) = p_g(j, e)$), and finally column (4) reports over-enrollment (difference between columns (3) and (2)).

Student loans and repayment delinquency among dropouts

Over-optimism leads to over-enrollment in college, with consequences for the mass of college dropouts in the population. Table 13 compares the calibrated baseline model equilibrium against untargeted moments related to educational attainment among those with education debt in the SCF. These moments report the distribution of debtors and student loan balances by educational attainment: dropouts, college graduates, and those not yet in repayment (i.e., enrolled in college). In the data, dropouts are over-represented among delinquent debtors, and the model matches this untargeted attribute of the data. Among all households with a positive balance of student loans, the share of dropout households is 19.53 percent in the data and 18.21 percent in the model. The share of outstanding student loan balances held by dropouts is 11.63 percent in the data and 15.04 percent in the model. Among households that are delinquent on repayment of student loans, the next set of moments, note that those enrolled in college are not relevant for this statistic—their share of this population is always zero because they have not yet entered repayment. The model rationalizes roughly a third of the share of delinquencies by dropouts observed in the data (50.54-57.31 percent). In this model, this statistic is 19.74 percent. Hence, even with over-optimism, the model somewhat understates the shares of student loans and delinquencies accounted for by dropouts. Finally, the model can also explain the share of delinquent debt held by dropouts, which is 29.18-37.13 percent in the data and 22.25 percent in the model baseline equilibrium.

Table 13: College dropouts in the population of student debtors

Population	Statistic	Group	Data	Baseline model
All debtors	Share of HH	Dropouts	19.53	18.21
		Graduates	65.35	69.88
		Enrolled or in grace period	15.74	11.90
	Share of $	Dropouts	11.63	15.04
		Graduates	74.45	72.73
		Enrolled or in grace period	14.81	12.23
Delinquent debtors	Share of HH	Dropouts	50.54-57.31	19.74
		Graduates	42.69-49.46	80.26
	Share of $	Dropouts	29.18-37.13	22.25
		Graduates	62.87-70.82	77.75

Notes: Table 13 compares the representation of college dropouts among student debtors and delinquent student debtors, for the data (SCF) and the model baseline equilibrium. The first group of statistics contains information on the educational attainment of households with a positive balance of educational debt, by share of households and by share of dollars owed. Units are percent of the population under consideration, which is specified in the first column. The second group of statistics reports the educational attainment of households that are not making payments on their educational loans because they cannot afford to do so (delinquent households), by share of households and by share of dollars owed. Units are percent of the total dollars owed for the population under consideration, which is specified in the first column.

5.2 The role of federal student loan borrowing limits

One of our main policy experiments is to study the impact of increasing the federal student loan limit in the presence of over-optimism. The impact of such a policy depends on the extent to which borrowing limits on federal student loans bind in equilibrium, and how this affects college enrollment. In this section, we review empirical evidence that current federal borrowing limits are not sufficient to finance the average total costs of college, and that college students frequently use up all of the federal loans they have access to. Next, we turn to the model to see how borrowing constraints affect the relationship between enrollment and family income in the baseline equilibrium.

Federal student loan borrowing limits in the data

To measure the extent to which students are constrained in their ability to finance college with public loans, we begin by noting that the current federal student loan limit is enough to pay for 1.49 years of average tuition, net of grants, if we include the average expenditure for college room and board.[50] A student who is attending a college that costs the average amount, and who does not access to financing other than federal loans (e.g., family transfers, other income, or housing provided by their family) will not be able to finance their entire college degree under current borrowing limits. The model is calibrated to match this attribute of current federal loan policy.[51]

Under current policy, to what extent are college students using all of the federal loans they have access to? To measure utilization rates in the data, we turn to the HSLS:09, which allows us to see the FAFSA

[50] If we exclude college room and board expenditure, the current federal student loan limit is enough to pay for 3.96 years of average tuition net of grants. See the citations used to construct this moment given in Table 9 in section .

[51] See Table 9 in section and surrounding discussion.

submission of every federal student loan recipient in the sample. With this information, we compute the federal loan utilization rate, which is the ratio of cumulative federal debt balance to total borrowing limits for three years of college.[52] The results are reported in Table 14: more than 50 percent of incoming seniors in the 2016-2017 academic year utilized more than half of their federal student loan limit, and about 1 in 3 students utilized all of the federal loans they were allowed access to.[53] This is direct empirical evidence that shows that college students are constrained with respect to the federal student loan limit. Although these moments are not targeted in the calibration, the model's baseline equilibrium also exhibits a sizable share of students hitting the federal borrowing limit, as is seen in the data. Since we underestimate this share in the model baseline, our welfare gains from loan expansions can be considered lower bounds (Table 13).

Table 14: Utilization rates for federal student loans

Utilization	Data	Baseline model
$\geq 50\%$	0.529	0.383
$\geq 100\%$	0.328	0.193
Obs	1940	

Notes: Table 14 reports utilization rates for federal student loans in the data (HSLS) and in the baseline model equilibrium. Sample: students who enrolled in a 4-year program in the Fall of 2013 and persisted to their fourth year. Utilization rates of federal student loans by end of 3rd academic year and are in fractions of the total limit. For the data moment, which is from the HSLS, the borrowing limit computed using reported dependency status and academic year from FAFSA. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

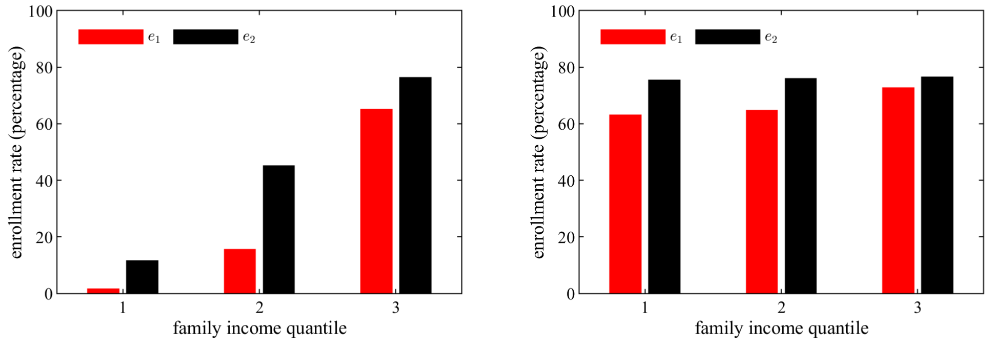
Federal student loan borrowing limits in the model's baseline equilibrium

Federal student loan borrowing limits play a sizable role in allowing the model framework to generate patterns of enrollment similar to the data. The qualitative properties of enrollment rate patterns are documented by Lochner and Monge-Naranjo (2011), who show that in the NLSY97 college enrollment rates are increasing in family income conditioning on student skill—especially for students with low skill. In Figure 2, we plot the model's college enrollment by family income quantile for students in the lower two skill quantile. Enrollment rates in the baseline equilibrium, shown in subfigure 2a, are increasing in family income for a given skill quantile, matching the patterns seen in the data.
The model framework captures this pattern from the data because it incorporates borrowing constraints and heterogeneous access to family transfers. To see the contribution of borrowing constraints, subfigure 2b shows enrollment rates in a counterfactual world where the federal student loan borrowing limit is expanded to finance four years of college costs (i.e., the entire college degree). Comparing subfigures in Figure 2, the positive relationship between family income and enrollment rates for low-skill students is substantially reduced by this policy change, indicating that a lack of access to credit significantly contributes to the observed empirical relationship between enrollment and family income for low-skill students.

The contribution of family transfers is shown in Figure 3, which plots the inter-vivos transfers received by students in the baseline equilibrium by family income and student skill. We see that conditioning on skill,

[52] Federal loans have set yearly borrowing limits; private loans do not have limits common to all holders of private loans. That is why we compute utilization rates only for federal loans.

[53] For a more detailed set of moments on federal loan utilization rates, see Table 49 in the Appendix.



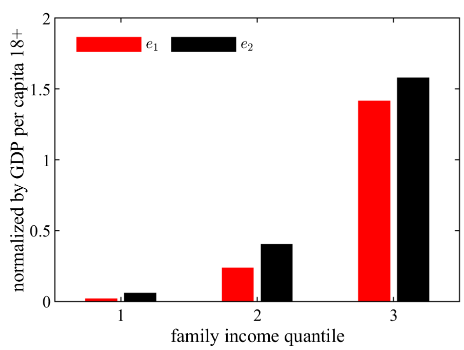
(a) Enrollment in baseline equilibrium

(b) Enrollment with federal loan limit raised to 4 years

Figure 2: College enrollment by family income and student skill quantile

Notes: Figure 2 shows enrollment rates in the baseline equilibrium and a counterfactual equilibrium. Subfigure 2a shows enrollment rates by family income quantile, for each student skill quantile e, in the model's baseline equilibrium. In the baseline, the federal student loan borrowing limit represents less than 2 years worth of college costs (tuition and fees plus room and board). Subfigure 2b shows enrollment rates in the model when the federal student loan borrowing limit is expanded to finance four years of college costs (the entire degree). The positive relationship between family income and enrollment rates for low-skill students is substantially reduced by this policy change, indicating that borrowing constraints contribute significantly to the observed empirical relationship between enrollment and family income for low-skill students.

inter-vivos transfers increase in family income. These transfers are what allows students from high income families to pay for college in the baseline equilibrium, despite the presence of borrowing constraints.



(a) Inter-vivos transfers in baseline model

Figure 3: Inter-vivos transfers family income and student skill quantile

Notes: Figure 3 shows family transfers by family income and student skill quantiles; transfers increase more strongly moving across family income bins for a given skill quantile than moving across skill quantiles within a family income bin.

6 Policy Experiments

This section presents and discusses the results of two policy experiments: an information intervention, analyzed in 6.1, and a federal student loan limit expansion, analyzed in section 6.2.

Specifically, section 6.1 reports the effects of eliminating over-optimism about college graduation probabilities by setting $\hat{p}(e) = p_g(j, e)$. This exercise is motivated by the model finding in section 5.1, which demonstrated that over-optimism leads to significant over-enrollment. This is especially true among low-skill students in the calibrated baseline equilibrium. Compared to the partial equilibrium analysis of section 5.1, in this policy experiment we incorporate general-equilibrium adjustments that occur when over-optimism is eliminated: namely, changes in prices, public transfers, the income tax, accidental bequests, and inter-vivos transfers from parents.[54] Besides illustrating the effect of an information intervention that eliminates over-optimism, this exercise also provides broad insight into the general-equilibrium impact of over-optimism about college completion in the model economy.

In section 6.2, we analyze the effects of expanding the federal student loan limit to $\bar{A} = 4$, so that federal loans are sufficient to pay for four years of college tuition plus room and board (net of grants). This expansion allows overly-optimistic high school students to access more credit if they enroll in college, potentially worsening over-enrollment and the debt burden of dropouts. At the same time, in the baseline (as in the data) many students are fully utilizing their federal student loan limits, so that expanding accessed to federal loans could increase their well-being by relaxing a binding borrowing constraint.[55] The welfare consequences of a federal student loan limit expansion is therefore ambiguous, with the model determining the relative magnitudes of each of these forces and sign of the net effect.

Note that, in both policy exercises, we assume that the economy is in its steady state in period 0. In period 1, the transition is announced unexpectedly, but there is perfect foresight thereafter.

6.1 Information intervention

First, we analyze the impact of an information intervention which eliminates over-optimism on education decisions, macro aggregates, and welfare. The effects of this policy on the model's steady-state equilibrium are shown in column (1) of Table 15. Effects on the model economy are summarized by changes in education and skill statistics (Panel A), macroeconomic aggregates (Panel B), and prices, the income tax rate, and transfers (Panel C).

Panel A indicates that an information intervention leads to lower enrollment in college, especially for low skill students. Enrollment decreases because, when students use the true probabilities of college graduation in making their college enrollment decision, the value of going to college decreases. By construction, over-enrollment in the new equilibrium goes to 0. That statistic was highest for the lowest-skilled students in the baseline equilibrium, so over-enrollment changes the most for those with the lowest skill. The elimination of over-enrollment leads to a higher graduation rate due to compositional changes: the average college student is now higher skill and therefore more likely to graduate. However, lower enrollment in college leads to fewer college graduates in the population at the new steady state. As a result, in the future generations, the

[54] Eliminating over-optimism not only eliminates over-enrollment, but also over-optimism about the child's likelihood of college graduation among parents. This affects the parent's inter-vivos transfer decision and the dynamic decisions leading up to the age in which the inter-vivos transfer is made.

[55] See section 5.2.

mass of 18-year-olds with low skill increases and the mass with high skill decreases.

Moving to Panel B, note that the drop in the mass of college graduates reduces the total efficiency units of labor, which reduces labor earnings. Lower earnings, in turn, lower both savings and aggregate capital. This reduction in factor inputs lowers output and, consequently, lowers consumption. Panel C of Table 15 indicates that the risk-free savings rate falls and the wage rate increases slightly, because aggregate labor falls more than aggregate capital. At the same time, the interest rate on private students loans decreases due to lower delinquency risk among private student loan borrowers. The information intervention's reduction of labor income, capital income, and consumption lead to lower tax revenue revenue for the government at any given tax rate, so that the income tax rate increases in the new steady state in order to balance the government's budget. Inter-vivos transfers decline significantly for three reasons: first, parents' altruistic incentive to provide inter-vivos transfers to their children decreases, because parents are no longer over-optimistic their children's ability to graduate from college; second, the economy will have less wealth and income in the new steady state; and, third, there will be fewer high-skill students, which will lower parents' incentive to pay for college (see Figure 3). The last two rows of this panel show that transfers such as accidental bequests and Social Security decrease, due to lower savings and lower labor earnings in the new steady state. The impact of the changes reported in panel C will matter for welfare, which we turn to next.

