Optimism About Graduation and College Financial Aid*

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

September 26, 2023

Abstract

In the United States, college dropout risk is sizable. We provide empirical evidence showing that beliefs about the likelihood of college graduation predict college enrollment and that the distribution of these beliefs exhibits widespread optimism and localized pessimism. We incorporate this distribution of beliefs into an overlapping generations model with college as a risky investment that can be financed via federal and private loans, grants, family transfers, or earnings. We then examine the welfare impact of expanding federal student loan limits. This expansion reduces welfare for young adults who are poor, low-skilled, and optimistic, due to their mistaken beliefs.

JEL classification numbers: I22, I26, E7, G28, G5
Keywords: Subjective beliefs, Optimism, Student loans, Post-secondary education.

*We are grateful to Dirk Krueger, Igor Livshits, Alexander Ludwig, Todd Schoellman, Adam Blandin, Christopher Herrington, Brant Abbott, and Sergio Ocampo Dìaz for their detailed comments on this paper, and also thank Stefania Albanesi, Manuel Amador, Anmol Bhandari, Maria Jose Carreras-Valle, V.V. Chari, Louphou Coulibaly, Doireann Fitzgerald, Alessandra Fogli, Kyle Herkenhoff, Loukas Karabarbounis, Tim Kehoe, Pawel Krolikowski, Marc-André Letendre, Ellen McGrattan, Alexander Monge-Naranjo, Simon Mongey, Juan Pablo Nicolini, Nicholas Petrosky-Nadeau, Richard Rogerson, and seminar and conference participants at the Minnesota alumni workshop, McMaster University, the 2022 North American Summer Meeting of the Econometric Society, 2022 LA Macro Conference, Federal Reserve Bank of Kansas City, Université de Montréal, Queen's University, Federal Reserve Bank of Philadelphia, Western University, 2022 Canadian Macro Study Group, Bureau of Labor Statistics, Lafayette College, William & Mary, University of Hong Kong, Federal Reserve Bank of Minneapolis, Federal Reserve Bank of Richmond, and the 2023 Society of Economic Dynamics. Sergio Salgado provided help with the Survey of Consumer Finances. All errors are ours. This paper draws on a previous working paper circulated under the title "Over-optimism About Graduation and College Financial Aid."

[†]William & Mary. E-mail: egmoschini@wm.edu.
[‡]McMaster University. E-mail: raveeg1@mcmaster.ca.
[§]Queen's University. E-mail: ming.xu@queensu.ca.

1 Introduction

In the United States, approximately one third of students who enroll in a bachelor's program fail to complete their degree. Although student loans may facilitate the financing of college costs, currently a significant amount of outstanding student debt burdens college dropouts.[1] We provide new empirical evidence showing that the expected likelihood of completing a college degree positively predicts college enrollment for high school graduates. When compared to realized college graduation rates, the distribution of these expectations exhibits widespread optimism: most potential college students underestimate dropout risk, although to varying extents. Furthermore, parents have similar patterns of beliefs about their child's educational attainment prospects. In light of this evidence, we build a general equilibrium overlapping generations model, where college is a risky investment that can be financed with federal and private student loans, grants, endogenous family transfers, and labor earnings. Consumers exhibit subjective beliefs about the likelihood of college graduation that may be incorrect, both when they choose whether to enroll in college and when they choose how much wealth to transfer to their child later in life. The estimated model can replicate the empirical responsiveness of college enrollment decisions to both beliefs about one's college graduation likelihood and college tuition subsidies. We then expand the federal student loan limit and examine the welfare impact of this policy change; in our analysis, welfare is measured using lifetime utilities computed with the correct college dropout probabilities, but taking as given consumer choices which are made based on their subjective beliefs. While many high-skill young adults experience welfare gains after the policy change, our analysis adds nuance by showing that many low-skill young adults experience welfare losses. These losses are driven by the enrollment responses of the especially optimistic poor, whose beliefs lead them to transition into college enrollment after the federal loan limit expansion but who are more likely to become high-debt dropouts than they anticipate. With correct expectations about college dropout risk, such welfare losses do not arise. Our results highlight the importance of incorporating subjective beliefs into the analysis of college financial aid policies.

Our main empirical findings are drawn from two nationally representative panel surveys of young people in the United States: the 1997 National Longitudinal Survey of Youth (NLSY97) and the High School Longitudinal Study of 2009 (HSLS:09). In the NLSY97, we observe expectations about the high school students' probability of earning a 4-year bachelor's degree (BA) by age 30, solicited from both the student and their parent. Among high school graduates who later enroll in a BA program, we construct the realized graduation rate by age 30, and impute it to non-enrollees within the same skill tercile (where skill is measured with high school grade point average, or

[1] Sources: 1997 National Longitudinal Survey of Youth, High School Longitudinal Study of 2009, and authors' calculations.

GPA). We show that expected probabilities of earning a BA positively predict college enrollment. We also find that there is widespread optimism about one's likelihood of completing a bachelor's degree among those who later enroll in college; for example, college enrollees believe they have a 92 percent chance of earning a BA by age 30, yet only 70 percent of this group actually go on to earn their degree. The extent of optimism is highest for college enrollees with low skill, a pattern that continues to hold even when we account for gender and parental education. Among those who never enroll in college, the low-skill exhibit sizable optimism and the high-skill slight pessimism. Furthermore, we document similar patterns of subjective beliefs among parents about their child's prospects. In the HSLS:09, our second main source of data, we observe uptake of federal financial aid and private student loans. By using the HSLS:09 to track a cohort of college enrollees until three years after college enrollment (before repayment begins), we confirm that the amount of student debt owed by college dropouts (federal or private) is economically significant at the individual level and in the aggregate. We also use information on student debt portfolios and private loan uptake in the HSLS:09 to provide new findings on the private student loan market, which we use to discipline our model.

Our calibrated model matches empirical moments related to the distribution of subjective beliefs, college enrollment and graduation, student loan uptake and repayment, and family transfers. Besides performing well in validation exercises related to enrollment responsiveness, the model also matches skill-specific college wage premiums. We additionally show that, in both the recent U.S. cohort of the HSLS:09 and in the model's baseline equilibrium, a significant share of college students fully utilize their federal student loans, which indicates that federal loan limits are binding for college enrollees. These high utilization rates—taken together with our new evidence on widespread optimism among potential college students—motivate our main experiment, in which we expand the federal student loan limit so that federal loans can be used to pay for 100 percent of college costs for all four years of college. This represents a significant expansion because federal loan limits in the baseline economy are only enough to finance 37.5 percent of annual college costs, reflecting current U.S. policy.[2]

We find that expanding the federal student loan limit leads to heterogeneous welfare changes, especially among 18-year-olds from poor families: those with high skill who are not especially pessimistic (representing 9 percent of young adults) see large welfare gains equivalent to roughly 4 percent of lifetime consumption, while those with low skill who are especially optimistic (7 percent of young adults) see welfare losses of more than 1 percent of lifetime consumption. To provide intuition for the source of these welfare losses, we introduce the concept of being "over-enrolled"

[2]Definition of college costs: the average value of tuition and fees, net of grants, plus room and board. Sources for borrowing limits: Smole (2019) and NCES (2019), authors' calculations. Sources for utilization rates: HSLS:09 and Smole (2019), authors' calculations.

in college, which describes a college enrollee who would not have enrolled if their beliefs were correct. We begin by focusing on partial equilibrium, where the crucial role of optimistic beliefs is clearest. We show that, in a simplified partial equilibrium without parental altruism, becoming over-enrolled after the policy change is perfectly correlated with experiencing welfare losses. In the partial equilibrium of our quantitative model with altruism, the close association between becoming over-enrolled and experiencing welfare losses is maintained (although the correlation is no longer perfect). In general equilibrium, subjective beliefs continue to have this effect; endogenous prices also introduce welfare losses for additional groups, stemming from the decline in wages for workers with a college degree.

We contribute to previous related work that studies college financial aid policies, which includes Caucutt and Kumar (2003), Ionescu (2009), Lochner and Monge-Naranjo (2011), Chatterjee and Ionescu (2012), Krueger and Ludwig (2016), Ionescu and Simpson (2016), Luo and Mongey (2019), Abbott, Gallipoli, Meghir, and Violante (2019), Caucutt and Lochner (2020), and Colas, Findeisen, and Sachs (2021). A key assumption maintained in these studies is that student and parent expectations about academic outcomes are consistent with realized outcomes. In such frameworks, in partial equilibrium without allowing for potential endogenous adjustments to parental transfers, every young adult is better off after an increase in federal student loan limits. We incorporate subjective beliefs about the likelihood of college graduation among potential college students and demonstrate that, for some consumers, expanding financial aid in the presence of mistaken subjective beliefs leads to welfare losses in partial equilibrium. These welfare losses persist in general equilibrium.[3]

Our new empirical evidence on the expected likelihood of college graduation complements previous work by Stinebrickner and Stinebrickner (2012). That influential study examines a panel survey of students at a small U.S. college, and finds evidence of over-optimism about future academic performance. Using this information, the authors then infer the extent of optimism about college graduation among college students in their sample, and find it to be sizable. We use reported expectations about education attainment in the NLSY97, a nationally representative survey, to provide new evidence on the distribution of subjective beliefs in the population of *potential* college students about the likelihood of completing a bachelor's degree. We show that these beliefs positively predict college enrollment. We also find that there is widespread optimism about one's likelihood of completing a bachelor's degree among college enrollees, especially for those with low skill, and that among those who never enroll in college the low-skill exhibit sizable optimism and the high-skill slight pessimism. Furthermore, we document similar patterns of subjective beliefs

[3]Related work that instead studies regulation of the credit card market includes Nakajima (2012, 2017), which incorporates time-inconsistent preferences, and Exler, Livshits, MacGee, and Tertilt (2021), which allows for optimism about earnings.

among parents about their child's prospects.[4]

We make an additional empirical contribution that provides discipline for the private student loan market in our model environment. As argued by Lochner and Monge-Naranjo (2011), including private student loans in studies of college financial aid policy is important because the private market provides an outside option to the government financial aid program. However, while the current literature has routinely incorporated key features of the federal student aid program into their model frameworks, there is less consensus about modeling the private student loan market. For example, Lochner and Monge-Naranjo (2011) assume that lenders set loan interest rates using repayment risk that depends on student skill, while Ionescu and Simpson (2016) assume that private lenders price the student loan based on the inherent credit risk of the borrower. Abbott, Gallipoli, Meghir, and Violante (2019) assume that students from low-income families do not have access to private student loans. We contribute to the aforementioned literature by using the HSLS:09, supplemented by the 2019 Survey of Consumer Finances (SCF), to document key attributes of the U.S. private student loan market which are then reflected in our model framework. Our approach puts empirical discipline on the nature of imperfect substitutability between private and federal student loans.