Our measure of welfare is consumption-equivalent variation. As our model includes psychic costs (i.e., search costs for student loans, an effort cost for college, and stigma costs for delinquency on student loans), we follow Abbott, Gallipoli, Meghir, and Violante (2019) and use the value of not going to college (which does not include any psychic costs) to compute consumption-equivalent variation. We compute the lifetime change in consumption required in the initial steady state, in every period and at every state, in order for an 18-year-old to be indifferent between the initial steady state value of not going to college and the lifetime value before going to college in the initial steady state, the transition path, or the final steady state. After computing the consumption-equivalent variation, we report the change in lifetime consumption relative to period 0, when the economy is at the initial steady state. Therefore, positive values indicate gains and negative values indicate losses. Furthermore, when measuring welfare, we assume that the government is paternalistic: the government knows the true payoff of choices but internalizes that the consumer is overly optimistic when making the college enrollment decision, when making the inter-vivos transfer decision, and when making the dynamic decisions leading up to the age in which the inter-vivos transfer is made. Figure 4 illustrates a welfare analysis of the information intervention in general and partial equilibrium, divided into four subfigures. Subfigure 4a shows welfare changesfor by skill quantile, and on average, over the transition path between the baseline equilibrium and the new equilibrium. In the initial periods of the transition path, students in the lowest skill quantile benefit from the information intervention. These are students who exhibit the most over-enrollment in the baseline equilibrium (Table 12). By contrast, students in the highest and the middle skill quantiles experience much smaller gains or even losses. Over time, the gains dampen or the losses amplify for students in all skill quantiles. This means that an average 18-year-old in initial periods of the transition experiences small welfare gains but incurs welfare losses in later periods.

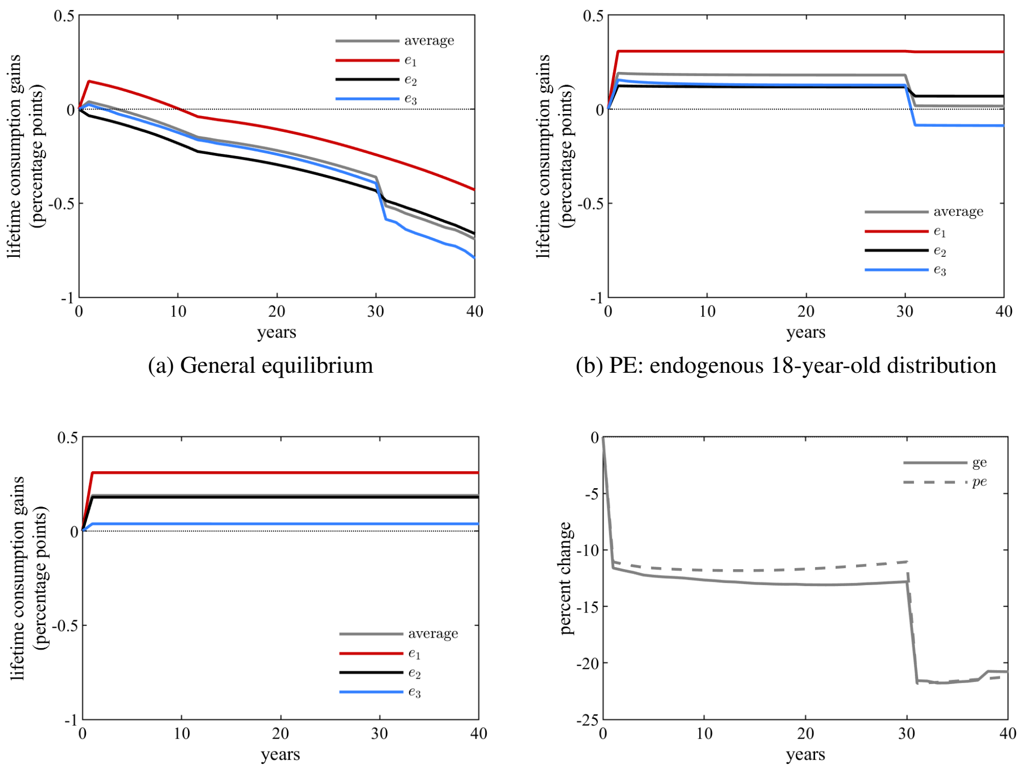
The remaining subfigures of Figure 4 serve to clarify the intuition behind the patterns of welfare changes

Table 15: Policy experiment results

Variable		Policy (1) Information intervention	Policy (2) Federal loan limit expansion
Panel A: Education and skill statistics			
Units: percentage point change			
College enrollment rate	e_1	-21.63	41.95
	e_2	-26.71	31.91
	e_3	0.00	0.00
Over-enrollment	e_1	-20.28	39.23
	e_2	-12.82	21.14
	e_3	-6.99	-6.99
Graduation rate		8.05	-4.99
Population share college graduates		-8.81	12.08
Share of 18-year-olds	e_1	2.10	-2.89
	e_2	0.19	-0.26
	e_3	-2.29	3.14
Panel B: Macroeconomic aggregates			
Units: percent change			
Output		-3.53	3.69
Capital		-3.03	1.05
Labor (efficiency units)		-3.81	5.21
Consumption		-3.50	3.72
Panel C: Prices, income tax rate, and transfers			
Units: percentage point/percent change			
Risk-free savings interest rate		-0.06	0.31
Wage rate		0.29	-1.44
Private student loan interest rate		-0.29	-
Income tax rate | Average income		0.49	-0.41
Inter-vivos transfers		-23.39	-14.24
Accidental bequests		-3.27	6.03
$\text{SS}_{\ell,e}$	e_1	-2.34	2.12
	e_2	-2.33	2.07
	e_3	-2.28	2.03
$\text{SS}_{h,e}$	e_1	-2.07	1.24
	e_2	-1.71	1.35
	e_3	-1.57	1.06

Notes: Table 15 provides results from a steady-state comparison of the baseline economy to: (1) an economy with an information intervention that eliminates over-optimism (i.e., $\hat{p}(e) = p_g(j, e)$), and (2) an economy with the federal student loan limit expanded to fund four years of college tuition plus room and board net of grants (i.e., $\bar{A} = 4$). Panels A, B, and C report changes in education and skill statistics, macroeconomic aggregates, and prices, the income tax rate, and transfers, respectively.

discussed in the previous paragraph. Subfigure 4c plots welfare changes in a partial equilibrium in which the income tax rate, prices, bequests, Social Security transfers, and the 18-year-old distribution are fixed at their initial steady state values. In this partial equilibrium, all high school students benefit in proportion to the extent of over-enrollment in their skill bin (e.g., the lowest skill quantile benefits the most). This is the direct impact of eliminating over-optimism and is not surprising: when over-optimism is eliminated, students correct their enrollment decisions and the transfers they make later in life as parents, both of which improve their well-being. In subfigure 4b, we plot welfare changes in a partial equilibrium in which the 18-year-old distribution is now endogenized (with other objects held constant). In this case, the 18-year-olds in the initial periods of the transition benefit. However, 18-year-olds in the later periods have their gains dampened or suffer losses. This is due to a change in inter-vivos transfers across generations: as subfigure 4d shows, in the later periods of the transition path inter-vivos transfers to 18-year-olds decrease. Once we



(a) General equilibrium

(b) PE: endogenous 18-year-old distribution

(c) PE: exogenous 18-year-old distribution

(d) Inter-vivos transfers

Figure 4: Information intervention welfare analysis for 18-year-olds: general and partial equilibrium effects

Notes: Figure 4 provides a welfare analysis of an information intervention for 18-year-old consumers in general and partial equilibrium. Subfigures 4a-4c report lifetime consumption gains and losses for the average 18-year-old and the average-18-year-old given skill in each period of the transition path under the following cases: (a) general equilibrium, (b) a partial equilibrium in which the income tax rate, prices, bequests, and Social Security transfers are fixed at their initial steady state values, but the 18-year-old distribution is endogenous, and (c) a partial equilibrium in which the income tax rate, prices, bequests, Social Security transfers, and the 18-year-old distribution are fixed at their initial steady state values. Subfigure 4d reports average inter-vivos transfers received by 18-year-olds in general equilibrium in each period of the transition path.

take general-equilibrium effects into account, as in subfigure 4a, the average 18-year-old's welfare gains are lowered or become losses. This is due to an increase in the income tax rate, a decrease in the savings interest rate, and a fall in accidental bequests and Social Security transfers.[56]

This policy experiment uncovers the complex role of over-optimism in the economy: although over-optimism leads to 18-year-old students to over-enroll in college, it also benefits them in two ways. First, as long as parents are also over-optimistic (as our empirical work indicates they are), this raises the inter-vivos transfers that 18-year-olds receive from their parents. Second, over-optimism raises the college education rate and lowers labor income taxes, thus benefitting 18-year-olds later via general-equilibrium effects. The results of our analysis indicate that a large scale information intervention with strong general-equilibrium effects

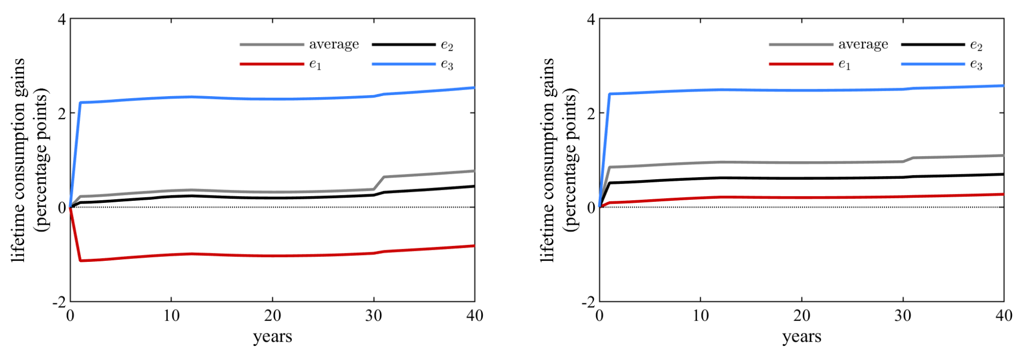
[56] In Appendix E, Table 59, we report welfare implications for consumers that are 19 and older in the period of the transition. Although the magnitudes of welfare implications are smaller, the qualitative takeaways are the same. In partial equilibrium, the cohort that is 19 and over will benefit. However, in general equilibrium, the same cohort will be hurt.

will in fact make the average person worse off when they are 18. In fact, even a small scale information intervention without general equilibrium effects could hurt 18-year-old consumers, if the parents respond to the information intervention by reducing inter-vivos transfers to a sufficient extent.

6.2 Federal loan limit expansion

In our second policy experiment, we analyze the effects of expanding the federal student loan limit to $\bar{A} = 4$, so that federal loans are sufficient to pay for four years of college tuition plus room and board (net of grants). The effects of this loan limit expansion on education and skill and macroeconomic aggregates, as well as prices, the income tax rate, and transfers, are shown in column (2) of Table 15. The mechanisms in this exercise parallel those from an information intervention, but move in the opposite direction.

Panel A of Table 15, column (2), shows that the expansion in the loan limit increases enrollment in college for the lowest two skill quantiles, which in turn raises over-enrollment for those quantiles. This happens because previously credit-constrained students can now enroll in college. Higher enrollment among low skill students leads to a lower graduation rate. However, higher enrollment also increases the share of college graduates in the population, which leads to more 18-year-olds with higher skill in the new steady state. Panel B reports the resulting increase in aggregate output, capital, labor, and consumption. In Panel C, we see that the risk-free savings interest rate rises and the wage rate drops, because efficiency units of labor rise by more than the capital stock. By construction, the private student loan market completely shuts down when students can use public loans to pay for college costs. Because of the expansion in the tax base, the income tax rate decreases. Public transfers from Social Security rise, and so do accidental bequests; however, because loans with generous borrowing terms are now sufficient for financing college, inter-vivos transfers decrease.



(a) Baseline

(b) No over-optimism

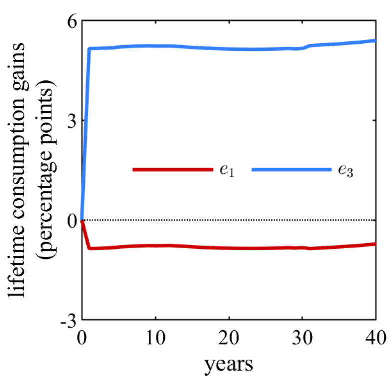
Figure 5: Federal loan limit expansion welfare analysis for 18-year-olds: baseline vs. no over-optimism

Notes: Figure 5 provides a welfare analysis of an expansion in the federal student loan limit to fund four years of college net tuition plus room and board for 18-year-old consumers in the baseline economy (5a) and an economy without over-optimism (5b). Both subfigures report lifetime consumption gains and losses for the average-18-year-old and the average-18-year-old given skill in each period of the transition path.

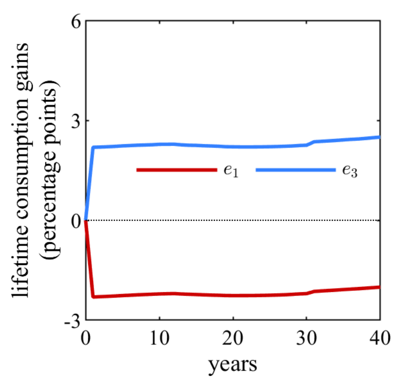
In Figure 5, we analyze the welfare implications of expanding the federal student loan limit for 18-year-old consumers in both the baseline economy (subfigure 5a) and in an economy without over-optimism (subfigure 5b). In the baseline economy with over-optimism, students in the bottom skill quantile experience welfare losses, while students in the highest skill quantile experience welfare gains. Students in the middle skill quantile benefit, albeit marginally. Overall, the average 18-year-old gains from the limit expansion.

When we perform the same experiment in an environment without over-optimism (subfigure 5b), 18-year-olds in all skill quantiles experience welfare gains. These gains are increasing in skill. This means that, without over-optimism, general-equilibrium effects and the public costs of subsidized federal student loans do not offset the welfare gains which stem from increased access to credit for financing college.

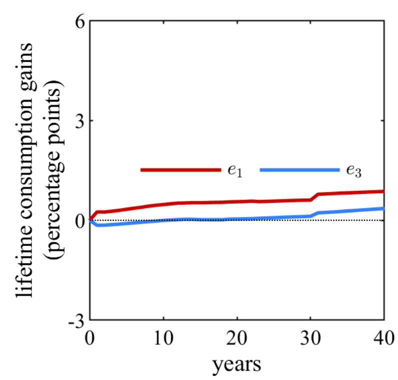
A comparison of subfigures 5a and 5b highlights the additional insight one gains from incorporating over-optimism about college graduation likelihood into the model environment. With this model ingredient, college enrollment increases for the lowest skill quantile when loan limits expand, but these students also drop out of college more often than they anticipate at enrollment (Table 15, Panel A). Low-skill students are therefore hurt by this policy when over-optimism is taken into account, in stark contrast to their outcomes in a model where expectations reflect outcomes more accurately.[57]



(a) Family income quantile 1



(b) Family income quantile 2



(c) Family income quantile 3

Figure 6: Federal loan limit expansion welfare analysis for 18-year-olds: by family income and student skill

Notes: Figure 6 provides a welfare analysis of an expansion in the federal student loan limit to fund four years of college net tuition plus room and board for 18-year-old consumers by family income quantile and skill quantile in general equilibrium. Panels 6a-6c report lifetime consumption gains and losses for the average-18-year-old in the lowest and highest skill quantiles in each period of the transition path by family income quantiles 1, 2, and 3, respectively.