This paper proceeds as follows. Section 2 overviews our empirical findings. Section 3 lays out the model, Section 4 describes the model parameterization, and Section 5 analyzes properties of the model's initial steady state equilibrium. Section 6 reports the results of our main experiment. Section 7 concludes.

2 Data

The two main datasets we draw on are the 1997 National Longitudinal Survey of Youth and the High School Longitudinal Study of 2009.[5] Both of these surveys are collected within the United States.

The NLSY97 is a nationally representative panel survey that follows young adults born between 1980 and 1984 ("sample members") from 1997 until 2019. It is collected by the U.S. Bureau of Labor Statistics (Bureau of Labor Statistics, U.S. Department of Labor, 2019). The NLSY97

[4]Previous structural studies that consider subjective beliefs in the context of post-secondary education have examined grant and tax progressivity policy (Matsuda, 2020, 2022), extrapolating from the empirical findings of Stinebrickner and Stinebrickner (2012) about college students at one U.S. college in order to motivate optimism in the model's population of potential college students along multiple margins. In addition to our empirical contributions described in the main text, we differ in our focus on federal student loan policy.

[5]We supplement our findings from these datasets with information on interest rates for education loans from the 2019 Survey of Consumer Finances.

provides information on expected probabilities of earning a 4-year bachelor's degree for sample members and their parents, as well as realized education outcomes. We use this information to quantify the extent to which expectations predict college enrollment, and to compare expected graduation likelihoods with realized college graduation rates.[6]

The HSLS:09 is a nationally representative panel survey that follows a sample of ninth-grade students from 2009 until 2016, although some information from post-secondary transcripts and student records is collected after 2016. It is conducted by the National Center for Education Statistics (NCES), a subsidiary of the U.S. Department of Education (National Center for Education Statistics, U.S. Department of Education, 2020a). Unlike the NLSY97, the HSLS:09 follows a cohort that interacted with the most recent iteration of U.S. financial aid policy, to which we calibrate our structural model (e.g. borrowing limits set in 2012). We use the HSLS:09 to document student loan uptake and balances by college persistence status. We also document the composition of student debt portfolios by loan type (i.e., federal or private) and private loan uptake patterns by high school GPA and family income.

2.1 Subjective beliefs about the likelihood of college graduation

The NLSY97 asks sample members twice about their expected probability of earning a BA by age 30: once in 1997 and again in 2001. The survey also asks parents the same question about their child, but only once, in 1997. This question can be paraphrased as: "What is the percent chance that [you/your child] will have a four-year college degree by the time [you/they] turn 30?" The response is a percentage value between 0 and 100. The NLSY97 also reports the high school GPA (our preferred measure of skill), college enrollment, and educational attainment of sample members over the course of the panel.[7] We use this information to flag those who had enrolled in a BA program, as well as those who had earned a BA, by age 30.[8]

Do beliefs about the likelihood of earning a BA reported in the NLSY97 predict actions? We apply this question to the college enrollment decision in particular, and in Table 1 we report results for a regression in which the dependent variable takes a value of 100 if the individual enrolled in a BA program before age 30 (and is set to 0 otherwise) and the independent variables include the sample member's expected probability of earning a BA degree before age 30 (a value between 0 and 100). Here, we use the most recent valid response to this question collected while the sample

[6] We use "college" to refer to a 4-year bachelor's degree program throughout this paper.

[7] We use high school GPA to measure skill because it is in both the NLSY97 and the HSLS:09, and we want to measure within-skill tercile values of various variables in both data sources. The HSLS:09 does not contain a variable recording either the Armed Services Vocational Aptitude Battery score of the NLSY97 or the Armed Forces Qualifications Test score of the 1979 NSLY.

[8] All tabulations of NLSY97 data do not use survey weights, following Abbott, Gallipoli, Meghir, and Violante (2019).

member was enrolled in high school. Additional controls are also included: model (1) controls for individual characteristics (i.e., the sample member's high school GPA, gender, and age), while model (2) adds family characteristics (i.e., family income and parent education). The estimator is Ordinary Least Squares.

The results presented in Table 1 indicate that the sample member's expected probability of earning a BA positively predicts college enrollment, even when controlling for individual and family characteristics. Specifically, in model (1) a 1 percentage point increase in the expected probability of earning a BA implies a 0.515 percentage point increase in the probability of enrolling in a BA program. This effect falls slightly to 0.473 when we control for family characteristics in model (2). In both model (1) and model (2), the marginal effect of beliefs is highly statistically significant.

We draw two additional conclusions from the results of Table 1. First, the fact that beliefs positively predict enrollment after conditioning on high school GPA indicates that there is heterogeneity in beliefs for high schoolers with the same GPA. Appendix A.1.1 reports a discretized distribution of beliefs conditional on skill in order to quantify this heterogeneity, which we incorporate into the model framework presented in Section 3. Second, because beliefs predict actions, students are not just providing socially desirable answers in their survey response (Krumpal, 2013). Nevertheless, because Table 1 does not rule out that "social desirability bias" may inflate reported expectations in the data, our model parameterization approach of Section 4 allows for such an upward bias in survey responses.

How do expectations about the likelihood of earning a BA compare with reality? Table 2 makes this comparison, beginning with the sample of survey respondents who enroll in a BA program before age 30. Here, we interpret survey responses as the expected probability of graduation conditional on enrollment in a BA program, whereas in the analysis of Table 1 we remained agnostic on this point.[9] For college enrollees, Panel A of Table 2 compares averages of sample member expectations with realized graduation rates by skill tercile, where we assign each sample member to a skill tercile using the distribution of high school GPA among high school graduates. The first column reports the skill tercile; the second column contains the number of observations in each skill tercile for the sample of college enrollees. In the remaining columns we report the within-tercile average, followed by its standard error in parentheses. Specifically, the third column reports the expected probability of earning a BA by age 30, using the most recent response collected while the sample member was enrolled in high school as in Table 2. The fourth column contains the

[9] A more general interpretation of the survey's reported expected probabilityis that it combines the expected probability of enrolling in college with the expected probability of completing college conditional on enrollment. Under this interpretation, for the same reported expected probability, lowering the expected probability of enrollment would raise the implied conditional probability of graduating once enrolled. In that sense, assuming an expected probability of enrollment of 100 percent, as we do in Tables 2 and 3, makes our optimism findings lower bounds.

Table 1: BA enrollment by age 30 as a function of the expected probability of earning a BA

	Enrolled in a BA program by age 30	
Controls	**(1)**	**(2)**
Expected probability of earning a BA by age 30	0.516	0.473
	(0.0329)	(0.0403)
High school GPA	30.38	26.05
	(1.652)	(1.975)
Male	-1.086	0.203
	(1.806)	(2.139)
Age in 1997	0.220	-0.0833
	(0.648)	(0.747)
Logged family income		6.395
		(1.099)
At least one parent BA+		13.96
		(2.695)
Constant	-80.32	-130.1
	(11.29)	(16.06)
R^2	0.259	0.297
Obs	2,367	1,656

Notes: Table 1 presents estimation results from two models. The dependent variable for both models (1) and (2) is a flag for enrollment in a BA program by age 30, which takes a value of 100 if the individual enrolled in a 4-year program BA program by age 30 and 0 otherwise. The expected probability of earning a BA by age 30 uses respondent beliefs, with a range between 0 and 100. Additional controls: Individual-level controls are the respondent's high school GPA (between 0 and 4), an indicator set equal to 1 if the respondent is male and equal to 0 otherwise, the respondent's age in 1997; Family-level controls are the log of family income while the sample member is in high school and an indicator equal to 1 if at least one resident parent has a bachelor's degree or higher, and equal to 0 otherwise. Both models include a constant. Samples: model (1) is high school graduates; model (2) is high school graduates, conditional on observing family income and parent education. Standard errors in parentheses. Source: NLSY97.

realized graduation rates computed as the frequency of BA attainment by age 30. The last column reports the percentage point difference between average expected probabilities and the realized probability, which represents the average extent of optimism for the skill tercile (negative values indicate pessimism). The standard errors in this column are computed using the delta method.

Panel A of Table 2 indicates that, within each skill tercile, the expected probability of earning a BA by age 30 is much higher than the realized rate of attaining that outcome: that is, respondents are optimistic. This is especially true for those with the lowest skill, whose extent of optimism is about 50 percentage points, compared to those with the highest skill, whose optimism is about 15 percentage points. In Table 20 of Supplementary Appendix A.1.1, we show how the extent of optimism for each skill tercile of college enrollees varies by gender and parental education and find that low-skill college students continue to exhibit sizable and relatively higher extent of optimism within each gender and parental education grouping.