The federal student aid program in the US is intended to provide funding that facilitates college attendance for students without other forms of financing, who tend to be from low-income families. In Figure 6, we plot welfare implications by family income quantile and student skill. Subfigures 6a and 6b show that students from low income families in the lowest skill quantile experience large welfare losses while students from low income families in the highest skill quantile experience large gains. Thus, among students raised by poor families, there are heterogeneous effects of the increased access to cheap credit for financing college.

[57] General equilibrium effects amplify the gains or dampen the losses slightly. See Appendix F for details.

At the same time, the impact of the loan limit expansion on students from high-income families is small. This is because students from high income families receive large inter-vivos transfers from their parents before the policy change, and so do not adjust their enrollment decision once access to credit is expanded.

The results from this policy experiment show that an expansion in the federal loan limit will not be welfare improving for all students. In particular, the impact of the loan limit expansion is felt most by students in families with low income; in this income group, low-skill students see welfare losses while high-skill students see welfare gains. This result suggests that switching to merit based student loan limits could generate large gains, a policy which we are currently investigating in ongoing research.

7 Conclusion

In this paper, we examine the optimal student loan limit in an environment where low-skill students are overly optimistic about their probability of finishing college. This over-optimism means that they are more likely to be repaying their student loan debt without enjoying the college wage premium. As a consequence, when given the opportunity, some students over-borrow, which means that low-skill students tend to benefit less from loan limit expansions than their higher-skilled peers.

Here, we focus on the student loan limit for our policy analysis, but we could examine other margins—for example, student loan forgiveness, penalties for non-repayment, or interest rates that incorporate non-completion risk. Focusing on the optimal use of other policy instruments would potentially highlight different consequences for over-optimism among young adults when they make the college enrollment decision.

References

Abbott, B., G. Gallipoli, C. Meghir, and G. Violante (2019). Education Policy and Intergenerational Transfers in Equilibrium. *Journal of Political Economy 127*(6).

Amir, E., J. Teslow, and C. Borders (2021). The MeasureOne Private Student Loan Report. https://www.measureone.com/resources#report. accessed 5-8-2022.

Caucutt, E. and L. Lochner (2020). Early and Late Human Capital Investments, Borrowing Constraints, and the Family. *Journal of Political Economy 128*(3), 1065–1147.

Chatterjee, S. and F. Ionescu (2012). Insuring student loans against the financial risk of failing to complete college. *Quantitative Economics 3*, 393–420.

Chetty, R. (2006). A New Method of Estimating Risk Aversion. *The American Economic Review 96*(5), 1821–1834.

Duprey, M., D. Pratt, D. Jewell, D. Brown, L. . Caves, S. Kinney, T. Mattox, M. Smith Ritchie, J. Rogers, C. Spagnardi, and J. Wescott (2020). High School Longitudinal Study of 2009 (HSLS:09) Postsecondary Education Transcript Study and Student Financial Aid Records Collection. Found at: https://nces.ed.gov/pubs2020/2020004.pdf (accessed: 05.02.2022).

Dynarski, S., C. Libassi, K. Michelmore, and S. Owen (2021). Closing the Gap: The Effect of Reducing Complexity and Uncertainty in College Pricing on the Choices of Low-Income Students. *American Economic Review 111*(6), 1721–1756.

Dynarski, S. and J. E. Scott-Clayton (2008). Complexity and Targeting in Federal Student Aid: A Quantitative Analysis. *NBER Working Paper 13801*.

Elsby, M. W., B. Hobijn, and A. Sahin (2013). The Decline of the U.S. Labor Share. *Brookings Papers on Economic Activity*.

Exler, F., I. Livshits, J. MacGee, and M. Tertilt (2021). Consumer credit with over-optimistic borrowers. *Unpublished manuscript*.

Federal Reserve Board and the Unites States Department of the Treasury (2019). Survey of Consumer Finances.

Federal Student Aid (posted 9.29.2021). National Default Rate Briefing for FY 2018 Official Cohort Default Rates. Found at: https://fsapartners.ed.gov/sites/default/files/2021-09/FY18OfficialNationalBriefing.pdf (accessed: 05.15.2022).

Grossman, G. and E. Oberfield (2021). The Elusive Explanation for the Declining Labor Share. *NBER Working Paper 29165*.

Heathcote, J., K. Storesletten, and G. L. Violante (2017). Optimal Tax Progressivity: An Analytical Framework. *Quarterly Journal of Economics 132*, 1693–1754.

Ionescu, F. (2009). The Federal Student Loan Program: Quantitative Implications for College Enrollment and Default Rates. *Review of Economic Dynamics 12(1)*, 205–231.

Ionescu, F. (2011). Risky Human Capital and Alternative Bankruptcy Regimes for Student Loans. *Journal of Human Capital 5(2)*, 153–206.

Ionescu, F. and N. Simpson (2016). Default Risk and Private Student Loans: Implications for Higher Education Policies. *Journal of Economic Dynamics and Control 64*, 119–147.

Jones, C. (2016). Facts About Economic Growth. *Handbook of Macroeconomics 2A*.

Krueger, D. and A. Ludwig (2016). On the optimal provision of social insurance: Progressive taxation versus education subsidies in general equilibrium. *Journal of Monetary Economics 77*, 72–98.

Lochner, L. J. and A. Monge-Naranjo (2011). The Nature of Credit Constraints and Human Capital. *The American Economic Review*, 2487–2529.

Luo, M. and S. Mongey (2019). On the optimal provision of social insurance: Progressive taxation versus education subsidies in general equilibrium. pp. Unpublished manuscript.

Nakajima, M. (2012). Rising indebtedness and temptation: A welfare analysis. *Quantitative Economics 3(2)*, 257–288.

Nakajima, M. (2017). Assessing bankruptcy reform in a model with temptation and equilibrium default. *Journal of Public Economics 145*, 42–64.

National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education (2007). High School Longitudinal Study of 2009 (restricted use license).

Smole, D. P. (2019). Federal Student Loans Made Through the William D. Ford Federal Direct Loan Program: Terms and Conditions for Borrowers. (Congressional Research Service of the United States R45931).

Social Security Administration (2020). The 2020 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds. Found at: https://www.ssa.gov/oact/TR/2021/tr2021.pdf (accessed 5.15.2022).

Stinebrickner, T. and R. Stinebrickner (2012). Learning about academic ability and the college dropout decision. *Journal of Labor Economics 30(4)*, 707–748.

U.S. Bureau of Economic Analysis (posted 4.28.2022a). Table 1.1.5 Gross Domestic Product. Found at: https://apps.bea.gov/iTable/iTable.cfm?reqid=19&step=3&isuri=1&select_all_years=0&nipa_table_list=5&series=a&first_year=2016&last_year=2018&scale=-9&categories=survey&thetable= (accessed: 05.15.2022).

U.S. Bureau of Economic Analysis (posted 4.28.2022b). Table 2.1 Personal Income and Its Disposition. Found at: https://apps.bea.gov/iTable/iTable.cfm?reqid=19&step=3&isuri=1&select_all_years=0&nipa_

table_list=58&series=a&first_year=2016&last_year=2018&scale=-9&categories=survey&thetable= (accessed: 05.15.2022).

U.S. Bureau of Economic Analysis (posted 4.28.2022c). Table 3.1 Gross Domestic Product. Found at: https://apps.bea.gov/iTable/iTable.cfm?reqid=19&step=3&isuri=1&select_all_years=0&nipa_table_list=86&series=a&first_year=2018&last_year=2016&scale=-9&categories=survey&thetable= (accessed: 05.15.2022).

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics (2019). Table 330.10. Found at: https://nces.ed.gov/programs/digest/d19/tables/dt19_330.10.asp (accessed: 01.04.2022).

U.S. Department of Education, Office of Federal Student Aid (2022). Understand how interest is calculated and what fees are associated with your federal student loan. Found at: https://studentaid.gov/understand-aid/types/loans/interest-rates (accessed: 03.14.2022).

Yannelis, C. (2020). Strategic default on student loans. *Unpublished manuscript*.

Appendix

Supplemental material for the paper "Over-optimism About Graduation and College Financial Aid"

by Emily G. Moschini, Gajendran Raveendranathan, and Ming Xu

Not intended for publication – to be made available online

Contents

Part I

Data Appendix

A The 1997 National Longitudinal Survey of Youth

The 1997 National Longitudinal Survey of Youth, referred to as the NLSY97, is a nationally representative sample of people born between 1980 and 1984 who lived in the United States in 1997. This survey collected data annually from 1997 to 2011 and biannually from 2011 to the present. The latest round of data collection is round 19 (2019-2020). We use the NLSY97 to examine college enrollment outcomes by age 30, as well as expectations about educational attainment for NLSY97 focal respondents as reported by them and by their parents. Besides these stylized facts used to motivate the paper, we also discipline the role of skill in the lifetime earnings process using variation in wage outcomes for a given education level by skill quantile from the NLSY97. In that exercise, as in the HSLS:09, focal individual's skill is measured by high school GPA.

This section proceeds as follows. Subsection A.1 addresses focal respondent bachelor's degree outcomes as well as respondent and parent beliefs about those outcomes. Subsection A.2 reports our results showing a lack of learning about college graduation likelihood. The earnings process estimation, which also uses the NLSY97, is described in section B.

A.1 Student enrollment in 4-year programs: expectations and outcomes

Table 16 reports statistics on enrollment in a 4-year postsecondary program by age 30, an outcome which can then be compared with beliefs. These statistics are computed within the pool of high school graduates. Note that checking enrollment at age 25 captures most of the enrollment realized by age 30, although the biggest change in those 5 years occurs in the lowest tercile of skill (high school GPA). Enrollment rates in 4-year postsecondary programs are increasing skill.

Table 16: Enrollment in BA by age 25 and 30

	Obs(Q)	By Age 25	By Age 30
1	798	22.81	28.45
2	821	44.09	48.72
3	723	76.49	78.15
Obs	2342		

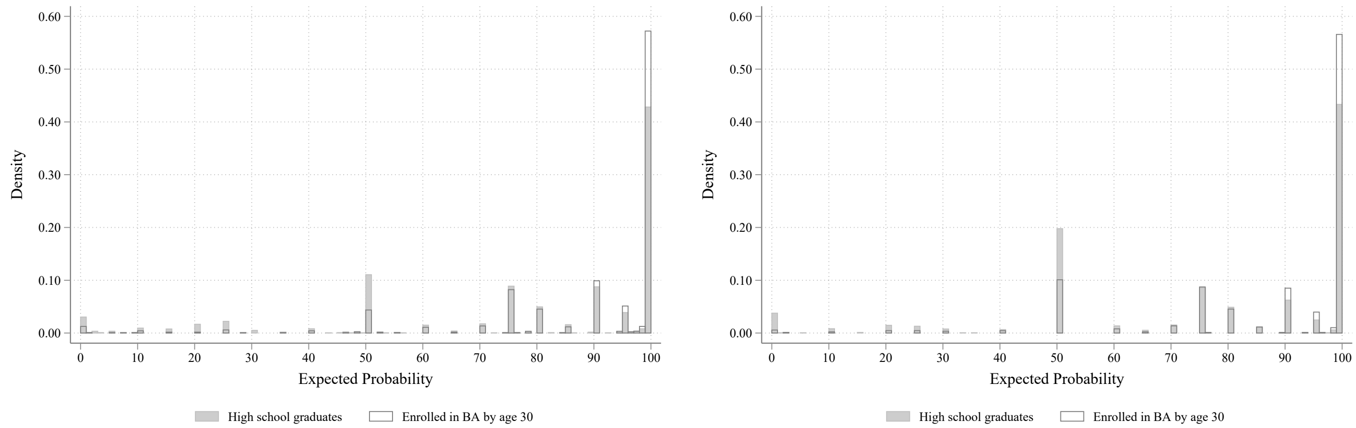
Sample: high school graduates. Source: NLSY97.

Figure 7 shows a histogram of reported expected probabilities from the focal respondent (panel 7a) and their parent (panel 7b), in regards to the respondents probability of earning a BA by age 30. In each of these panels, the distribution of beliefs is reported both for the population of high school graduates and for those who enroll in college before age 30.

Figure 7: Distribution of Expectations About BA Attainment Probability

(a) Respondent response frequencies: probability of earning BA by 30

(b) Parent response frequencies: probability of earning BA by 30



Notes: Figure 7a plots the distribution of student responses about college graduation probability for the sample of high school gradutes and the sample of those who enroll in a 4-year program before age 30. Figure 7b plots the parent response to the question about their child for the same samples.

Parent beliefs, when compared with realized outcomes in Table 17, exhibit overconfidence. While the expected probability of the parent's child earning a BA by age 30 does increase in the child's skill, the relationship is much weaker than for the realized outcomes by skill reported in Table 16. Thus, overconfidence among parents is higher for respondents with lower levels of skill.

Table 17: Overconfidence about student BA attainment: parent beliefs

	Obs(Q)	Exp Pr Earned BA by 30	Overconfidence
1	168	80.50	47.90
2	301	84.45	27.70
3	413	92.92	13.09
Obs	882		

Sample: parents of 4-year college enrollees. Source: NLSY97.

Finally, parent expectations about their child's educational attainment are compared with the child's realized outcomes in Table 18 (high school graduates) and Table 19 (college enrollees). These tables indicate that for both samples, parents' stated likelihood of graduating from college for their child (the respondent) closely align with the respondent's own stated beliefs, because the median difference in their stated probabilities is 0 and the mean difference in responses is at most about 3 percentage points.

Table 18: Comparison student and parent beliefs (student - parent): high school graduates

	Obs(Q)	mean	p10	p50	p90
1	580	2.14	-40	0	50
2	591	3.06	-30	0	40
3	528	0.36	-20	0	24
Obs	1699				

Sample: parents of high school graduates and their children. Source: NLSY97.