Panel A documents optimistic beliefs collected while the respondents are in high school, condi-

Table 2: Subjective beliefs of college enrollees

Panel A Student optimism by student skill tercile among college enrollees	**Skill**	**Obs**	(a) **Expected graduation prob.**	(b) **Realized graduation rate**	**Difference** $(a)-(b)$
	1	222	81.78 (1.70)	31.98 (3.14)	49.80 (3.57)
	2	395	87.42 (1.05)	55.95 (2.50)	31.47 (2.71)
	3	587	93.56 (0.53)	78.19 (1.71)	15.36 (1.79)
	Obs	1,204			

Panel B Student optimism by response timing among college enrollees	**Response timing**	(a) **Expected graduation prob.**	(b) **Realized graduation rate**	**Difference** $(a)-(b)$
	Before enrollment	92.07 (0.89)	69.62 (2.59)	22.45 (2.74)
	After enrollment	93.14 (1.16)	69.62 (2.59)	23.52 (2.84)
	Obs 316			

Panel C Parent optimism by student skill tercile among parents of college enrollees	**Skill**	**Obs**	(a) **Expected graduation prob.**	(b) **Realized graduation rate**	**Difference** $(a)-(b)$
	1	166	80.93 (1.88)	31.33 (3.61)	49.61 (4.07)
	2	297	84.79 (1.20)	54.88 (2.89)	29.91 (3.13)
	3	429	93.03 (0.70)	78.79 (1.98)	14.24 (2.10)
	Obs	892			

Notes: Panel A of Table 2 compares students' mean expected probability of earning a BA program by age 30 with the realized graduation rate within each student skill tercile for the sample of respondents who enrolled in a BA program by age 30; Panel B compares the expected probability of earning a BA collected before and after college enrollment with the realized graduation rate for the sample of respondents who were enrolled in high school in 1997, were enrolled in a BA program in 2001, and who also answered the education expectations question in both years; Panel C compares mean parental expectations for their child's likelihood of earning a BA with realized graduation rates by student skill tercile for the sample of students who enroll in a BA before age 30 whose parents were asked the expected education question while their child was in high school. Table entries for probabilities report within-group means with standard errors in parentheses; standard errors for the extent of optimism are computed using the delta method. Skill terciles are assigned using the distribution of high school GPA among high school graduates. Expectations, graduation rates, and their differences are all in units of percentages. Source: NLSY97.

tional on their eventually enrolling in a BA program. Does the optimism documented in Panel A persist until the college enrollment decision? We argue that it does and offer supporting evidence by examining a group of respondents for whom we can measure expectations on both sides of the college enrollment decision. Specifically, we restrict attention to sample members who answer the 1997 question while still in high school and also answer the 2001 question while enrolled in a BA program. The results are shown in Panel B. If these individuals were changing their expectations right before college enrollment to be closer to the realized probability of graduation, then one could safely presume that the expected probability after enrolling would be closer to the realized probability of graduating, which is about 70 percent. In fact, the expected graduation likelihood increases slightly from 92 to 93 percent.[10]

[10]We do not break down these statistics by skill tercile in Panel B because the sample size is small, which occurs for two reasons. First, a small proportion of respondents meet the education timing criteria. Second, in the 2001 NLSY97 questionnaire, respondents were divided into four groups for the beliefs questions, and only groups 1 and

Panel C reports the same statistics as Panel A but for parental expectations about their child's prospects. Because this panel conditions on observing the parents' expected probabilities of their child earning a BA, the sample differs slightly from that of Panel A. Consequently, the college completion rates by skill tercile change slightly. Panel C indicates that parents, like their children, are optimistic about their child's prospects for earning a BA, and to a similar extent as their child. In fact, expectated beliefs are very similar within families, regardless of college enrollment outcomes for the child, as shown in Table 19 of Supplementary Appendix A.1.

Next, we turn to the sample of those who never enroll in college; enrollment responses due to financial aid expansions will be affected by the beliefs of this group. Table 3 compares the average expected probabilities of those who never enroll with the realized graduation rates of those who enroll in each skill tercile, drawn from Table 2.[11] The results in Table 3 make it evident that in the NSLY97 optimism is present and sizable for the lowest levels of skill regardless of college enrollment outcomes. Additionally, the last row of Table 3 indicates that non-enrollees in the top skill tercile exhibit slight pessimism about the probability of earning a BA (the difference between expectations and reality is negative).[12] Motivated by these findings, the heterogeneity in subjective beliefs that we include in the model of Section 3 allows for both optimism and pessimism about the likelihood of earning a BA among potential college students.

Table 3: Subjective beliefs of non-enrollees

Skill	**Obs**	(a) **Expected graduation prob.**	(b) **Realized graduation rate**	**Difference** $(a)-(b)$
1	585	64.72 (1.31)	31.98 (3.14)	32.74 (2.87)
2	417	69.31 (1.54)	55.95 (2.50)	13.36 (2.90)
3	161	72.17 (2.46)	78.19 (1.71)	-6.02 (3.50)
Obs	1,163			

Notes: Table 3 reports sample counts and beliefs about college graduation likelihood by skill tercile for high school graduates who did not enroll in a 4-year BA program before age 30, along with the realized graduation rate of the skill tercile for those who enroll from Panel A of Table 2. Table entries for probabilities report within-group means with standard errors in parentheses; standard errors for the extent of optimism are computed using the delta method. Skill terciles are assigned using the distribution of high school GPA among high school graduates. Expectations, graduation rates, and optimism are all in units of percentages. Source: NLSY97.

We show supporting evidence for our optimism findings in the NLSY97 from an additional dataset, the HSLS:09, in Table 28 of Supplementary Appendix A.2.2.[13] However, our main findings from

3 were asked about educational attainment expectations.

[11]We tabulate this group seperately from those who enroll in college because we do not directly observe college graduation rates for non-enrollees.

[12]This finding of pessimism among high-achieving non-enrollees speaks to a population of interest in studies such as Hoxby and Avery (2013), Hoxby and Turner (2015), and Dynarski, Libassi, Michelmore, and Owen (2021).

[13]The HSLS:09 is not our preferred source for education expectations because there is no age limit condition on the

the HSLS:09 relate to student loans, and in the next section we use that dataset to document how uptake of student loans varies by college persistence status.

2.2 Student loan uptake and balances

The HSLS:09 contains information about the focal ninth-grade high school student (e.g., their total high school GPA and their expected educational attainment) as well as about their family (e.g., family income and parental education). For the vast majority of sample members, high school graduation occurs in the spring of 2013. The HSLS:09 also contains information on student loan balances, if any, collected from student records submitted by post-secondary institutions. We use the HSLS:09 to demonstrate that there is sizable student loan uptake among those who enroll in a BA program but do not persist toward graduation.

We restrict our sample to students who graduated from high school by the summer of 2013 and enrolled in a BA program in the fall of 2013. Among this group, we additionally restrict attention to individuals for whom we observe family (parent) income, biological parental educational attainment, and the student's high school GPA. We also require that the student reports their educational attainment expectations in the spring of their junior year of high school.[14] In Table 4, we report loan statistics by persistence status; by "persisting" we mean maintaining enrollment in their program from the first year (the 2013-2014 academic year) through their third year (the 2015-2016 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, so we largely avoid using terms such as "dropout" in our discussion of the HSLS:09 findings.

Table 4 shows that 24 percent of the enrolled population fail to persist toward college completion. Those who do not persist owe 19 percent of the sample's student debt balances (either federal or private) and are more likely to have student debt relative to those who persist. Conditional on having student debt, the average and median student loan balance is economically significant several years after enrollment, regardless of persistence status. This is true despite non-persisters using that money to finance fewer years of tuition, compared to students who persist toward degree completion.[15] In the next section, we focus particularly on private loans using information from

outcome being asked about, and the response is categorical (e.g., "Bachelor's") rather than a continuous probability.

[14]This allows us to use a consistent sample for both the student debt findings and the robustness exercise of Supplementary Appendix A.2.2.

[15]Our findings documenting student debt among dropouts in the HSLS:09 complementf the work of Chatterjee and Ionescu (2012), which uses the SCF to show that outstanding balances held by college dropouts are significant. We expand this analysis in two ways and reach similar conclusions. First, we use the HSLS:09 to document significant balances among dropouts by tracking a single cohort of college students until three academic years after enrollment. These attributes are an advantage compared to a cross-sectional sample like that of the SCF, with the potentially

the HSLS:09 and the 2019 SCF, interpreted using several additional sources.

Table 4: Student loan incidence by persistence status

Persistence status	% of enrollees	% of SL $	% with SL	Average $	Median $
Did not persist	24	19	78	15,270	12,238
Persisted	76	81	65	24,648	19,500
Obs	2,356				

Notes: Table 4 divides the pool of 2013 bachelor's degree enrollees into students who persisted in college and those who did not persist. Persistence status is assigned based on whether their student record indicates that they were enrolled for each academic year between 2013-2014 and 2015-2016. Within each persistence status group, the table reports the group's percentage of 2013 enrollees, the dollars owed by the group as a percentage of aggregated student debt among 2013 enrollees, the percentage of the group with a positive student debt balance, and the average and median student loan balance owed by debtors in the group after three academic years, in 2016 dollars. Percentages are rounded to the nearest percentage point. Weights are PETS-SR student records longitudinal weights. Source: HSLS:09.