Table 19: Comparison student and parent beliefs (student - parent): BA enrollees

	Obs(Q)	mean	p10	p50	p90
1	168	0.42	-40	0	40
2	301	2.26	-25	0	30
3	413	0.40	-15	0	20
Obs	882				

Sample: parents of 4-year college enrollees and their children. Source: NLSY97.

A.1.1 Summary statistics

Tables 20 and 21 show summary statistics for the samples used in the NLSY97 expectations versus outcomes tabulations.

Table 20: Overconfidence about BA attainment: sample summary statistics

	Cleaned sample		With student beliefs		Enrolled in BA by 30	
	count	mean	count	mean	count	mean
HS GPA	4588	2.90	2342	2.89	1192	3.12
Age in 1997	4588	14.41	2342	15.26	1192	15.21
Pr BA by 30 (parent)	1731	76.70	1699	76.84	882	87.66
Pr BA by 30 (student)	2342	78.52	2342	78.52	1192	88.97
Enr BA by 25	4588	48.54	2342	46.84	1192	92.03
Enr BA by 30	4588	52.33	2342	50.90	1192	100.00
BA by 30	4588	33.63	2342	32.11	1192	63.09
Obs	4588		2342		1192	

Notes: Notes: summary statistics for cleaned sample, sample additionally imposing student beliefs, and sample also additionally imposing enrollment by age 30. Source: NLSY97.

Table 21: Overconfidence about BA attainment: sample summary statistics student vs parent

	Enrolled in BA by 30			
	Student beliefs		Parent beliefs	
	count	mean	count	mean
HS GPA	1192	3.12	882	3.11
Age in 1997	1192	15.21	882	15.75
Pr. BA by 30 (parent)	882	87.66	882	87.66
Pr. BA by 30 (student)	1192	88.97	882	88.70
Enr BA by 25	1192	92.03	882	91.38
Enr BA by 30	1192	100.00	882	100.00
BA by 30	1192	63.09	882	62.93
Obs	1192		882	

Notes: Notes: summary statistics for sample with student beliefs, and also with parent beliefs observed. Source: NLSY97.

A.2 Learning about graduation likelihood

Table 2 in the main text shows that overconfidence about college graduation likelihood does not wane upon college enrollment. The summary statistics for this estimation sample are shown in Table 22. The sample for this question is quite restricted because it requires that the student be enrolled in high school in 1997 and in college in 2001, in addition to having valid responses to the beliefs questions in each round.

A.2.1 Summary statistics

Table 22: Learning and overconfidence about BA attainment: sample summary statistics

	Learning Sample	
	count	mean
HS GPA	272	3.23
Age in 1997	316	15.83
Pr. BA by 30 (parent)	276	90.16
Pr. BA by 30 (student)	316	92.07
Enr BA by 25	316	100.00
Enr BA by 30	316	100.00
BA by 30	316	69.62
Obs	316	

Notes: Notes: summary statistics for sample where the student was in high school in 1997 and had enrolled in a 4-year program by 2001. Source: NLSY97.

B Earnings process estimation

The earnings process we use in our structural model, $\epsilon_{j,s,e}\nu_j$ has a deterministic component and a stochastic component. The deterministic component depends on the consumer's age, j, their educational attainment, s, and their skill endowment, e,:

$$\epsilon_{j,s,e} \quad = \quad \exp\left(\beta^A_{s,1}j + \beta^A_{s,2}j^2 + \beta^A_{s,3}j^3 + \beta^e_{s,e}\right)$$

The stochastic component is an AR(1) process

$$\begin{aligned} \nu_j &= \rho_s \nu_{j-1} + \eta_{s,j} \\ \eta_{s,j} &\sim \mathbb{N}(0, \sigma_{s,\eta}) \end{aligned}$$

To estimate the total earnings process for each education category (with and without a college degree) we proceed as follows. First, we estimate how log wages depend on a third order polynomial of age for a given education group, $s = \ell$ (HS or some college) or $s = h$ (BA or higher), using the Panel Study of Income Dynamics (PSID). This identifies $\beta^A_{s,1}$, $\beta^A_{s,2}$, and $\beta^A_{s,3}$ for each education group s. Next, we take log wages in the National Longitudinal Study of Youth 1997 (NLSY97), clean them of age effects with the PSID estimation results, and regress age-free log wages on indicators for skill terciles. The coefficients on these terciles are the factor loadings on skill e for a given education s, $\beta^e_{s,e}$. Finally, we use the residuals from the NLSY97 regression and jointly estimate $\rho_s$and $\sigma_{s,\eta}$ for each education group.

B.1 Estimating age profiles in the PSID

We clean the PSID data before estimation in the following way. The PSID collects data on the household head and resident spouse, if there is one. We use information at the individual level on the educational attainment of the household head and resident spouse (if any), as well as their sex, total income, total income from transfers, total labor earnings, labor component of business income,[58] hours worked, marital status (a flag equal to 1 if married with spouse present, 0 if not) and employment situation (which is used to identify the self-employed). Using this information, we construct unearned income as total income net of earnings and transfers. We construct hourly wages by dividing the individual's labor earnings (plus labor component of business income when necessary) by total hours worked for the individual. We correct all income and wage variables for inflation using the CPI and use real values in all of our analysis. Since the PSID data has a different variable per year per individual for all of these measures, we reshape the data to be at the individual-year level, with one variable per measure.

We exclude observations from the SEO census sample, which leaves us with 1,429,096 individual-year observations with 34,856 unique individuals. Next, we drop observations without a valid state of residence value, or with marital status or sex of the household head missing. This leaves us with 34,854 individuals. We count the number of times an individual is observed in the PSID sample, and drop observations observed less than 8 times. This leaves us with 32,016 individuals in the sample. Next, we drop observations with a yearly wage growth of more than 4 percent or less than -2 percent, or where the level of real wages exceeds 400. This does not eliminate any individuals from the sample, while the total number of individual-year observations remaining is 1,030,945. We define those with a high school education as individuals with between 12 and 15 years of schooling; those with a college education are individuals with 16 years of schooling or more. We restrict the sample to those 65 and younger who are greater than 17 if they have a high school degree, greater than 19 if they have some college, and greater than 22 if they have a BA or more.

[58] This is not included in labor earnings for some years of the PSID. For years when it is not included, we add it to reported labor earnings.

This leaves us with 17,215 individuals. Dropping those who are self employed reduces the sample further to 16,692 individuals. Conditioning on having valid observations for all of the regrssion controls described below, our sample has 9,204 individuals for the high school education category and 7,249 individuals in the college education category.

Using this sample, we proceed in two stages for observations within each education category and in that age range. In the first stage, we regress an indicator for working positive hours on a set of standard controls (age polynomial, an indicator for being married, a set of dummies for the year, and a set of dummies for the state of residence) for those with a given educational attainment who are not self-employed. In addition to these standard controls, X (where Xincludes a constant), in the first stage we also control for Z, which is unearned income after correcting this variable for inflation. This regression can be written as:

$$\mathbb{I}_{hrs>0} \quad = \quad \gamma_{s,Z}Z + \alpha_s X + \epsilon$$

where ϵ is the residual. This regression is estimated using a probit estimator and the result is used to construct an inverse mills ratio, which is included in a second stage regression that has all of the same controls but with unearned income replaced with the estimated mills ratio, IM, from the first stage. In this second stage regression, the dependent variable is the log of the real wage, w. This regression estimated on a given education group can be written as:

$$w \quad = \quad \gamma_{s,IM}IM + \beta^A_{s,0} + \beta^A_{s,1}age + \beta^A_{s,2}age^2 + \beta^A_{s,3}age^3 + \gamma_s \times [i.state + i.year + i.married] + u$$

where u is the i.i.d. residual. The coefficients of interest for estimating the age profile of education category s are $\beta_{s,1}$, $\beta_{s,2}$, and $\beta_{s,3}$. Since the average offerred wages which is rejected is likely lower than the average offer wage that is accepted, the expected sign of the inverse mills ratio coefficient in the second stage, $\gamma_{s,IM}$, is positive. In our estimation, this coefficient has the expected sign for both education groups.

Table 23 presents the results from implementing this regression in the PSID.

Table 23: Wages as a function of age

	(1) HS or Some College	(2) BA
$\beta^A_{s,1}$	0.114*** (0.0117)	0.181*** (0.0233)
$\beta^A_{s,2}$	-0.00199*** (0.000339)	-0.00306*** (0.000614)
$\beta^A_{s,3}$	0.0000109*** (0.00000319)	0.0000163** (0.00000522)
R^2	0.109	0.167
Obs	86109	65089

Notes: The effect of age on earnings by education level $s = \ell$ (HS or Some College) and $s = h$ (BA). Not shown in table but included as controls: state and year FE, married indicator, inverse mills ratio.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Data source: PSID.

B.2 Estimating skill loadings in the NLSY97

We clean the NLSY97 data before estimation in the following way. We begin with 8,984 individual observations (170,696 individual-year observations). We condition on observing high school GPA, which leaves us with 6,004 individual observations. Conditioning on observing a valid wage and educational attainment reduces the sample further to 5,115 individuals. Next, we restrict to observations who have completed high school (4,930 individuals). We drop obsrvations with wages in dollar units above 400 and below 1 (4,600 observations remain) and we also drop those with wage group above 4 percent and below -2 percent and those with either some high school (which we define as a high school degree or some college - GED recipients and those with an associate's degree are not included) or a BA or more (which we define as college graduates or those with a graduate degree) leaves 4,429 individuals. We include those currently enrolled in a graduate program but not those currently enrolled in a BA or currently enrolled in high school in these education categories. We restrict ages to be above 24 and below 39. This leaves 3,903 individuals: 2,345 with a high school degree and 1,656 with a BA or more. Since the NLSY97 records information about individuals using separate variables for each year, we reshape the data so that it isa panel with a single variable recorded across years. We estimate the factor loadings on skill using these remaining observations in the panel data.

Using the estimated age contributions to log wages, we take log wages in the NLSY97 and, using the observation's associated age, clean log wages of their age component. These age-free log wages, w_{AF}, are then regressed on dummies for high school GPA terciles, a set of controls X which includes indicators for the year, a set of indicators for the number of children (top-coded at 4), an indicator for being married, and a control for being in the supplemental sample for the NLSY. The age restriction on the NLSY97 sample is ages 25 and above. Standard errors in this regression are clustered at the individual level. The estimation equation can be written as:

$$w_{AF} \quad = \quad \beta^e_{s,0} + \beta^e_{s,e} \times i.\left[GPA_Q = e\right] + \chi X + u$$

where u is the i.i.d. residual.

Table 24 presents the skill loadings estimated in the NLSY97; Tables 25 and 26 show summary statistics of the estimation sample used in Table 24.

Table 24: Age-free wages as a function of skill

	(1) HS or Some College	(2) BA
$\beta^e_{s,e=1}$	-0.0433	-0.196***
	(0.0245)	(0.0380)
$\beta^e_{s,e=2}$	-0.0278	-0.148***
	(0.0253)	(0.0269)
r2	0.0408	0.0576
N	14771	8499

Notes: Baseline skill quantile is the top quantile ($e = 3$). Not shown in table but included as controls: year FE, number of children (topcoded at 4) indicators, married indicator.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Data Source: NLSY97.

Table 25: Earnings process estimation: sample summary statistics at individual-year level

	Cleaned sample	Regression sample	HS or some college only	BA+
Real wage	11.20	12.84	10.74	16.50
Age	26.85	29.98	29.81	30.29
HS GPA	2.88	2.92	2.72	3.28
No. children	0.74	0.88	0.99	0.69
Married	0.31	0.41	0.36	0.49
BA +	0.28	0.37		
HS or some college	0.72	0.63		
Meets age restriction	0.62			
Obs	37435	23270	14771	8499

Notes: summary statistics highlighting effects of restrictions on sample composition for skill loading regression. The level of observation in this table is the individual-year, so being in one estimation sample precludes being in the other. Source: NLSY97.

Table 26: Earnings process estimation: sample summary statistics at individual level

	Regression sample		HS or Some College		BA+	
	count	mean	count	mean	count	mean
HS GPA	4222	2.91	2734	2.72	1620	3.24
Age in 1997	4222	14.42	2734	14.45	1620	14.38
Pr. BA by 30 (parent)	1599	76.72	1052	69.54	609	89.69
Pr. BA by 30 (student)	2162	78.66	1438	72.21	803	90.77
Enr BA by 25	4222	49.74	2734	24.51	1620	92.28
Enr BA by 30	4222	53.39	2734	29.19	1620	96.17
BA by 30	4222	34.96	2734	2.01	1620	90.68
BA + regression sample	4222	0.38				
HS only regression sample	4222	0.65				
Obs	4222		2734		1620	

Notes: summary statistics highlighting effects of restrictions on sample composition for skill loading regression. Being in one education category regression does not exclude being in the other over the lifecycle. The level of observation in this table is that of individual; the regression level of observation is at the individual-year level. Source: NLSY97.

B.3 Estimating the stochastic component of earnings

After estimating the skill loadings in the NLSY97, we use residuals implied by the regression results and estimate a shock process for each education category using this output. To do this, given a guess of parameters, we construct a variance-covariance matrix between lags of the residual component, and compare it with an analogous matrix constructed on the empirical residuals. We iterate on the parameter guess until the two matrices converge. In our estimation, we use 500 bootstraps.

We find that the stochastic component of the earnings process is more persistent for those with more education. The higher this persistence coefficient, the closer the stochastic process is to a random walk, and the harder it is for individuals to self-insure. The fact that self-insurance is harder for those with more education in our results means that higher educational attainment does not insulate workers from risk. Finally, in our estimation the random shock $\eta_{s,j}$ has a similar variance across the two education groups, although it is slightly higher for the $s = h$ (BA+) group. Specific point estimates are reported in Table 27.

B.4 Summary of results

The results of the age profile from the PSID, the loadings on skill from the NLSY97, and the residual process estimation are presented in Table 27.

Table 27: Earnings process estimation results

Parameter	Description	Value
Lifecycle productivity: high school or some college ($\epsilon_{j,s=\ell,e}$)		
$\rho_{\eta\|n}$	Persistence AR(1)	0.854980
$\sigma^2_{\eta\|n}$	Variance AR(1)	0.082005
$\beta^A_{\ell,1}$	Age third-order polynomial	0.114
$\beta^A_{\ell,2}$		-0.00199
$\beta^A_{\ell,3}$		0.0000109
$\beta^e_{\ell,1}$	Skill endowment shifter	-0.0433
$\beta^e_{\ell,2}$		-0.0278
Lifecycle productivity: college graduate ($\epsilon_{j,s=h,e}$)		
$\rho_{\eta\|c}$	Persistence AR(1)	0.871885
$\sigma^2_{\eta\|c}$	Variance AR(1)	0.083397
$\beta^A_{h,1}$	Age third-order polynomial	0.181
$\beta^A_{h,2}$		-0.00306
$\beta^A_{h,3}$		0.0000163
$\beta^e_{h,1}$	Skill endowment shifter	-0.196
$\beta^e_{h,2}$		-0.148

Notes: This Table summarizes results from the earnings process estimation. Data sources are the PSID and the NLSY97.

C The High School Longitudinal Study of 2009

The High School Longitudinal Study of 2009 (HSLS:09) is a representative panel of ninth-grade students in the United States beginning in 2009 who attended high schools that had both ninth and eleventh grades (HSLS (2009)). The survey collection occurs over several waves, with the most recent wave being the collection of post-secondary transcripts and student records up to and including in the 2016-2017 academic year (HSLS (2020)).[59]

As the outline indicates, the focal sample member is referred to as "Student" while they are in high school and as "Sample member" during the 2013 update, because they are between educational programs. Regardless of the focal individual's educational status after the base year, the HSLS:09 makes an effort to collect data from them. Thus the second follow-up in 2016 includes information from students who are currently enrolled in post-secondary education as well as those who are not and used to be enrolled and those who did not pursue post-secondary education. Another benefit of this survey is that the survey includes those who began, but did not complete, post-secondary education. If sample members does begin a 4-year degree program in the Fall after high school graduation (the Fall of 2013), and does not take time off from

[59]Note: this transcript data collection only generates observations for those enrolled in post-secondary education.

school, then students complete a survey after their third year of college. This information includes their high school GPA as well as financial aid and private loans they took out to pay for the post-secondary education they did complete. Such financial aid information in the student questionnaire is supplemented by information collected from institutions themselves in the post-secondary transcripts and student records data collection wave, implemented after the second follow-up. Table 28 contains an outline of the structure of the HSLS:09.[60]

Table 28: Structure of the HSLS

Calendar Year(s) Academic Year (if enrolled) Questionnaire	Base Year 2009 (Fall) 1st year HS (Fall)	1^{st} Follow-up 2012 (Spring) 3rd year HS (Spring)	2013 Update 2013 (Summer) Graduated HS	HS Transcripts 2013-2014 Graduated HS	2^{nd} Follow-up 2016 3rd year PS (Spring)	PS Transcripts + SR 2017-2018 4th year PS (Fall)
Student	X	X				
Parent	X	X				
Student/Parent			X			
Sample Member					X	
Counselor [1]	X	X				
Aministrator [2]	X	X				
Teacher [3]	X					
Institution Attended				X		X

Notes: Table 28 describes the survey structure of the HSLS:09. In this table, HS stands for "high school" and PS stands for "post-secondary". Calendar years and academic years are distinguished because academic years overlap two calendar years and the semester of data collection is indicated for the academic year except for the 2013 Update and HS transcript collection when this is not relevant. The focal sample member is referred to as "Student" while they are in high school and as "Sample member" during the 2013 update, because they are between educational programs. Regardless of the focal individual's educational status after the base year, the HSLS:09 makes an effort to collect data from them. This ensures that the second follow-up in 2016 includes information from students who are currently enrolled in post-secondary education as well as those who are not.

[1] Lead counselor at student's high school.
[2] Administrator (principal) at student's high school.
[3] Math or science teacher at student's high school.

This section proceeds as follows. Subsection C.2 tabulates student expected educational attainment and their measured outcomes (given the restrictions of the HSLS:09 design); subsection C.3 reports HSLS:09 statistics used in model specification, calibration, and validation. Subsection C.4 presents additional tabulations, besides those in the main text, on the student aid (loans and grants) and private loan uptake, as measured in the HSLS:09. Finally, subsection C.5 presents summary statistics for the various subsamples used in the tabulations from the HSLS:09. Note that, in all of the tabulations, the data moments are weighted using survey weights and the sample counts are rounded to the nearest 10 as per NCES requirements (HSLS (2020)). The specific survey weights used vary and are noted in the Table captions.

C.1 Enrollment rates in 4-year programs

Table 29 shows enrollment rates in 4-year programs in the Fall of 2013 or any time until the 2016-2017 academic year.

[60] Questionnaires are available here: https://nces.ed.gov/surveys/hsls09/questionnaires.asp.

Table 29: Enrollment rates and parent education by skill

	Obs(Q)	% Enr Fall 2013	% Ever Enrolled by 2016-17 AY	% Parent High Ed
1	2980	6.67	12.02	19.73
2	3550	26.38	35.88	34.56
3	4320	48.17	58.70	54.41
Obs	10850			

Notes: Sample: high school graduates. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to the nearest 10 per NCES requirements. Weights are 2013 Update longitudinal weights.

C.2 Expected educational attainment versus outcomes

In the second wave of data collection, which occurs in what would be the students' junior year of high school, the HSLS:09 asks students what their expected educational attainment is.[61] This survey subsequently also records not only high school educational outcomes but also the post-secondary educational outcome of the focal student: whether they enrolled in a 4-year program after high school, whether they took out student loans to pay for this program, and whether they persisted in their program after enrollment. With this information, we examine the relationship between student skill (high school honors-weighted GPA) and educational outcomes (both expected and realized). In the process we establish that low-students are more likely to over-estimate their future educational attainment, and that individuals who do not persist in college have sizable student debt balances three years after high school graduation (as addressed in the main text). This evidence suggests that mistaken expectations are related to the debt burden faced by students who do not complete the educational program the loan was taken out to pay for.

Unfortunately, the phrasing of the question in the HSLS:09 on BA attainment does not permit straightforwardly refutable, unlike the phrasing of the similar question in the NLSY97. Because of this, in the main text we rely on our preferred evidence for this point from the NLSY97 to establish overconfidence about BA attainment. Specifically, for this question, the HSLS:09 asks "As things stand now, how far in school do you think you will actually get [in your education]?" of students.[62] The possible answers to this question are reported in Table 30. A valid response is one in which students select a code between 1 and 13 ("Don't Know" is a valid response). To flag those who expect to enroll in a 4-year program, a flag is set to 0 for all observations with a valid response to this educational attainment question, and the flag is replaced with a 1 if the response value x was such that $7 \leq x < 13$. To flag those who expect to *persist and complete* the 4-year program, a new flag is created, which is set to 0 for valid responses and replaced with a 1 if the response x is such that $8 \leq x < 13$. A flag for those who expect to graduate from high school or enroll in a masters is constructed the same way.

[61] Note that students who drop out of high school are still surveyed.

[62] The survey also asks the same question about the focal student of the student's parent.

Table 30: Possible responses to expected educational attainment question in the HSLS

Value	Response
1	Less than high school completion
2	Complete a high school diploma, GED or alternative high school credential
3	Start, but not complete a certificate or diploma from a school that provides occupational training
4	Complete a certificate or diploma from a school that provides occupational training
5	Start, but not complete an Associate's degree
6	Complete an Associate's degree
7	Start, but not complete a Bachelor's degree
8	Complete a Bachelor's degree
9	Start, but not complete a Master's degree
10	Complete a Master's degree
11	Start, but not complete a Ph.D., M.D., law degree, or other high level professional degree
12	Complete a Ph.D., M.D., law degree, or other high level professional degree
13	You don't know

Table 31 shows educational attainment expectations and outcomes (using our preferred measure of outcomes for the HSLS:09), for the sample of students who enrolled in a 4-year program in the Fall of 2013, right after high school graduation. This Table presents, by high school GPA quantile, the percent of each skill bin that expected to complete college and the percent of the bin that persisted in college into their fourth year. Note that, since in the HSLS:09 we observe the student up to four years after high school graduation, we cannot definitively say if they permanently drop out of college or fail to ever enroll. For this reason, we use terms like "persistence" and "non-persistence" when discussing findings from the HSLS:09, as opposed to more definitive terms like "dropping out" and "graduating", respectively, which are terms we favor in when describing our model framework. Table 31 demonstrates that students at all level of skill tend to overestimate their educational attainment, given their skill. This is especially the case for those in the first (lowest) skill quantile

Table 31: College degree: outcomes vs. expectations

	Obs(Q)	% Persisted	% Expect BA
1	190	38.81	76.33
2	800	56.30	82.22
3	1680	78.70	92.27
Obs	2670		

Notes: Sample: students who enrolled in a 4-year program in the Fall of 2013. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to the nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

One concern with the findings reported in Table 31 is that perhaps restricting to only those who enroll in a 4-year program right after college (in the Fall of 2013) restricts to a distorted sample of students. Perhaps these students are the most overconfident and those who take a gap year (or several) are more realistic. In Table 32 we present the results of a robustness exercise where we broaden the sample to include those who enroll in a 4-year program at any point between the 2013-2014 academic year and the 2016-2017 academic year. Persistence in this context means continued enrollment once enrolled, for however long that interval is. So, someone who enrolls in the 2016-2017 academic year automatically counts as a student who persists. This is a looser criterion for persistence to compare with expected educational attainment, compared to Table 31. Even with this relaxed criterion, although persistence rates do increase, students still tend to overestimate their persistence on average given their level of skill.

Table 32: College degree: outcomes vs. expectations

	Obs(Q)	% Persisted	% Expect BA
1	290	54.38	72.25
2	1030	60.00	81.62
3	1950	79.98	91.33
Obs	3270		

Notes: Sample: students who ever enrolled in a 4-year program between 13-14 AY and 16-17 AY. Not persisting means leaving the program after enrolling. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Weights are PETS-SR student records longitudinal weights.

Another concern is that when students answer that they expect a BA (or a higher degree) they are simply responding in this way because admitting they probably will stop after high school is embarrassing (that is, it generates a stigma cost). To try to control for this, in Table 33 we show a tabulation restring to those who expect a master's (MA) degree or higher. It seems less likely that stating you expect to get an MA is driven by fear of stigma costs. Note that, by implication, in this group everyone expects to get a BA. The tabulation demonstrates that the fraction who persist in each quantile still remains well below the expected graduation rate from college.

Table 33: College degree: outcomes vs. expectations

	Obs(Q)	% Persisted	% Expect BA
1	70	37.83	100.00
2	390	57.33	100.00
3	1110	79.47	100.00
Obs	1580		

Notes: Sample: students who enrolled in a 4-year program in the Fall of 2013 who expected to enroll in MA+. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to the nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

Finally, what do parents think of their child's likelihood of graduating from a 4-year degree? In Table 34 we tabulated the parent responses to what they expect their child's educational attainment will be, with the response choice set shown in Table 30.[63] Table 34 shows that, for the sample of students who enrolled in a BA program in the Fall of 2013, most parents were not surprised (they expected their children to do so). Nevertheless, parents in this group tend to overestimate the likelihood of college graduation for their children, and this is especially true when their child belongs in a lower skill quantile.

[63]The sample size of families with responses to this questionnaire is much smaller than the sample of valid student responses because the parent questionnaire was only administered to a random sample of 48% of families in the sample.

Table 34: Parents: college degree: outcomes vs. expectations

	Obs(Q)	% Expected enrollment	% Persisted	% Expected BA
1	80	73.43	38.81	71.20
2	370	95.91	56.30	94.62
3	790	94.59	78.70	94.38
Obs	1240			

Notes: Sample: parents of students who enrolled in a 4-year program in the Fall of 2013, where the parents answered the question about child's expected educational attainment. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to the nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

C.2.1 Summary statistics

This section presents summary statistics for the samples used to generate Tables 31, 32, and 33. Table 34 is the sample for Table 31 conditioning on observing valid parent responses to the expected educational attainment question. Variables in the summary statistics tables are divided into two groups for each sample: binary (0 or 1) variables where only the mean reported, and continuous support variables where the mean and beneath it the median in parentheses are reported. In the tables below, Table 35 and Table 36 show summary statistics for Tables 31 and 32. Tables 37 and 38 show summary statistics for Table 33.

Table 35: Educational attainment expectations and outcomes

	4yr Enrolled Fall 2013	4yr Enrolled Fall 2013 + Expect MA
Student Exp. HS	0.938	1.000
HS Degree or Equivalent	1.000	1.000
Student Exp College Enr.	0.884	1.000
Enr 4yr Fall 2013	1.000	1.000
Ever Enr 4yr by Summer 2016	1.000	1.000
Student Exp College Pers.	0.876	1.000
Persisted to 2016-2017 AY	0.679	0.709
Std Pers (Ever Enr 4yr)	0.679	0.709
PETS: bachelor's deg received as of June 2016	0.031	0.036
Obs	2670	1580

Notes: for each variable in each sample, fraction of total. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to nearest 10 per NCES requirements. Sample size rounded to the nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

Table 36: Demographics

	4yr Enrolled Fall 2013	4yr Enrolled Fall 2013 + Expect MA
Household income	106,750	111,787
	(85,000)	(87,500)
HS GPA (Honors Wgtd)	3.18	3.28
	(3.26)	(3.35)
Female	0.53	0.55
	(-)	(-)
At least 1 Parent BA+	0.56	0.60
	(-)	(-)
Fed SL Balance Y1	3,425	3,315
	(3,500)	(3,500)
Prv SL Balance Y1	821	786
	(0)	(0)
T+F (1st Inst)	17,253	17,695
	(11,127)	(11,373)
Obs	2670	1580

Notes: for each variable in each sample, medians in parentheses below means. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to nearest 10 per NCES requirements. Sample size rounded to the nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

Table 37: Educational attainment expectations and outcomes

	4yr Ever Enrolled
Student Exp. HS	0.932
HS Degree or Equivalent	1.000
Student Exp College Enr.	0.866
Enr 4yr Fall 2013	0.798
Ever Enr 4yr by Summer 2016	1.000
Student Exp College Pers.	0.858
Persisted to 2016-2017 AY	0.679
Std Pers (Ever Enr 4yr)	0.703
PETS: bachelor's deg received as of June 2016	0.026
Obs	3270

Notes: for each variable in each sample, fraction of total. Sample: ever enrolled in a 4-year program between 13-14 AY and 16-17 AY. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

Table 38: Demographics

	4yr Ever enrolled
Household income	103,603
	(82,500)
HS GPA (Honors Wgtd)	3.11
	(3.20)
Female	0.52
	(-)
At least 1 Parent BA+	0.54
	(-)
Fed SL Balance Y1	3,182
	(3,175)
Prv SL Balance Y1	786
	(0)
T+F (1st Inst)	15,501
	(10,002)
Obs	3270

Notes: for each variable in each sample, medians in parentheses below means. Sample: ever enrolled in a 4-year program between 13-14 AY and 16-17 AY. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

C.3 Calibration targets and model primitives

This section contains tabulations reporting HSLS:09 moments used for calibration targets in the main text. Table 39 contains the conditional distribution of child skill given parent education. Table 40 contains enrollment rates and the fraction of students with parents where at least one resident parent has a BA or higher. In this Table, enrollment rates are defined as enrolling in a 4-year program in the Fall of 2013, the academic year after graduation. This definition of enrollment is more strict than the measure used in the main text (before age 25) and measured in the NLSY97. Table 41 shows that student skill is positively and significantly correlated with the income of the household that raised them (their family), and that this correlation is positive and significant even after selection into a 4-year degree.