2.3 Private student loans

The private student loan market warrants further examination; it is the source of a potential substitute for federal loans and, therefore, relevant for our loan limit expansion exercise. We begin with information from the HSLS:09 reported in Table 5, which summarizes sources of student loans three academic years after enrollment among 2013 college enrollees. Results are broken down separately for each persistence status. Moving from left to right, the columns report, first, the percentage of the group that has either federal or private student loans; second, the percentage that has only federal loans; third, the percentage with only private loans; and fourth, the percentage with debt from both kinds of student loans. This table has two main takeaways, which hold for both persistence statuses: first, more than one in five students take out a private student loan during college, indicating that using this source of financing is somewhat common; and, second, there is a pecking order for loan types, where students tend to take out a federal loan and then sometimes turn to private loans. For intuition about the second takeaway, note that if students often took out private loans without first using federal loans, then the share of student debtors with only private loans would be more similar to the share with only federal loans. However, Table 5 shows that this is not the case in the data: for both persistence groups the share with only private student loans is almost 0, whereas the share with only federal loans is quite large.

The HSLS:09 also sheds light on access to private student loans by key student characteristics.

large heterogeneity in federal policy regimes at loan issuance, time in repayment, labor market experience, and other factors that such a sample implies. Second, information in the HSLS:09 on student debt balances is drawn from student records submitted by post-secondary institutions, which are likely to be a more reliable source of information than self-reported balances recorded in the SCF.

Table 5: Student loan portfolio composition

Persistence status	**Either**	**Only federal**	**Only private**	**Both**
Did not persist	78	53	1	24
Persisted	65	44	2	20
Obs	2,356			

Notes: Table 5 reports, by persistence status, the percentage of all 2013 bachelor's degree enrollees who owe money for either, only federal, only private, or both types of student loans three years after enrollment. Percentages are rounded to the nearest percentage point, so the sum of the last three columns may not exactly equal the value in the first column. Weights are PETS-SR student records longitudinal weights. Source: HSLS:09.

Table 6 reports uptake rates for private loans, computed as the percentage of each tercile of the joint distribution of family income and skill that has taken out a private student loan three academic years after they began college. Family income and skill terciles are assigned using the distribution of each variable among high school graduates. This table illustrates that college students from the poorest families and college students in the lowest skill tercile take out private loans like their richer and higher-skill peers. Because access is a necessary condition for uptake, and terciles are assigned using the distribution of high school graduates, the results in Table 6 reject the hypothesis maintained in previous studies that low-skill or low-income prospective college students are barred from the private student loan market.

To be clear, in Table 6 we do not claim to demonstrate that all prospective college students necessarily have access to private student loans. Based on industry reports and guides for potential private loan borrowers, it seems that with most private lenders having a cosigner is a sufficient condition for access to private student loans at good terms, yet the presence of a qualifying cosigner is likely not highly correlated with skill or family income. Indeed, among the five largest private student lenders, 90 percent of undergraduate student loans issued since 2010 have had a cosigner (Amir, Teslow, and Borders, 2021).[16] Most adults qualify as cosigners for private student loans: for loan approval, the minimum credit score requirements range from none to 680, and even cosigners without a credit score could still qualify with some lenders if their income is steady and meets a low threshold level (Holhoski, Clark, and Beresford, 2022).

Because the HSLS:09 does not report the student loan's interest rate, we turn to the 2019 SCF to compare interest rates on private and federal student loans, as well as to examine how loan interest rates vary with borrower attributes within private loans. We find that the mean and median interest rate is very similar across the two loan types, and that the interest rate on private student loans

[16] For students without a cosigner, it is much more difficult to get any private student loan in the freshman and sophomore years of college. However, in the junior and senior years of college, students with a credit score and a steady income can get a private student loan. For an example of a private student loan that does not require a cosigner, see Funding U., Inc. (2022).

Table 6: Private loan uptake rates

		GPA		
	Q	**1**	**2**	**3**
Income	**1**	19	25	16
	2	26	31	24
	3	39	21	19
Obs	2,356			

Notes: Table 6 reports the percentage of each cell that has a positive private student loan balance three academic years after enrollment in the fall of 2013. Percentages are rounded to the nearest percentage point. Rows are student family income terciles using parents' income during high school; columns are high school GPA terciles. terciles are assigned using the distribution of high school graduates. Weights are PETS-SR student records longitudinal weights. Source: HSLS:09.

exhibits little variation across borrowers with different attributes. Details of these findings are reported in Supplementary Appendix A.3.

In the next section, we build a model framework that incorporates our empirical findings from Sections 2.1 and 2.3 on subjective beliefs, financial aid, and the private student loan market. Our findings on student loan uptake in Section 2.2 are used to validate the calibrated model in Supplementary Appendix C.1.

3 Model

Motivated by our findings in Section 2.1, we enrich the general equilibrium life cycle model with college choice of Krueger and Ludwig (2016) by incorporating subjective beliefs about the likelihood of college graduation. We also incorporate endogenous and exogenous college dropout, as in Chatterjee and Ionescu (2012). The features of the federal student loan program are largely based on Luo and Mongey (2019), and the features of the private student loan market are based on empirical patterns documented in Section 2.3.

3.1 Overview

Time is discrete and runs forever; each period lasts one year. There are four main kinds of agents in the economy: consumers, the government, private lenders, and a final goods firm.

Consumers Let j denote the age of consumers; consumers start making decisions when they turn 18 at $j = 1$. Figure 1 illustrates the phases of the consumer's adult life cycle.

Let s denote the skill endowment; at the start of adulthood, with an exogenous probability $q(s)$, 18-year-old consumers may choose whether or not to enroll in college; otherwise, college is not an

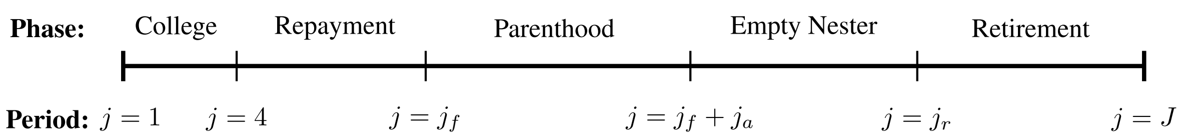

Figure 1: Phases of the consumer's life cycle

option, and they join the workforce without a college degree.[17] The enrollment choice depends on the skill endowment as well as an idiosyncratic earnings productivity, η, initial net assets, a, and the subjective belief about the annual probability of being allowed to continue in college, $\hat{p}$.

The skill endowment is drawn once from a conditional distribution that depends on parental education. The skill endowment indexes the distribution from which the consumer's subjective belief is drawn at the start of adulthood, as well as the true probability of being allowed to continue given enrollment, deterministic earnings productivity, and proportional grants for college from the government and private sources. The idiosyncratic productivity component of earnings follows a lag-1 auto-regressive, or AR(1), process that depends on completed education. Net assets at the start of adulthood are determined by a one-time inter vivos transfer from the consumer's parent.

Earning a college degree requires four completed years of enrollment. Consumers learn their true annual probability of being allowed to continue in college, denoted by $p_c(j, s)$, immediately after enrollment; they may then choose to leave college any time after their first academic year.[18,19] Education is recorded with e; a college student or college graduate has a high level of education, indicated with $e = h$, although only a college graduate enjoys the college wage premium. If the consumer never enrolls, or drops out of college, then they have a low level of education and e is set to ℓ. College enrollees have access to federal student loans, where the stock of debt is indicated with $a < 0$. Enrollees also have access to private student loans, denoted with x; this variable takes a positive value to record the stock of private debt, which is set to zero for all consumers at the

[17] This model feature captures academic, personal, or family reasons that prevented college enrollment (see Table 26 in Supplementary Appendix A.2.1 for suggestive empirical evidence). This modeling approach is a nested version of stochastic utility costs, where with probability $1 - q(s)$, the realized cost draw is large enough to prevent enrollment.

[18] The arguments of $p_c\,(j, s)$ are motivated by our empirical findings presented in Tables 25 and 27 in Supplementary Appendix A.2. The findings of Table 25 indicate that the annual conditional probabilities of persisting in college are increasing during one's college career, while the findings reported in Table 27 indicate that high school GPA rather than other student, family, or institution characteristics, predicts continuation in college enrollment in the HSLS:09. As for attributing this role (at least partly) to an exogenous shock, we build on the findings of Stinebrickner and Stinebrickner (2012), which show that heterogeneity in ability, rather than heterogeneity in effort, leads to college dropout. For example, even for students in the same major who put in the same hours of study, the paper finds significant differences in academic performance.

[19] In principle, the assumption that consumers update their beliefs to the truth immediately after enrollment minimizes the impact of subjective beliefs on consumer behavior. However, quantitatively, this assumption does not matter: in Supplementary Appendix C.3, we consider a sensitivity analysis where college students never learn about their true likelihood of being allowed to continue.

beginning of $j = 1$ and is fixed at that value for those who never enroll.

Figure 2 depicts the college phase from the perspective of an 18-year-old who decided to enroll given their initial state $(s, \eta, a, \hat{p})$, shown in the "State" row at the top of the figure. The thick black arrow traces the already-realized path of this fictional student; the figure shows the student's possible paths at the end of their first academic year, after they have learned their true probability of being allowed to continue but before they know if they will have the option. In the figure, possible outcomes for this draw are indicated by the orange arrows. At this point, their state space reflects the education choice ($e = h$), and includes possible federal and private student loan balances (reflected in a and x, respectively). Not being allowed to continue enrollment generates exogenous dropout, which represents being forced to leave because of a lack of academic ability. If allowed to continue, the student may also choose to endogenously drop out at the start of the next academic year (dotted blue arrow). Although not shown in the figure, in the model a college enrollee faces a similar situation at the end of each academic year of college—the only difference being that after year 4 there is no choice to continue, because the student has graduated.

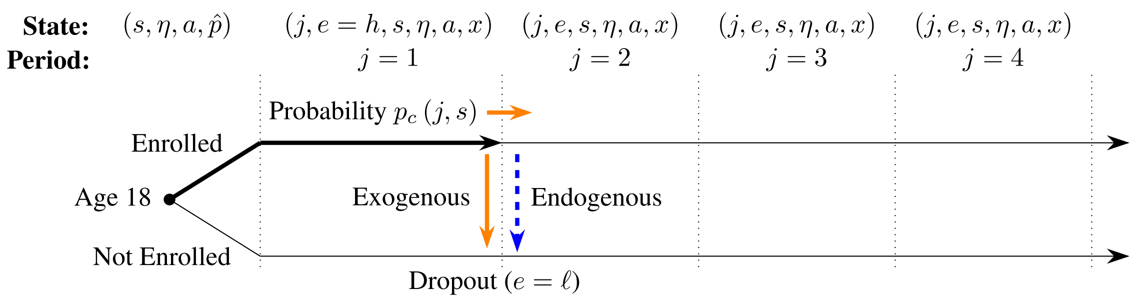

Figure 2: The college phase for a first-year college enrollee

Notes: Figure 2 illustrates the college phase of the consumer's life cycle from the perspective of a college enrollee who is completing their first academic year. The first row of the figure ("State:") indicates the state space for that period; the second row ("Period:") indicates the period of adulthood, given by consumer age j. The thick black arrow traces the path of a consumer who, at the start of period $j = 1$ when they are 18 years old, decides to enroll given their state vector $(s, \eta, a, \hat{p})$, so that $e = h$ in period $j = 1$. At the end of their first academic year, the student faces a probability $p_c(j, s)$ that they will be allowed to continue to their second academic year; otherwise, they exogenously drop out (solid orange arrows). This probability may differ from their expected probability at the point of enrollment, $\hat{p}$. If allowed to continue, consumers may also choose to leave college by endogenously dropping out (dashed blue arrow). To complete college, students must allowed to continue their enrollment (and choose to do so) through all four periods of the college phase, and be allowed to graduate.

The benefits of graduating from college are higher labor earnings, a higher probability of having high-skill children, and higher Social Security transfers.[20] The costs of college are foregone earnings due to part-time work, an effort cost net of any consumption value of college, and an annual

[20]Consumers must graduate from college to enjoy these benefits. In Table 22 of Supplementary Appendix A.1.2, we show that, relative to having only a high school degree, the marginal effect of some college (college dropouts or those with an associate's degree) on the age profiles of earnings is approximately zero.

pecuniary cost (tuition and fees). College expenses including room and board can be financed with student loans borrowed from the federal student loan program and the private loan market, as well as inter vivos transfers from parents, grants from public and private sources, and earnings from part-time work while enrolled.

After the age of college graduation, consumers with an outstanding student loan balance may be either college graduates or college dropouts; student debt is the only form of debt in the economy. At this point, consumers begin to make decisions on whether to make loan payments: in particular, they may choose to repay only federal loans, only private loans, both types of loans, or neither type of loan.[21] Upon paying off student loans, consumers solve a standard consumption-savings problem.[22] Consumers who do not make payments on their student loans are considered delinquent, and their disposable income above the amount $\bar{y}$ is garnished at the rate τ_g. In the period they are delinquent, delinquent debtors also incur a collection fee and a utility cost whose value is indexed to the type of loan on which they are delinquent.[23]

All consumers have a child at the fertile age, j_f; this child will leave the household j_a years after birth. At the beginning of the period when the child leaves, as in Krueger and Ludwig (2016) and Abbott, Gallipoli, Meghir, and Violante (2019), each parent makes an inter vivos transfer to their child after observing the child's skill, s_c, and the subjective belief about the child being allowed to continue in each year of college, $\hat{p}$. Note that the subjective belief is the same for both the parent and the child. The transfer is motivated by parental altruism, where the parent's subjective beliefs about the annual likelihood of their child being allowed to continue toward college completion are built into the altruism term included in their objective function. Consumers retire at age j_r. At this point, they stop working and receive Social Security transfers. Consumers survive each period with probability ψ_j, and live for a maximum of J periods.

The subjective beliefs discussed thus far imply that consumers in our model deviate from rational expectations in the following way: 18-year-olds making the college enrollment decision (and their parents choosing inter vivos transfers) believe that they are unique when it comes to their probability (or their child's probability) of being allowed to continue from one academic year to

[21] In our model, it is not possible to have an either federal or private student loan debt written off via default. This is consistent with U.S. policy, where student loans may eventually be classified as defaulted loans but are almost never discharged.

[22] We assume that student loans must be paid off for consumers to save because this reduces the state space necessary to represent asset positions from three to two elements. This assumption is consistent with optimizing behavior by the consumer in an environment in which consumers cannot be delinquent, because in that case, the optimal strategy would be to pay off all loans before saving as long as the interest rates on loans are higher than the savings interest rate. The interest rates are ordered in this way in our framework by construction. This incentive is somewhat offset because of the delinquency choice we incorporate, but that is not a quantitatively significant concern.

[23] These garnishment rules reflect the U.S. system, where both federal and private lenders may garnish earnings (private lenders require a court order).

the next. Consumers understand everything else about their environment: they know their own skill, how skill affects earnings, and that others have subjective beliefs. Because individuals are atomistic, they can believe that their own continuation probability is uniquely different from others of the same type, while taking as given aggregate endogenous states which are computed using decision rules of consumers with subjective beliefs and then simulated with the true continuation probabilities.[24]

Government The federal student loan program is characterized by a cumulative student loan limit $\bar{A}$ and a student loan interest rate $r_{SL} = r + \tau_{SL}$, where r is the risk-free interest rate on savings and τ_{SL} is the add-on set by the government. To expand federal student loan limits in our policy experiment, we increase $\bar{A}$ from its baseline value. Federal student loans are assessed interest starting from the year after the age of college graduation ($j > 4$).[25]

In addition to running the federal student loan program, the government provides grants for college education and funding for Social Security, and also faces an exogenous government consumption requirement expressed as a fixed fraction g of gross domestic product (GDP). Expenditures are financed with revenue generated from progressive income taxes and a flat-rate consumption tax.

Private lenders The private student loan market is characterized by a continuum of risk-neutral competitive lenders. The features of the private loan market are based on findings from our empirical analysis in Section 2.3. First, to capture the pecking order from federal to private student loans shown in Table 5, we introduce a loan uptake cost specifically for acquiring private student loans. This cost makes private student loans an imperfect substitute for federal student loans; it represents the additional effort required in the private student loan market to avoid predatory lending and hidden fees, as well as potential difficulties in acquiring a cosigner or even finding a lender. Second, we do not explicitly exclude any consumer from access to the private student loan market, which is consistent with positive private loan uptake patterns observed in Table 6. Third, we incorporate a student loan issuance cost that is common to both private and federal student loans, τ_{is}, to capture the fact that the mean and median of private student loan interest rates are roughly the same as federal student loan interest rates, as shown in Table 29. Fourth and finally, to capture the lack of variation in private student loan interest rates along key characteristics (Table 29), we assume lenders cannot price-discriminate by skill or any other characteristic. Consequently, lenders pool each cohort of students to price loans, which leads to a single interest rate, r_{SL}^{pr} (see equation (22)

[24] Consumers in this model have rational expectations if $\hat{p} = p_c(j, s)$ for all j and s.

[25] All loans in this model are subsidized; the federal student loan program we model abstracts from features such as unsubsidized loans, loan fees, and the Expected Family Contribution (EFC). In Supplementary Appendix C.3, we show that our main findings do not change if we incorporate a higher add-on for the student loan interest rate as a sensitivity analysis for the lack of unsubsidized loans and loan fees. The EFC would introduce heterogeneity in access to need-based aid like subsidized loans; this represents an intermediate case between the exercise in Supplementary Appendix C.3 and the specification of the main text.

in Supplementary Appendix B.2).[26]

Final goods firm Output is produced by a final goods firm, which operates a production technology in which the inputs are capital, efficiency units of low-skill labor, and efficiency units of high-skill labor.

3.2 Consumer life cycle problem

This section presents the main value functions; remaining value functions are presented in Supplementary Appendix B.1. Recall that $e \in \{h, \ell\}$ denotes education status where h refers to a high-education consumer who either is enrolled in college or is a college graduate, and ℓ refers to a low-education consumer who did not go to college or who dropped out of college. Recall that $a \geq 0$ indicates positive net assets that earn an interest rate r and $a < 0$ indicates federal student loan balances, while $x > 0$ denotes the outstanding balance of private student loans.

Consumer problems before college graduation age (j ≤ 4) Given their type, $(s, \eta, a, \hat{p})$, which reports skill, idiosyncratic AR(1) productivity, net assets, and the subjective belief about being allowed to continue in each of year of college, respectively, an 18-year-old (age $j = 1$) has a value function given by

$$\begin{aligned} \hat{W}(s, \eta, a, \hat{p}) = & q(s)[\max_{\hat{d}_e}(1 - \hat{d}_e)V(j = 1, \ell, s, \eta, a, x = 0) + \\ & \hat{d}_e \hat{V}(j = 1, h, s, \eta, a, x = 0, \hat{p})] + (1 - q(s))V(1, \ell, s, \eta, a, x = 0) \end{aligned} \tag{1}$$

With probability $q(s)$, the consumer may choose whether to enroll in college or not by selecting selecting $\hat{d}_e \in \{0, 1\}$, where $V(j = 1, \ell, s, \eta, a, x = 0)$ is the value of not going to college and $\hat{V}(j = 1, h, s, \eta, a, x = 0, \hat{p})$ is the value of going to college given subjective beliefs about being allowed to continue in college (hereafter referred to as the subjective value of college). At this point, the balance of private student loans is set to 0. With exogenous probability $1 - q(s)$, the consumer does not have the option to enroll and proceeds through life as a low-education worker