C.3.1 Moments using sample of high school graduates

Table 39: Distribution over HS GPA quantile given parent education

e_c	$\pi\left(e_c \mid s=\ell\right)$	$\pi\left(e_c \mid s=h\right)$
1	42.43	18.53
2	34.32	32.19
3	23.25	49.28
Obs	5870	4980

Notes: this Table shows the conditional distribution of student skill in percentage points, given parental education. Parent education is "low" (i.e., $s = \ell$) if neither resident parent has a BA or higher. If either parent has a BA or higher, the parent education for that household is set to h. Sample: high school graduates. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to the nearest 10 per NCES requirements. Weights are 2013 update longitudinal weights.

Table 40: Enrollment rates and parent education by skill

	Obs(Q)	% Enrolled Fall 2013	% with parent high education
1	2980	6.67	19.73
2	3550	26.38	34.56
3	4320	48.17	54.41
Obs	10850		

Notes: Sample: high school graduates. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to the nearest 10 per NCES requirements. Weights are 2013 update longitudinal weights.

Table 41: Correlation of student skill and student's household income while in high school

	High school graduates High school GPA	Fall 2013 enrollees High school GPA
Household income	0.129^{***}	0.0630^{**}
Obs	10850	2670

Notes: Sample: high school graduates (left), and students who enrolled in a 4-year program in the Fall of 2013 (right). Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to the nearest 10 per NCES requirements. Weights are 2013 Update longitudinal weights (left) and PETS-SR student records longitudinal weights (right). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

C.3.2 Moments using sample of 2013 college enrollees with various persistence requirements

Table 42 shows the fraction who persist to their second year of college, by skill quantile, for students who enroll in college in the Fall of 2013. Table 43 shows the fraction who persist to their fourth academic year of college, by skill quantile, for students who enroll in college in the Fall of 2013 and also persist to their second year of college. Table 44 shows the average hours worked for all students enrolled in their third year of college, given that they began college in the Fall of 2013.

Table 42: Persistence to second year of college

	Obs(Q)	Fraction persisted to Y2
1	190	0.577
2	800	0.743
3	1680	0.913
Obs	2670	

Notes: Sample: students who enrolled in a 4-year program in the Fall of 2013. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to the nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

Table 43: 2013 Enrollees: persistence to 2016-2017 AY (Y4) given persistence to 2014-2015 AY (Y2)

	Obs(Q)	Fraction Persisted to Y4
1	120	0.672
2	630	0.757
3	1560	0.862
Obs	2310	

Notes: Sample: students who enrolled in a 4-year program in the Fall of 2013 and persisted to Y2. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to the nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

Table 44: Average time spent working junior year (fraction of full-time)

	Hrs: Enr	(Hrs: Enr)/40
mean	12.760	0.319
Obs	2090	

Notes: Sample: students who enrolled in a 4-year program in the Fall of 2013 who were enrolled their third year as well. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to the nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

C.4 Student loans and grant aid

This subsection presents information from the HSLS:09 on various attributes of student debt's patterns in the cohort studied by that survey, to supplement the discussion in the main text.

Using the distribution of high school GPA observed within the sample of high school graduates, we then assign students to one of three quantiles of skill.[64] Next, for those who enroll in a 4-year postsecondary program in the Fall of 2013, we tabulate the extensive and intensive margin of student loan uptake for those who persist in the program until their 4th academic year, and for those who do not persist. Persisting means maintaining enrollment in a 4-year program for each academic year after their first year (the 2013-2014 academic year). The results are reported in Tables 45 and 46.[65]

To begin, Table 45 reports the extensive margin of student loan usage among HSLS:09sample members who enrolled in a 4-year program in the Fall of 2013. This sample criterion does not condition on persistence, so we can observe student loan uptake rates for persisters and non-persisters in the sample. Here, student loan uptake means having at least one dollar of student debt either from federal loans or from private loans. Table 45 also reports the share of students who persist (or do not), for each skill quantile. The findings reported in Table 45 findings indicate that non-persisters exhibit loan uptake rates similar to those of persisters, and that for the whole sample taking out student loans to pay for college is very common. Not persisting in college

[64] For summary statistics of the samples analyzed in this section, see Tables 55 and 56 in Appendix C.5. The quantiles are assigned using cross-sectional 2013 Update survey weights.

[65] Note that all tabulations using HSLS:09 data reported in this document use survey weights. Specific weights used are noted in Table captions.

is also very common, especially for those with low levels of skill.[66]

Table 45: Extensive margin of student loan uptake

	Obs(Q)	Persisted		Did not persist	
		% of Total	% with SL	% of Total	% with SL
1	190	38.81	89.00	61.19	80.98
2	800	56.30	74.42	43.70	76.63
3	1680	78.70	59.36	21.30	67.54
Obs	2670				

Notes: Sample: students who enrolled in a 4-year program in the Fall of 2013. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to the nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

Table 46 shows the average amount of student loan balances, conditional on having any federal or privates student loan debt, by skill quantile (rows) and persistence status (columns) as well as overall. These debt balances are measured in the second follow-up wave, which is collected in 2016 and measures the cumulative debt after at most 3 years of post-secondary enrollment. Evidently, those who persist have more debt than those who do not persist, which is because they have paid for more years of tuition with loans by the summer of 2016.[67]

Table 46: Intensive margin of student loan uptake

	Obs(Q)	All	Persisted	Did not persist
1	190	17,509	25,795	11,826
2	800	22,923	27,568	17,056
3	1680	21,684	22,288	19,767
Obs	2670			

Notes: average student loan balance, conditional on having a positive balance, by skill quantile. Sample: students who enrolled in a 4-year program in the Fall of 2013. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to the nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

To better map the intensive margin of student loan uptake from Table 46 into the model for the purpose of comparison, Table C.4 shows the unconditional uptake rates for student loans and the intensive margin moments normalized by per capita GDP. The intensive margin is humped shaped in skill, while the extensive margin is decreasing in skill.

[66] Table C.4 in the Appendix shows extensive and intensive margins of student loan uptake for the pooled population of 2013 enrollees, without conditioning on persistence status as of the 2016-2017 academic year.

[67] These figures pool federal and private student loan debt. Federal student loans in the United States have yearly borrowing limits, which depend on the student's dependency status as reported on the Free Application for Federal Student Aid (FAFSA) form (see the post 2012 panel of Table C.1 in CRS (2019) for borrowing limits). Table 49 in the Appendix shows utilization rates of federal student loans by the end of the third year of college for students who persist.

Table 47: Intensive margin of student loan uptake: normalization

	Obs(Q)	Any SL	SL balance	$\frac{\text{SL balance}}{\text{GDP p.c.}}$ (%)
1	190	84.06	17,509	23.35
2	800	75.38	22,923	30.57
3	1680	61.13	21,684	28.91
Obs	2670			

Notes: Sample: students who enrolled in a 4-year program in the Fall of 2013. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to the nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

In contrast to the uptake pattern of loans reported in , which are not most common among students with the highest skill, Table 48 shows that grant aid is more generous relative to tuition for those with high skill in the HSLS:09.

Table 48: Ratio of total grant aid to total tuition

	Obs(Q)	$\frac{\text{Agg merit + need grants}}{\text{Agg tuition + fees}}$
1	190	0.408
2	800	0.474
3	1680	0.514
Obs	2670	

Notes: for each skill quantile (row) this table shows the ratio of dollars of merit and need grants received to dollars of tuition and fees paid within that quantile. Sample: students who enrolled in a 4-year program in the Fall of 2013. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to the nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

Note that federal loans in the United States have both yearly and aggregate borrowing limits, which vary by the dependency status of the student aid applicant.[68] This dependency status is reported in the Free Application for Federal Student Aid (FAFSA) form whose submission is a necessary condition to receive aid. We are interested in the degree to which college students are borrowing-constrained when it comes to federal student loans. To compute statistics on utilization rates which are relevant to examine such questions, we capitalize on the FAFSA filing information reported by the HSLS:09for survey respondents to flag those who file as a dependent in each academic year. Then, we construct the sum of the yearly borrowing limits for the first three years of college incorporating the information on dependency status. We then take the ratio of the federal student loan balance at the end of the student's third year (collected in the second follow-up wave of data collection) and divide the cumulative debt balance by the borrowing limit. The utilization rates that result are reported in Table 49, for all filers and restricting attention to those whose most recent FAFSA filing was as a dependent. For either group, a large fraction of students have borrowed up to the limit for their first three years of college.

[68] Independent FAFSA filers have higher borrowing limits, both per year and overall, than dependent filers.

Table 49: Utilization rates for federal student loans

	All	Dependents
$\geq 50\%$	0.529	0.529
$\geq 75\%$	0.444	0.445
$\geq 90\%$	0.378	0.377
$\geq 95\%$	0.341	0.338
$\geq 99\%$	0.332	0.329
$\geq 100\%$	0.328	0.325
Obs	1940	1870

Notes: Sample: students who enrolled in a 4-year program in the Fall of 2013 and persisted to their 4th year. Utilization rates are of federal student loans by end of their third academic year. Borrowing limits are computed using reported dependency status and academic year from FAFSA filings. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

Most students file a FAFSA, and most students who file a FAFSA are filing as dependents. This is documented using FAFSA filing information from the HSLS:09 in Table 50.

Table 50: Share of students who are FAFSA filers + share of FAFSA filers who are dependents

	FAFSA filer | enrolled	Dependent| FAFSA filer
13-14 AY	0.797	0.992
14-15 AY	0.739	0.987
15-16 AY	0.676	0.982
16-17 AY	0.689	0.948
Obs	1940	1550

Notes: Sample: students who enrolled in a 4-year program in the Fall of 2013 and persisted to their 4th year. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

To supplement the discussion in the main text, next we break down grant aid receipt by type of grant and persistence status. Table 51 reports grant aid incidence and type by persistence status. Many students in either persistence status group receive aid, both need and merit. It is quite common to receive only need or only merit based grants, perhaps because student skill is positively correlated with household income for the student so that qualifying for one type of grant makes the student less likely to qualify for the other.[69]

[69] See Table 41 in the Appendix for the correlation coefficient between household (e.g., parent) income and student high school GPA in the HSLS:09.

Table 51: Grant aid incidence by persistence status

	Obs(Q)	% need only	% merit only	% both	% either
Did Not Persist	730	17	32	26	76
Persisted	1940	21	28	24	75
Obs	2670				

Notes: this table reports the rate of grant aid receipt by the type of grant (need or merit), by persistence status. Percentages are rounded to the nearest percentage point. Sample: students who enrolled in a 4-year program in the Fall of 2013. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to the nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

Along the intensive margin of grant aid, Table 52 shows that the dollar amounts received by non-persisters are quite sizable. The fact that, like loan aid, grant aid to non-persisting students amounts to nearly 1 in 4 dollars of aid means that these resources have been allocated to students who did not progress in their program and whose earnings likely will not reflect the college wage premium.

Table 52: Grant aid dollar amount by type and share of aggregate grants received by persistence status

	Obs(Q)	Total	Merit	Need	% of agg
Did Not Persist	730	9,573	7,302	6,980	23
Persisted	1940	12,630	8,856	10,106	77
Obs	2670				

Notes: this table reports the amoung of grant aid received by the type of grant (need or merit), by persistence status. Percentages are rounded to the nearest percentage point. Sample: students who enrolled in a 4-year program in the Fall of 2013. Averages are conditional on having a positive amount of aid. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to the nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

C.5 Summary statistics for moments used in main text

This section reports summary statistics for the samples used to compute the HSLS:09 findings reported in the main text.

C.5.1 High school graduates

Table 53 shows demographic variables for the sample of high school students and those who have graduated high school by the 2013 Update data collection. Comparing these two samples shows that they are very similar in composition: high school completion rates are quite high, so little selection occurs on this margin. This is reinforced by Table 54 which shows expectations and several measures of outcomes for multiple levels of educational attainment.

Table 53: Demographics

	HS Students	HS Grads
Household income	79,108	81,594
	(57,000)	(60,000)
HS GPA (Honors Wgtd)	2.61	2.69
	(2.66)	(2.72)
Female	0.49	0.50
	(-)	(-)
At least 1 Parent BA+	0.35	0.36
	(-)	(-)
Obs	11330	10850

Notes: for each variable in each sample, medians in parentheses below means. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to nearest 10 per NCES requirements. Weights are 2013 Update longitudinal weights.

Table 54: Educational attainment expectations and outcomes

	HS Students	HS Graduates
Student Exp. HS	0.900	0.903
HS Degree or Equivalent	0.944	1.000
Student Exp College Enr.	0.659	0.684
Enr 4yr Fall 2013	0.253	0.268
Ever Enr 4yr by Summer 2016	0.333	0.353
Student Exp College Pers.	0.646	0.672
Persisted to 2016-2017 AY	0.679	0.679
Std Pers (Ever Enr 4yr)	0.705	0.705
PETS: bachelor's deg received as of June 2016	0.017	0.017
Obs	11330	10850

Notes: for each variable in each sample, fraction of total. Samples: see column heading. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to nearest 10 per NCES requirements. Weights are 2013 update longitudinal weights.

C.5.2 Fall of 2013 college enrollees

Table 55 shows demographic variables for the sample of students who enroll in a 4-year program in the Fall of 2013 (column 1), those who additionally persist to their second academic year (column 2) and those who persist to their fourth academic year (column 3). Table 56 shows expectations and several measures of outcomes for multiple levels of educational attainment for these same three samples. Comparing these samples shows that they differ along margins of household income and parental education (Table 55) and that most college students expect to complete the program although fewer than 70 percent persist to their second year (Table 56.