[26] We have one market for private student loans because most loans are co-signed. We could incorporate another market for loans that are not co-signed. These loans would have worse terms than co-signed loans. This would make private loans even more of an imperfect substitute for federal loans. Hence, our model specification likely imposes a lower bound for the welfare changes from the federal loan limit expansion experiment.

with no student debt. The value of not going to college or dropping out for $j \leq 4$ is given by

$$V(j,\ell,s,\eta,a,x) = \max_{c\geq 0,a'} U(c,j,\ell) + \beta\psi_j E_{\eta'|\ell,\eta} V(j+1,\ell,s,\eta',a',x) \tag{2}$$

$$s.t.$$

$$(1+\tau_c)c + a' = y_{j,\ell,s,\eta,a} + a + Tr_j - T(y_{j,\ell,s,\eta,a})$$

$$a' \begin{cases} = a & \text{if } \ a < 0 \\ \geq 0 & \text{otherwise} \end{cases}$$

where c is consumption, a' is next period assets or federal student loans, $U(\cdot)$ is the utility function, β is the discount factor, τ_c is the consumption tax rate, $y_{j,\ell,s,\eta,a}$ is pretax income, Tr_j is accidental bequests, and $T(y)$ is the income tax function. For college dropouts solving (2), the stock of student debt (federal or private) is held fixed until $j = 5$, at which point they begin repayment. For consumers who never enroll in college, net assets are always weakly positive because student loans are the only form of borrowing. When making the college enrollment decision, the subjective value of college for $j = 1, 2, 3$ is given by

$$\hat{V}(j,h,s,\eta,a,x,\hat{p}) = \max_{\hat{c}\geq 0,\hat{a}',\hat{x}'} U(c,j,h) - \xi_L \mathbb{I}_{a\geq 0 \text{ and } x=0 \text{ and } (\hat{a}'<0 \text{ or } \hat{x}'>0)} - \xi_L^{pr}\mathbb{I}_{x=0 \text{ and } \hat{x}'>0} \tag{3}$$

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

$$s.t.$$

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

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

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

$$\hat{x}' - x \in \big[0, [(1-\theta(s)-\theta^{pr}(s))\kappa + \bar{c}] - [\max(-\hat{a}',0) - \max(-a,0)]\big]$$

where $\hat{x}'$ is next period private student loans, ξ_L is the loan search and debt aversion cost of acquiring any student loan and ξ_L^{pr} is the additional cost of acquiring a private student loan.[27] The parameter κ is annual tuition and fees;[28] $\theta(s)$ and $\theta^{pr}(s)$ are the share of tuition and fees paid for by public and private grants given skill, respectively; and $\bar{c}$ is the amount that can be borrowed for room and board expenses while in college.[29] These consumers may choose to drop out before the start of the next academic year, which is captured by the $\max$ expression in the continuation value.

[27] The utility costs of taking out any student loans can be due to excessively complicated paperwork, as noted by Dynarski and Scott-Clayton (2008).
[28] In Supplementary Appendix C.3, we analyze the case where tuition depends on skill. The main results do not change.
[29] Room and board is not a mandatory expenditure in our model because most students live off campus in practice, as shown in NCES (2020b).

College students can borrow from federal student loans, where $\bar{A}$ represents the number of years worth of net tuition and fees plus room and board expenses that the federal student loan limit is sufficient to finance.[30] The last constraint in equation (3) is the limit constraint for private student loans, which requires that the flow amount borrowed from private student loans in a given year must not exceed tuition plus room and board costs net of any financial aid.[31] The subjective value for the final year of college, when $j = 4$, is presented in equation (15) in Supplementary Appendix B.1. When constructing this value, the post-college continuation value conditional on graduation is based on $E_{\eta'|h,\eta}$ rather than $E_{\eta'|\ell,\eta}$. Furthermore, no endogenous dropout decision will be made in the continuation value because in the next period, the consumer will have graduated from college. The rest of the value function for the final year of college remains unchanged from previous years.

When consumers make the college entrance decision in equation (1), they have subjective beliefs and will use the subjective value of college from (3) to compute their expected value. However, consumers learn the true probabilities of being allowed to continue in the first year of college so that, while enrolled, the consumer's realized consumption-savings and dropout decisions are based on the following value function for $j = 1, 2, 3$:

$$V(j, h, s, \eta, a, x) = \max_{c\geq 0, a', x'} U(c, j, h) - \xi_L \mathbb{I}_{a\geq 0 \text{ and } x=0 \text{ and } (a'<0 \text{ or } x'>0)} - \xi_L^{pr} \mathbb{I}_{x=0 \text{ and } x'>0} + \tag{4}$$

$$\beta\psi_j E_{\eta'|\ell,\eta}\left[p_c(j,s)\max[V(j+1,h,s,\eta',a',x'), V(j+1,\ell,s,\eta',a',x')] + (1-p_c(j,s))V(j+1,\ell,s,\eta',a',x')\right]$$

where the control variables and constraints (omitted for the purpose of exposition) for this value function are the same as in the subjective value function given by (3), but without the hats. The only difference between this value function and the subjective value function is that (4) incorporates true probabilities of being allowed to continue in each year of college, $p_c(j, s)$, rather than the subjective belief probability, $\hat{p}$. Again, in the final year of college ($j = 4$), the consumer's value of college will be computed using equation (15) in Supplementary Appendix B.1, with the exception that the consumer will use the true probability of being allowed to continue rather than the subjective belief about being allowed to continue in each year of college.

[30] For example, if $\bar{A}$ is equal to four, then the limit is equal to four years of net tuition and fees, plus room and board, or 100% of annual college costs. The multiplier $\frac{j}{4}$ is an adjustment for the fact that the cumulative limit increases with each year of college.

[31] In our model, the only benefit of a private loan over a federal loan is that with a private loan, a college enrollee can keep their savings, whereas with a federal loan, a college enrollee must exhaust savings to borrow. With this feature, our model can generate uptake of only private loans by a small minority of students, a pattern we see in the data (see Table 37 in Supplementary Appendix C.3).

Consumer problems after college graduation age (j > 4) Consumers begin student loan payments the year after college graduation age, regardless of whether or not they complete college.[32] For the remainder of this section, we focus on the parent's problem of choosing between repayment and delinquency and their value of repayment at age $j_f + j_a$, the age at which they make an inter vivos transfer. At the start of age $j_f + j_a$, the parent draws their child's type and the family's subjective belief and then chooses whether or not to be delinquent on any student debt payments. The value function before the draw of child type and subjective belief is given by

$$V(j,e,s,\eta,a,x) = \sum_{s_c} \pi_{s_c}(s_c|e) \sum_{\hat{p}} \pi_{\hat{p}}(\hat{p}|s_c)\,[\max_{d_f,d_x}(1-d_f)(1-d_x)V^R(j,e,s,\eta,a,x,s_c,\hat{p}) \tag{5}$$
$$+\, d_f(1-d_x)V^{D_f}(j,e,s,\eta,a,x,s_c,\hat{p}) + (1-d_f)d_xV^{D_x}(j,e,s,\eta,a,x,s_c,\hat{p})$$
$$+\, d_fd_xV^D(j,e,s,\eta,a,x,s_c,\hat{p})\,],$$

where $\pi_{s_c}(s_c|e)$ is the conditional probability over child skill given parental education level, $\pi_{\hat{p}}(\hat{p}|s_c)$ is the conditional probability over the subjective belief about the child being allowed to continue in each year of college given child skill, and $d_f \in \{0,1\}$ and $d_x \in \{0,1\}$ denote the federal and private student loan delinquency decisions, respectively. The terms $V^R(\cdot)$, $V^{D_f}(\cdot)$, $V^{D_x}(\cdot)$, and $V^D(\cdot)$ denote the value of repayment on both loans, the value of delinquency on only federal loans, the value of delinquency on only private loans, and the value of delinquency on both types of loans, respectively. Here, we show the value of repayment for $j = j_f + j_a$, given by

$$V^R(j,e,s,\eta,a,x,s_c,\hat{p}) = \max_{c\geq 0,a',b} U(c,j,e) + \beta\psi_j E_{\eta'|e,\eta}V(j+1,e,s,\eta',a',x')+ \tag{6}$$
$$\beta_c E_{\eta'|\ell}\hat{W}(s_c,\eta',b,\hat{p})$$
$$s.t.$$
$$(1+\tau_c)c + a' + b = y_{j,e,s,\eta,a} + a + r_{SL}a\mathbb{I}_{a<0} + Tr_j - T(y_{j,e,s,\eta,a}) - \rho_R^{pr}(j,x)$$
$$a' \begin{cases} = (1+r_{SL})a + \rho_R(j,a) & \text{if } a < 0 \\ \geq 0 & \text{if } a \geq 0 \text{ and } x = 0 \\ = 0 & \text{otherwise } (a \geq 0 \text{ and } x > 0) \end{cases}$$
$$x' = (1+r_{SL}^{pr})x - \rho_R^{pr}(j,x)$$
$$b \begin{cases} = 0 & \text{if } a < 0 \text{ or } x > 0 \\ \geq 0 & \text{otherwise } (a \geq 0 \text{ and } x = 0) \end{cases}$$

[32] In the United States, federal student loans typically have a six-month grace period after graduation before repayment begins. A period lasts one year in our model, so we set repayment to begin the first period after graduation. We assume that payments begin in the same age for dropouts for simplicity.

where b is the inter vivos transfer to the child, $\hat{W}(\cdot)$ is the child's value function, and β_c disciplines the intensity of parental altruism toward the child. Because the parent uses $\hat{W}(\cdot)$ for their child's lifetime utility, the parent also has the same subjective belief as the child about the likelihood of the child being allowed to continue in college. The child's AR(1) productivity η' is drawn from the stationary distribution for a consumer without a college degree. The objects $\rho_R(j, a)$ and $\rho_R^{pr}(j, x)$ are full payment functions for federal and private student loans, respectively. If the parent has outstanding federal loans, then $a' = (1 + r_{SL})a + \rho_R(j, a)$. As in Ionescu and Simpson (2016), we assume consumers cannot choose to pay down their federal or private loans faster than the required payment amount. If the parent has paid off their student loans, then they may save and make an inter vivos transfer to their child.[33]

The definition of the equilibrium and the computational algorithm are provided in Supplementary Appendix B.2 and B.3, respectively.

3.3 Functional forms

Probability of being allowed to continue college The true probability of being allowed to continue to the next year of college, $p_c(j, s)$, is determined by two objects, $p(s)$ and $\rho_d(s)$, both of which depend on skill:

$$p_c(j, s) = 1 - (1 - p(s))\rho_d(s)^{j-1} \tag{7}$$

Equation (7) implies that the exogenous dropout probability is $(1 - p(s))\rho_d(s)^{j-1}$. The object $p(s)$ determines the common probability of being allowed to continue in college in any year of enrollment, while the object $\rho_d(s)$ determines the persistence of the exogenous drop out probability. Note that the functional form of equation (7) uses two parameters per skill bin to assign a probability for each year because we observe two conditional persistence probabilities in the HSLS:09.

Student loan payments The full payment function $\rho_R(j, a)$ for federal student loans is given by

$$\rho_R(j, a) = \begin{cases} -\dfrac{r_{SL}}{1 - (1 + r_{SL})^{-(T_{SL}+5-j)}}a & \text{if } a < 0 \text{ and } 4 < j \leq T_{SL} + 4 \\ -(1 + r_{SL})a & \text{if } a < 0 \text{ and } j > T_{SL} + 4 \\ 0 & \text{otherwise } (a \geq 0) \end{cases} \tag{8}$$

If there is an outstanding balance and j is still within the standard repayment period, T_{SL}, the loan is amortized with an interest rate of r_{SL}; if there is an outstanding loan balance and the standard repayment period has expired, the outstanding principal plus interest is due; and, if there

[33] In the initial stationary equilibrium, the situation where a parent cannot make an inter vivos transfer due to outstanding student loans is rare: only 0.39 percent of consumers at $j_f + j_a$ have student loans.

is no outstanding loan balance, the payment amount is zero. Instead of repayment, if a consumer chooses delinquency, their disposable income above $\bar{y}$ is garnished at the rate τ_g. This leads to a partial payment function given by

$$\rho_D(j,a,y) = \min[\tau_g \max[y - T(y) - \bar{y}, 0], \rho_R(j,a)] \tag{9}$$

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

The payment structure for private student loans parallels the payment structure for federal student loans with the full payment function, $\rho_R^{pr}(j,x)$, and the partial payment function, $\rho_D^{pr}(j,x,y)$, defined analogously.

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

$$U(c,j,e) = \frac{\left(\frac{c}{1+\zeta \mathbb{I}_{j\in\{j_f,..,j_f+j_a-1\}}}\right)^{1-\sigma}}{1-\sigma} - \lambda \mathbb{I}_{e=h \text{ and } j\in\{1,2,3,4\}} \tag{10}$$

Together with j, e indicates whether or not a consumer is in college. Utility 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 is included in total household consumption with an adult equivalence parameter ζ. College students, for whom $e = h$ and $j \in \{1, 2, 3, 4\}$, are subject to an effort cost net of college consumption value, λ.

Income Age, education, skill, stochastic earnings productivity, and net assets, summarized by the tuple (j, e, s, η, a), determine income, y, given by

$$y_{j,e,s,\eta,a} = [w_\ell \epsilon_{j,\ell,s} \ell_{pt} \mathbb{I}_{j\le 4} \mathbb{I}_{e=h} + w_e \epsilon_{j,e,s} \mathbb{I}_{j>4 \text{ or } e=\ell} \mathbb{I}_{j<j_r}]\eta + ss_{e,s}\mathbb{I}_{j\ge j_r} + r\left[a\mathbb{I}_{j>1}\mathbb{I}_{a>0} + Tr_j\right] \tag{11}$$

where w_e is the wage rate that depends on completed education, $\epsilon_{j,e,s}$ is a deterministic life cycle productivity that depends on age, completed education, and skill, ℓ_{pt} is part-time hours, and $ss_{e,s}$ is the Social Security transfer that depends on completed education and skill as defined in equation (23) in Supplementary Appendix B.2.[34]

Income tax The income tax function follows the specification from Heathcote, Storesletten, and Violante (2017) and is given by

$$T(y) = y - \gamma y^{1-\tau_p} \tag{12}$$

[34]The indicator $\mathbb{I}_{j>1}\mathbb{I}_{a>0}$ implies that interest income on the inter vivos transfer accrues to the parents and not the newly emancipated child aged $j = 1$. This is a neutral assumption about timing: If the interest accrued to the child instead, the parent simply would choose an alternative b so that the net amount transfered to the child was the same.

where τ_p governs the tax progressivity and γ is used to balance the government budget constraint in every period as shown in equation (25) in Supplementary Appendix B.2.

Technology The production function for capital and aggregate labor is Cobb-Douglas, given by

$$K^{\alpha}(ZL)^{1-\alpha} \tag{13}$$

where K is aggregate capital stock, Z is aggregate labor productivity, L is aggregate labor, and α is the capital share. The capital stock depreciates at rate δ. Aggregate labor is a composite of efficiency units of labor with low education, L_ℓ, and efficiency units of labor with high education, L_h, given by

$$(\nu L_\ell^{\iota} + (1-\nu)L_h^{\iota})^{1/\iota} \tag{14}$$

where $1/(1-\iota)$ is the constant elasticity of substitution and ν is a share parameter.

4 Model Parameterization

The parameters of this model are divided into those estimated outside of the model, shown in Tables 7 and 8, and those calibrated inside of the model, shown in Table 9.

Table 7 presents externally estimated parameters related to education. Panel A begins with parameters governing the federal student loan program: first, the aggregate federal student loan limit, $\bar{A}$, is set to the current cumulative borrowing limit for four years of college, normalized by the average annual net tuition and fees plus room and board based on Smole (2019) and NCES (2019);[35] second, the add-on for the federal student loan interest rate, τ_{SL}, is set to the most recent value of 2.1 percentage points as reported by the Chief Operating Officer for Federal Student Aid (FSA) in Chief Operating Officer for FSA (2021); third, the number of years for repayment on a student loan, T_{SL}, is set to 10 based on Smole (2019);[36] fourth, the garnishment rate conditional on delinquency for both federal and private student loans, τ_g, is set to 15 percent, as reported in Yannelis (2020);[37] and, fifth, the student loan collection fee, ϕ_D, is set to 0.185 following Luo and Mongey (2019). The last row of Panel A reports working hours while in college, ℓ_{pt}, set to the average

[35] This limit has been in place since July 1, 2012. The U.S. federal student loan program sets yearly limits and lifetime limits on borrowing. Yearly limits depend on one's academic year (e.g., freshman) and dependency status. We assume borrowers are dependents because most undergraduate students are less than 24 years old, and use the cumulative limit over the first four years because college in our model lasts for four years.

[36] In the U.S., those with student loans may choose between a standard repayment plan of 10 years and an income-based repayment plan, which may have a repayment time frame ranging from 10 to 25 years.

[37] We set the garnishment rate for private loans equal to the garnishment rate for federal loans. This is consistent with the U.S. system, where garnishment is allowed for delinquent private loans as long as the loan provider obtains a court order.

weekly time spent working for third-year college students in the HSLS:09, expressed as a fraction of full-time work, as reported see Table 26 in Supplementary Appendix A.2.1.

Table 7: Externally estimated parameters related to education

Symbol	Parameter description	Data source	Parameter value
Panel A: Federal student loan program and college working hours			
$\bar{A}$	Limit	Smole (2019) and NCES (2019)	1.493
τ_{SL}	Interest rate add-on	Chief Operating Officer for FSA (2021)	0.021
T_{SL}	Maximum years to repay	Smole (2019)	10
τ_g	Federal SL garnishment rate	Yannelis (2020)	0.150
ϕ_D	Student loan collection fee	Luo and Mongey (2019)	0.185
ℓ_{pt}	Working hours while in college	HSLS:09	0.347
Panel B: Grant tuition subsidies, by skill endowment s			(s_1 , s_2 , s_3)
$\theta(s)$	Public tuition subsidy	HSLS:09 and Krueger and Ludwig (2016)	(0.285, 0.323, 0.364)
$\theta^{pr}(s)$	Private tuition subsidy		(0.122, 0.139, 0.156)
Panel C: Child skill distribution given parent education, by child skill endowment s_c			($s_{c,1}$, $s_{c,2}$, $s_{c,3}$)
$\pi_{s_c}(s_c\|e=\ell)$	Parent does not have BA	HSLS:09	(0.426, 0.341, 0.233)
$\pi_{s_c}(s_c\|e=h)$	Parent has BA		(0.176, 0.311, 0.512)
Panel D: Distribution of subjective beliefs, by skill endowment s			(s_1 , s_2 , s_3)
$\pi_{\hat{p}}(\hat{p}\|s)$	Mass in each subjective belief bin	NLSY97	Panel A of Table 16
$\hat{p}_1(s)$	Annualized subjective belief: probability		(0.449, 0.473, 0.411)
$\hat{p}_2(s)$	of being allowed to continue in college		(0.704, 0.692, 0.702)
$\hat{p}_3(s)$			(0.838, 0.838, 0.836)
$\hat{p}_4(s)$			(0.921, 0.923, 0.924)
$\hat{p}_5(s)$			(0.988, 0.990, 0.993)

Panel B of Table 7 reports the estimated share of tuition and fees paid with grants and scholarships from public sources, $\theta(s)$, and private sources, $\theta^{pr}(s)$. To assign these values, using data from the HSLS:09 we first express grants from any source as a share of tuition and fees for each skill tercile. Next, we multiply the total share of tuition subsidized by grants by 0.7 to assign values to $\theta(s)$ and assign the residual to $\theta^{pr}(s)$, incorporating estimates from Krueger and Ludwig (2016) on grants from public versus private sources.