Table 55: Demographics

	4yr Enr 2013	4yr Enr 2013 + Persisted 14-15	4yr Enr 2013 + Persisted 16-17
Household income	106,750	112,775	116,512
	(85,000)	(90,000)	(90,000)
HS GPA (Honors Wgtd)	3.18	3.25	3.28
	(3.26)	(3.34)	(3.36)
Female	0.53	0.53	0.53
	(-)	(-)	(-)
At least 1 Parent BA+	0.56	0.60	0.63
	(-)	(-)	(-)
Fed SL Balance Y1	3,425	3,454	3,308
	(3,500)	(3,500)	(3,500)
Prv SL Balance Y1	821	857	769
	(0)	(0)	(0)
T+F (1st Inst)	17,253		
	(11,127)		
Obs	2670	2310	1940

Notes: for each variable in each sample, medians in parentheses below means. Samples: see column heading. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to nearest 10 per NCES requirements. Weights are PETS-SR student records longitudinal weights.

Table 56: Educational attainment expectations and outcomes

	4yr Enr 2013	4yr Enr 2013 + Persisted 14-15	4yr Enr 2013 + Persisted 16-17
Student Exp. HS	0.938	0.945	0.946
HS Degree or Equivalent	1.000	1.000	1.000
Student Exp College Enr.	0.884	0.899	0.902
Enr 4yr Fall 2013	1.000	1.000	1.000
Ever Enr 4yr by Summer 2016	1.000	1.000	1.000
Student Exp College Pers.	0.876	0.893	0.898
Persisted to 2016-2017 AY	0.679	0.820	1.000
Std Pers (Ever Enr 4yr)	0.679	0.820	1.000
PETS: bachelor's deg received as of June 2016	0.031	0.037	0.045
Obs	2670	2310	1940

Notes: for each variable in each sample, fraction of total. Samples: see column heading. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Restricted-Use Data File. Sample size rounded to nearest 10 per NCES requirements.Weights are PETS-SR student records longitudinal weights.

D SCF

The Survey of Consumer Finances (SCF) is a triennial survey sponsored by the Federal Reserve Board and the US Department of the Treasury (SCF (2019)). We use the 2019 wave of the SCF.

D.1 Variables used

Table 57 contains variable codes corresponding to conceptual categories discussed in the text. Each individual loan is assigned a flag (non-completion, non-repayment) and then the portfolio of loans is assigned to a category (non-completion, non-repayment) according to a threshold rule. The preferred specification we use for this threshold rule is described in the text.

Table 57: 2019 Survey of Consumer Finances: variables used

Variable	Description
Survey	
Weight	X42001
Income + Components	
Total HH Income (manual)	(A)+(B)+(C)+(D)
Wages and Salaries (A)	X5702
Self Emp. Income (B)	X5704
Capital Income (C)	X5706 + X5708 + X5710 + X5712 + X5714
Other (D)	X5724 + X6558 + X6566 + X6574 + max(0,X6464) + max(0,X6469)+ max(0,X6474) + max(0,X6479) + max(0,X6965) + max(0,X6971) + max(0,X6977) + max(0,X6983)
Wealth: Education Loans	
Current balance	X7824 +X7847 +X7870 +X7924 +X7947 +X7970 +X7179
Loan for: $\mathbb{I}_{\text{resp,sp}}$	if=1 or 2: X7978, X7883, X7888, X7893, X7898, X7993
Flag fed loan	if=1: X7879, X7884, X7889, X7894, X7899, X7994
Flag non-fed (private) loan	if=5: X7879, X7884, X7889, X7894, X7899, X7994
Flag late pymt can't afford	if =3: X7821, X7844, X7867, X7921, X7944 , X7967
Flag for did not complete program loan was taken out to pay for	if =5: X7881, X7886, X7891, X7896, X7901, X7996

D.2 Interest rates on federal and private education loans

In this section, we use the 2019 SCF to document that there is little variation by type of loan in the interest rates charged on education loans (federal and private loan panels, column 3 in Table 58). Within a type of loan, we show that interest rates also exhibit little variation across income quantiles or educational attainment (columns 4-6 and 7-10, respectively). This is true lookig within a loan type and across income quantiles, or across loan types within an income quantile. Conditional on being in delinquency, student loan interest rates are slightly higher for college dropouts than for college graduates, but the difference is not significant (columns 11 and 12).

Overall, the findings presented in Table 58 motivate our specification of the private student loan market, in which the interest rate on private student loans solves a pooling equilibrium.

Table 58: Interest rates of education loans in the 2019 Survey of Consumer Finances

			By income quantile				By education status (all)			By education status (delinquent)	
Loan type	Statistic	All	1	2	3	4	Graduates	Dropouts	Not in repayment	Graduates	Dropouts
Federal	Mean	5.97	5.95	6.08	6.26	5.45	5.92	6.25	5.69	6.02	6.80
	Median	5.50	5.50	5.32	5.96	5.00	5.50	5.60	5.20	6.00	6.00
	SD	3.19	3.19	3.25	3.55	2.44	3.26	3.33	2.55	2.58	3.59
Private	Mean	5.85	5.65	5.95	6.78	5.23	5.86	6.07	6.03	6.18	6.90
	Median	5.84	6.00	4.85	6.38	5.00	5.84	5.40	6.00	6.70	6.00
	SD	2.62	1.48	3.24	2.91	2.18	2.50	3.18	3.55	2.35	2.88

Notes: Table 58 presents interest rates of federal (top panel) and private (bottom panel) student loans from the 2019 SCF.

Part II

Model Appendix

D.3 Value functions

The over-optimistic value of college for $j = 4$ is given by

$$\hat{V}(j,h,e,\eta,a,x) = \max_{\hat{c},\hat{a}',\hat{x}'} U(c,j,h) \; - \xi 1_{a\geq 0 \text{ and } x=0 \text{ and } (\hat{a}'<0 \text{ or } \hat{x}'>0)} - \xi^{pr} 1_{x=0,\hat{x}'>0} \tag{23}$$

$$+ \beta\psi_j \Big[\hat{p}(e) E_{\eta'|h,\eta} V(j+1,h,e,\eta',\hat{a}',\hat{x}') \; + (1-\hat{p}(e)) E_{\eta'|\ell,\eta} V(j+1,\ell,e,\eta',\hat{a}',\hat{x}')\Big]$$

$s.t.$

$$(1+\tau_c)\hat{c} + \hat{a}' + (1-\theta(e)-\theta^{pr}(e))\kappa = y_{j,h,e,\eta,a} + a + Tr_j - T(y_{j,h,e,\eta,a}) + (\hat{x}'-x)$$

$$\hat{c} \geq 0$$

$$\hat{a}' \geq -\frac{\bar{A}[(1-\theta(e)-\theta^{pr}(e))\kappa + \bar{c}]j}{4}$$

$$\hat{a}' \leq a \;\; \text{if} \;\;\; a \leq 0$$

$$\hat{x}' - x \in \left[0, [(1-\theta(e)-\theta^{pr}(e))\kappa + \bar{c}] - \left(\frac{\bar{A}[(1-\theta(e)-\theta^{pr}(e))\kappa + \bar{c}]j}{4} - \big(\max[-a',0] - \max[-a,0]\big)\right)\right].$$

The value of college for $j = 4$ with the true probability of graduation is given by

$$V(j,h,e,\eta,a,x) = \max_{c,a',x'} U(c,j,h) \; - \xi 1_{a\geq 0 \text{ and } x=0 \text{ and } (a'<0 \text{ or } x'>0)} - \xi^{pr} 1_{x=0,x'>0} \tag{24}$$

$$+ \beta\psi_j \Big[p_g(j,e) E_{\eta'|h,\eta} V(j+1,h,e,\eta',a',x') + (1-p_g(j,e)) E_{\eta'|\ell,\eta} V(j+1,\ell,e,\eta',a',x')\Big]$$

$s.t.$

$$(1+\tau_c)c + a' + (1-\theta(e)-\theta^{pr}(e))\kappa = y_{j,h,e,\eta,a} + a + Tr_j - T(y_{j,h,e,\eta,a}) + (x'-x)$$

$$c \geq 0$$

$$a' \geq -\frac{\bar{A}[(1-\theta(e)-\theta^{pr}(e))\kappa + \bar{c}]j}{4}$$

$$a' \leq a \;\; \text{if} \;\;\; a \leq 0$$

$$x' - x \in \left[0, [(1-\theta(e)-\theta^{pr}(e))\kappa + \bar{c}] - \left(\frac{\bar{A}[(1-\theta(e)-\theta^{pr}(e))\kappa + \bar{c}]j}{4} - \big(\max[-a',0] - \max[-a,0]\big)\right)\right].$$

The only difference between 23 and 24 is the use of the over-optimistic and the true probability of graduation, respectively.

When $j = j_f + j_a$ and the agent chooses delinquency, we assume those agents cannot make an familial inter-vivos transfer to be consistent with our assumption that agents cannot save until they have paid off their student loans. Therefore, the value functions for delinquency presented below are the same as in equations 14, 15, and 16 with the difference that the parent has altruistic utility towards their child with an inter-vivos transfer $b = 0$.

The value of delinquency only on federal student loans for $j = j_f + j_a$ is given by

$$\begin{aligned}
&V^{D_f}(j, s, e, \eta, a, x, e_c) = U(c, j, s) - \xi_D + \beta\psi_j E_{\eta'|s,\eta} V(j+1, s, e, \eta', a', x') \\
&+ \beta_c E_{\eta'|l} W(e_c, \eta', b) \\
&s.t. \\
&(1+\tau_c)c = y_{j,s,e,\eta,a} + Tr_j - T(y_{j,s,e,\eta,a}) - \rho_D(j, a, y_{j,s,e,\eta,a}) - \rho_R^{pr}(j, x) \\
&a' = (1 + r_{SL})a + \rho_D(j, a, y_{j,s,e,\eta,a}) - \phi_D[\rho_R(j, a) - \rho_D(j, a, y_{j,s,e,\eta,a})] \\
&x' = x(1 + r_{SL}^{pr}) - \rho_R^{pr}(j, x). \\
&b = 0.
\end{aligned} \tag{25}$$

The value of delinquency only on private student loans only for $j = j_f + j_a$ is given by

$$\begin{aligned}
&V^{D_x}(j, s, e, \eta, a, x, e_c) = U(c, j, s) - \xi_D^{pr} + \beta\psi_j E_{\eta'|s,\eta} V(j+1, s, e, \eta', a', x') \\
&+ \beta_c E_{\eta'|l} W(e_c, \eta', b) \\
&s.t. \\
&(1+\tau_c)c + a' = y_{j,s,e,\eta,a} + a + 1_{\{a<0\}} r_{SL} a + Tr_j - T(y_{j,s,e,\eta,a}) - \rho_D^{pr}(j, x, y_{j,s,e,\eta,a}) \\
&a' = 1_{a<0}(1 + r_{SL})a + \rho_R(j, a) \\
&x' = (1 + r_{SL}^{pr})x - \rho_D^{pr}(j, x, y_{j,s,e,\eta,a}) + \phi_D \max[\rho_R^{pr}(j, x) - \rho_D^{pr}(j, x, y_{j,s,e,\eta,a})] \\
&b = 0.
\end{aligned} \tag{26}$$

The value of delinquency on both the federal and private student loans for $j = j_f + j_a$ is given by

$$\begin{aligned}
&V^{D}(j, s, e, \eta, a, e, x, e_c) = U(c, j, s) - \xi_D - \xi_D^{pr} + \beta\psi_j E_{\eta'|s,\eta} V(j+1, s, e, \eta', a', x') \\
&+ \beta_c E_{\eta'|l} W(e_c, \eta', b) \\
&s.t. \\
&(1+\tau_c)c = y_{j,s,e,\eta,a} + Tr_j - T(y_{j,s,e,\eta,a}) - \rho_D(j, a, y_{j,h,e,\eta,a}) - \rho_D^{pr}(j, x, y_{j,s,e,\eta,a}) \\
&a' = (1 + r_{SL})a + \rho_D(j, a, y_{j,h,e,\eta,a}) - \phi[\rho_R(j, a) - \rho_D(j, a, y_{j,h,e,\eta,a})] \\
&x' = (1 + r_{SL}^{pr})x - \rho_D^{pr}(j, x, y_{j,s,e,\eta,a}) + \phi_D[\rho_R^{pr}(j, x) - \rho_D^{pr}(j, x, y_{j,s,e,\eta,a})] \\
&b = 0.
\end{aligned} \tag{27}$$

D.4 Equilibrium definition

To define the equilibrium, we must first discuss more notation and define the social security function. Let $\overrightarrow{\omega}$ denote the idiosyncratic state of an agent. This state depends on age and enrollment status in the following way:

$$\overrightarrow{\omega} = \begin{cases} (e, \eta, a) & \text{for 18-year-olds, before making the college entrance decision} \\ (j, h, e, \eta, a, x) & \text{for agents in college} \\ (j, s, e, \eta, a, x) & \text{for agents not enrolled, dropouts, or graduates, unless } j = j_f + j_a \\ (j, s, e, \eta, a, x, e_c) & \text{if } j = j_f + j_a \end{cases} \tag{28}$$

Private loan interest rate determination: $r^{pr}_{SL,t}$ is such that the lender makes zero expected profits in pooling equilibrium for each cohort of 18-year-old-consumers:

$$\left[\sum_{i=1}^{4}(\beta)^{i-1}\int(x'_{t+i-1}(\overrightarrow{\omega}) - x)\Omega_{t+i-1}d(\overrightarrow{\omega}|j=i)\right] = \sum_{i=5}^{J}(\beta)^{i-1}\int\Big[(1 - d_{x,t+i-1}(\overrightarrow{\omega}))\rho^{pr}_{R}(j,x) \tag{29}$$

$$+ d_{x,t+i-1}(\overrightarrow{\omega})\left[\rho^{pr}_{D}(x, y_{j,s,e,\eta,a}) - \phi_D[\rho^{pr}_{R}(j,x) - \rho^{pr}_{D}(x, y_{j,s,e,\eta,a})]\right]\Big]\Omega_{t+i-1}d(\overrightarrow{\omega}|j=i),$$

where β is the lender's discount factor.

The social security function is given by

$$ss_{s,e} = \chi\left[\frac{.5\int w\eta\epsilon_{j,s,e}[1_{s=h}z + 1_{s=\ell}]\Omega_t d(\overrightarrow{\omega}|18 \le j < j_r, s)}{\int \Omega_t d(\overrightarrow{\omega}|18 \le j < j_r, s)} + \tag{30}$$

$$\frac{.5\int w\eta\epsilon_{j,s,e}[1_{s=h}z + 1_{s=\ell}]\Omega_t d(\overrightarrow{\omega}|18 \le j < j_r)}{\int \Omega_t d(\overrightarrow{\omega}|18 \le j < j_r)}\right],$$

where it replaces χ fraction of the average of the average labor earnings for the 30 years before retirement conditional on education and high school GPA and the average unconditional labor earnings for the 30 years before retirement.

Definition Given an initial level of capital stock K_0 and initial distribution over idiosyncratic states $\Omega_0(\overrightarrow{\omega})$, a competitive equilibrium consists of sequences of household value functions $\{W_t(\overrightarrow{\omega}), V_t(\overrightarrow{\omega}), \hat{V}_t(\overrightarrow{\omega}), V^R_t(\overrightarrow{\omega}), V^D_t(\overrightarrow{\omega})\}$, household college entrance and dropout policy functions $\{\hat{d}_{s,t}(\overrightarrow{\omega}), \hat{d}_{d,t}(\overrightarrow{\omega}), d_{d,t}(\overrightarrow{\omega})\}$, household consumption and next period asset policy functions $\{\hat{c}_t(\overrightarrow{\omega}), \hat{a}'_t(\overrightarrow{\omega}), c_t(\overrightarrow{\omega}), a'_t(\overrightarrow{\omega})\}$, household default policy function $\{d_{f,t}(\overrightarrow{\omega})\}$, household inter vivos transfer policy function $\{b_t(\overrightarrow{\omega})\}$, production plans $\{Y_t, K_t, L_t\}$, sequence of federal student loan policies $\{\bar{A}_t, r_{SL}, \bar{y}, \tau_g, T_{SL}\}$, sequence of tax, government consumption, public education subsidy, and social security policies $\{\tau_p, \gamma_t, \tau_c, g, \theta(e), \chi\}$, sequence of prices $\{r_t, w_t\}$, sequence of accidental bequests $\{Tr_{t,j}\}$, and sequence of measures $\{\Omega_t(\overrightarrow{\omega})\}$ such that:

(i) Given prices, transfers, and policies, the value functions and household policy functions solve the consumer problems in equations 8, 9, 10, 11, 12, 13, 16, 17, 18, 23, 24, 27;

(ii) Interest rate and wage rate satisfy equations 21 and 22, respectively;

(iii) Accidental bequests are transferred to households between ages 50 and 60 ($33 \leq j \leq 43$) after deducting expenditure on private education subsidies:[70]

$$Tr_{t+1,j} = \frac{\int (1-\psi_j) a'_t(\overrightarrow{\omega}) \Omega_t d(\overrightarrow{\omega}) - \kappa \int \theta^{pr}(e) 1_{s=h \text{ and } j \in \{1,2,3,4\}} \Omega_{t+1} d(\overrightarrow{\omega})}{\sum_{j=33}^{43} N_{t+1,j}},$$

where $N_{t,j}$ denotes the mass of population at time t of age j;

(iv) Government budget constraint balances as follows by adjusting γ:

$$\int [\tau_c c_t(\overrightarrow{\omega}) + T(y_{t,j,s,e,\eta,a})] \Omega_t d(\overrightarrow{\omega}) = G_t + E_t + D_t + SS_t,$$

where G_t, E_t, D_t, and SS_t are government consumption, total public education subsidy, federal student loan program expenditure, and social security expenditure, and are computed as follows:

$$\begin{aligned}
G_t =& gY_t = gK_t^{\alpha}(zL_t)^{1-\alpha} \\
E_t =& \kappa \int \theta(e) 1_{s=h \text{ and } j \in \{1,2,3,4\}} \Omega_t d(\overrightarrow{\omega}) \\
D_t =& \int \Big[1_{j \leq 4} [\min[a, 0] - \min[a'_t(\overrightarrow{\omega}), 0]] + \\
& 1_{j>4}(1 - d_{f,t}(\overrightarrow{\omega}))[\min[a,0](1+r_{SL}) - \min[a'_t(\overrightarrow{\omega}), 0]] + \\
& 1_{j>4} d_{f,t}(\overrightarrow{\omega})[-\rho_D(a, y_{t,j,s,e,\eta,a}) + \phi \max[\rho_R(j,a) - \rho_D(a, y_{t,j,s,e,\eta,a}), 0]\Big] \Omega_t d(\overrightarrow{\omega}) \\
SS_t =& \int 1_{j \geq j_r} ss_{t,s,e} \Omega_t d(\overrightarrow{\omega});
\end{aligned}$$

[70] In our baseline calibration, and all of the counterfactual exercises, accidental bequests are always positive, because the assets of those who die exceed the expenditure on private subsidies to education costs. If they did not, then bequests would be negative, which is equivalent to a lump-sum tax.

(v) Labor, capital, and goods markets clears in every period t:

$$
\begin{aligned}
L_t =& \int [1_{j \leq 4, s=h} \eta \epsilon_{j,\ell,e} l_{pt} + 1_{4<j<j_r, s=h} \eta \epsilon_{j,s,e} + 1_{j<j_r, s=\ell} \eta \epsilon_{j,s,e}] \Omega_t d(\overrightarrow{\omega}) \\
K_{t+1} =& \int a'_t(\overrightarrow{\omega}) \Omega_t d(\overrightarrow{\omega}) \\
Y_t =& C_t + K_{t+1} - (1-\delta) K_t + G_t + \kappa \int 1_{j \leq 4, s=h} \Omega_t d(\overrightarrow{\omega}) + \\
& \int \Big[\phi 1_{j>4} d_{f,t}(\overrightarrow{\omega}) \max[\rho_R(j,a) - \rho_D(a, y_{t,j,s,e,\eta,a}), 0] + \\
& \int \Big[\phi 1_{j>4} d_{x,t}(\overrightarrow{\omega}) \max[\rho_R^{pr}(j,x) - \rho_D^{pr}(x, y_{t,j,s,e,\eta,a}), 0] \Big] \Omega_t d(\overrightarrow{\omega}),
\end{aligned}
$$

where C_t is aggregate consumption; and

(vi) $\Omega_{t+1} = \Pi_t\Big(\Omega_t\Big)$, where Π_t is the law of motion that is consistent with the household policy functions and the exogenous processes for population, labor productivities, high school GPA, and college dropouts.

Part III

Results Appendix

E Policy experiments: additional welfare change statistics

Table 59 reports additional welfare change statistics for the two main policy exercises from section 6: (1) information intervention and (2) federal loan limit expansion. These statistics are computed in the period of the transition for two populations, those 19 and over and parents. We report welfare in general and partial equilibrium.

In the case of an information intervention, shown in column (1) of Table 59, in partial equilibrium the cohort of consumers in the period of the transition experience gains (Panel A). However, once we take general equilibrium effects into account, the gains turn to losses (Panel B). In the case of a federal loan limit expansion, which is shown in column (2), the cohort that is 19 and over benefits during the period of the transition in general and partial equilibrium. General equilibrium effects amplify the gains.

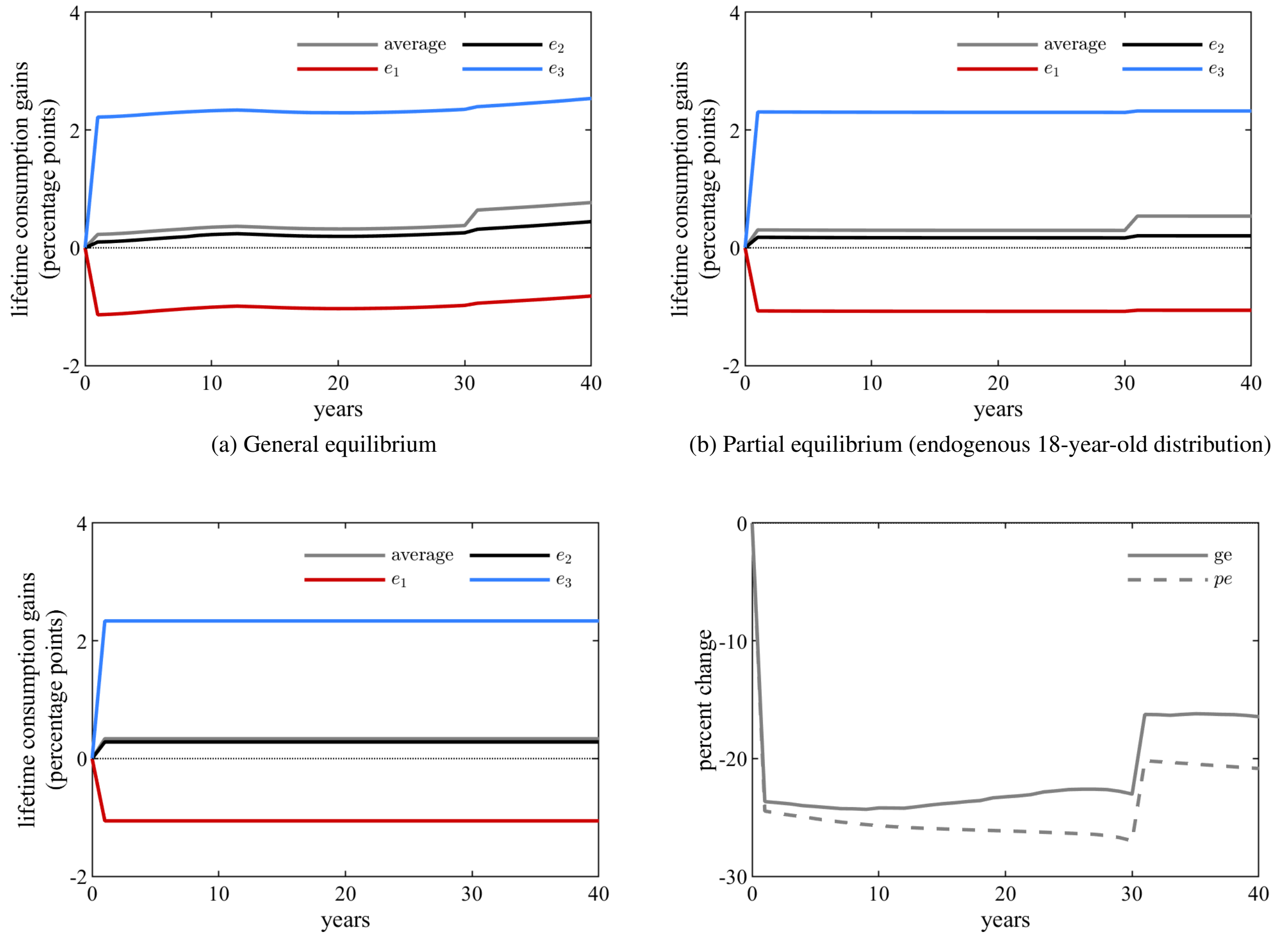
Table 59: Policy experiments: additional welfare change statistics

Variable	(1) Information intervention		(2) Federal loan limit expansion	
	%Δ welfare	%Δ cons.	%Δ welfare	%Δ cons.
Panel A: Partial equilibrium				
Population 19 and over	0.04	0.04	0.09	0.09
Parents	0.11	0.11	0.17	0.17
Panel B: General equilibrium				
Population 19 and over	-0.15	-0.15	0.16	0.16
Parents	-0.12	-0.12	0.32	0.32

Notes: Table 59 provides welfare implications for the population 19 and over and the parents in the period of the transition under the following two exercises: (1) information intervention that eliminates over-optimism, and (2) federal loan limit expansion to fund four years of college net tuition plus room and board. Panels A and B reports lifetime welfare and consumption gains and losses under partial equilibrium and a general equilibrium. In partial equilibrium, the income tax rate, prices, bequests, and Social Security transfers are fixed at their initial steady state values.

F Additional analysis of federal loan limit expansion: general and partial equilibrium effects

In this section, we further analyze the welfare implications of the federal loan limit expansion from section 6.2 in general and partial equilibrium for the population of 18-year-old consumers (Figure 8). Subfigure 8c plots welfare implications in a partial equilibrium in which the income tax rate, prices, bequests, Social Security transfers, and the 18-year-old distribution are fixed at their initial steady state values. In this case, students in the top two skill quantiles benefit while students in the bottom skill quantile are hurt. Further, by construction, the welfare estimates do not change over time.



(a) General equilibrium

(b) Partial equilibrium (endogenous 18-year-old distribution)

(c) Partial equilibrium (exogenous 18-year-old distribution)

(d) Inter-vivos transfers

Figure 8: Federal loan limit expansion welfare analysis for 18-year-olds: general and partial equilibrium effects

Notes: Figure 8 provides a welfare analysis of an expansion in the federal student loan limit to fund four years of college net tuition plus room and board for 18-year-old consumers in general and partial equilibrium. Subfigures 8a-8c report lifetime consumption gains and losses for the average 18-year-old and the average-18-year-old given skill in each period of the transition path under the following cases: (a) general equilibrium, (b) a partial equilibrium in which the income tax rate, prices, bequests, and Social Security transfers are fixed at their initial steady state values, but the 18-year-old distribution is endogenous, and (c) a partial equilibrium in which the income tax rate, prices, bequests, Social Security transfers, and the 18-year-old distribution are fixed at their initial steady state values. Subfigure 8d reports average inter-vivos transfers received by 18-year-olds in general equilibrium in each period of the transition path.

In subfigure 8b, we plot welfare implications in a partial equilibrium in which the income tax rate, prices, bequests, and Social Security transfers are fixed at their initial steady state values, but the 18-year-old distribution is endogenous. In this case, 18-year-olds in the initial periods experience the same gains or losses as in subfigure 8c. However, in the later periods of the transition, the gains to 18-year-olds increase (and losses dampen) because they receive higher inter-vivos transfers (subfigure 8d). Finally, once we take general equilibrium effects into account in subfigure 8a, the 18-year-old consumers' welfare gains are further amplified (or losses dampened). This is due to a fall in the income tax rate, an increase in the savings interest rate, and an increase in accidental bequests and Social Security transfers (see Table 15, column (2)).