Panel C of Table 7 reports the conditional distribution of child skill given parental education, $\pi(s_c|e)$. Note that the parameterized model reflects the fact that, in the HSLS:09, parent education and child high school GPA are positively correlated. Panels B and C draw on HSLS:09 findings reported in Table 25 of Supplementary Appendix A.2.1.

Panel D of Table 7 reports parameters governing the conditional distribution of subjective beliefs from which $\hat{p}$ is drawn. Specifically, for each skill bin, we discretize the distribution of beliefs into five bins of equal width (0-19, 20-39, etc.). The mass of responses in each element of this grid, $\pi_{\hat{p}}(\hat{p}|s)$, is estimated directly from the data and is reported in Panel A of Table 16 in Supplementary Appendix A.1. To estimate the five grid-point values of each conditional distribution, $\hat{p}_1(s)$ through $\hat{p}_5(s)$, we set each annual probability as a value between zero and one so that over four years it is equal to the average expected probability in the same bin reported in the NLSY97. The empirical targets for these values are reported in Panel B of Table 16 of Supplementary Appendix A.1.

This parameterization approach implements the following logic. Imagine that one is collecting a "survey" in the model that asks the same question as the NLSY97 about the likelihood of earning a bachelor's degree. The goal is to construct a model statistic that we map to the data on subjective beliefs. We make two assumptions about the behavior of survey respondents in the model. First, respondents report the likelihood of earning a four-year BA conditional on enrollment. This (conservative) assumption is the same as our treatment of the NLSY97 beliefs data in the discussion of Tables 2 and 3 of Section 2.1. Second, respondents lie in the model survey and over-report the expected likelihood of graduation by ignoring that dropout may arise endogenously. This assumption introduces an upward social desirability bias in the reported graduation likelihoods. Of course, in the model, when 18-year-olds decide whether or not to enroll they incorporate endogenous dropout into their value of college. Only their responses to the model survey are inflated by social desirability bias.

In the discussion of our empirical findings in Section 2.1, we pointed out that the observed difference between expected and realized graduation likelihoods could be inflated by bias in survey responses. The two assumptions that we apply to parameterize subjective beliefs in the model lead to conservative estimates for the true extent of optimism about graduation, which we do not measure directly in the data.[38] In model validation exercises presented in Section 5.2, we show that our mapping of the model to the NLSY97 data on subjective beliefs introduces a reasonable role for reported beliefs in determining college enrollment within the model, compared to our regression results in Table 1. We also show that the model exhibits a reasonable elasticity of enrollment to college tuition subsidies compared to estimates in the literature. In Table 33 of Supplementary Appendix C.1, we show that the model matches the extent of optimism and pessimism by enrollment status and skill tercile in the NLSY97, as reported in Tables 2 and 3. These moments are not targeted directly in our calibration because we did not calibrate the subjective beliefs by enrollment status.

Table 8 presents externally estimated parameters unrelated to education. Panel A governs demographics: the fertility period, j_f, is set to 13 so that consumers have a child when they turn 30; the age adulthood begins, j_a, is set to 18; j_r is chosen so that the retirement age is 65; and, finally, J sets maximum life span to 100 years. For $j < j_f + j_a$, we set survival probabilities ψ_j to one to rule out children without parents; ages $j \geq j_f + j_a$ use estimates from Bell and Miller (2020). Panel B, which covers preferences and technologies, begins with the relative risk aversion param-

[38] In the model, if we shut off endogenous college dropout, a consumer at the time of enrollment is optimistic about graduating college as long as $\hat{p}^4 > \prod_{j=1}^{4} p_c(j,s)$. For pessimism, the condition is $\hat{p}^4 < \prod_{j=1}^{4} p_c(j,s)$. With endogenous college dropout, even with social desirability bias the extent of optimism or pessimism about the likelihood of graduation cannot be identified using only $p_c(j,s)$ and the reported belief $\hat{p}$.

eter, σ, set to 2 based on Chetty (2006). The adult equivalence scale, ζ, is set to 0.3 following the Organization for Economic Co-operation and Development (OECD) modified scale. The capital share parameter, α, is set to 0.36 following Kydland and Prescott (1982). The depreciation rate of capital, δ, is set to 0.076, as in Krueger and Ludwig (2016). The parameter that dictates the elasticity of substitution between low- and high-education labor, ι, is set to 0.8, which implies an elasticity of substitution of 5. This value is in the middle of the range (between 4 and 6) reported in Card and Lemieux (2001) after controlling for imperfect substitutability across age groups. In Appendix C.3, we perform sensitivity analyses with a higher and lower value for the elasticity of substitution. Life cycle productivities $\epsilon_{j,e,s}$, are estimated and reported in Table 23, Supplementary Appendix A.1.2. Panel C contains government policy parameters: the consumption tax rate τ_c is set to 5 percent (Krueger and Ludwig, 2016), the progressivity of the income tax function, τ_p, is set to 0.177 following our estimation presented in Table 30 in Supplementary Appendix A.4, and government consumption as a share of GDP, g, is set to 0.141 using estimates from the Bureau of Economic Analysis (BEA) in BEA (2022, T1.1.5) and BEA (2022, T3.1).

Table 8: Externally estimated parameters not related to education

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	Bell and Miller (2020)	-
Panel B: Preferences & technology			
σ	Risk aversion	Chetty (2006)	2
ζ	Adult equivalence scale	OECD modified scale	0.3
α	Capital share	Kydland and Prescott (1982)	0.360
δ	Depreciation rate	Krueger and Ludwig (2016)	0.076
ι	Elasticity of substitution	Card and Lemieux (2001)	0.800
$\epsilon_{j,e,s}$	Earnings life cycle profile	Table 23	-
Panel C: Government			
τ_c	Consumption tax rate	Krueger and Ludwig (2016)	0.050
τ_p	Income tax progressivity	Table 30	0.177
g	Government consumption	BEA (2022, T1.1.5) and BEA (2022, T3.1)	0.141

Table 9 reports internally calibrated parameters.[39] The first column contains the parameter symbol; the second column, the parameter description; and the third column, the parameter value. Columns 4 through 6 contain the target moment's description, the moment in the data, and the moment in the calibrated model, respectively. Panel A of Table 9 presents parameters governed by moments from the HSLS:09. The first two objects are $p(s)$, which determines the true average probability of being allowed to continue in college, and $\rho_d(s)$, which determines the persistence of the true

[39] Although parameters and moments are grouped in Table 9 using the most significant one-to-one relationship between each parameter and target moment, and are discussed accordingly, the parameters are calibrated jointly and each parameter can affect all target moments.

exogenous dropout probability. Together, $p(s)$ and $\rho_d(s)$ determine the true probability of being allowed to continue in each academic year of college, $p_c(j, s)$. These objects are governed by persistence rates to the end of the third academic year (Y3), given enrollment in a four-year degree (Y1) and persistence to the end of the third academic year conditional on persisting to the second academic year (Y2), respectively, which are reported in Table 25 in Supplementary Appendix A.2.1. We cannot externally estimate $p(s)$ and $\rho_d(s)$ from the data because of the endogenous dropout decision.[40] The last two rows in Panel A contain the fixed utility costs for taking out any student debt, ξ_L, and for taking out private student loans specifically, ξ_L^{pr}. The values of ξ_L and ξ_L^{pr} are set so that the model matches the share of 2013 college enrollees who have any student loan debt and any private student loan debt, respectively, using the empirical moments reported in Table 5 of Section 2.3.[41]

Table 9: Internally calibrated parameters

Symbol	Parameter description	Parameter value	Moment description	Data moment	Model moment
Panel A: Moments from the HSLS:09					
$p(s)$	Continuation prob. average	(0.642,0.824,0.908)	Persist to Y3 | Y1	*(0.476,0.711,0.829)*	*(0.474,0.711,0.829)*
$\rho_d(s)$	Dropout prob. persistence	(0.631,0.779,0.940)	Persist to Y3 | Y2	*(0.774,0.863,0.913)*	*(0.774,0.863,0.913)*
ξ_L	Loan search cost	0.000	Loan uptake	*0.650*	*0.584*
ξ_L^{pr}	Private loan uptake cost	3.146	Private loan uptake	*0.220*	*0.221*
Panel B: Moments from the NLSY97					
λ	Net college effort cost	-0.169	Enr. by age 25	*0.478*	*0.478*
$q(s)$	Enrollment option shock	(0.508,0.777,0.912)	Enr. by age 25 | High fam. inc.	*(0.318,0.659,0.883)*	*(0.318,0.659,0.883)*
β_c	Parent altruism toward child	0.199	Ave. transfer, normalized	*0.578*	*0.578*
ν	Low-education labor share	0.523	College wage premium | s_2	*1.410*	*1.411*
Panel C: Moments from other sources					
$\bar{c}$	College room and board	0.147	Room + board, normalized	*0.147*	*0.147*
κ	Annual tuition	0.173	Net tuition + fees, normalized	*0.088*	*0.088*
$\bar{y}$	Garnishment-exempt income	0.151	Exempt earnings, normalized	*0.151*	*0.151*
ξ_D	Federal delinquency cost	0.181	Federal delinquency rate	*0.090*	*0.085*
ξ_D^{pr}	Private delinquency cost	1.195	Private delinquency rate	*0.074*	*0.075*
τ_{is}	Student loan issuance cost	0.038	Interest rate comparison	-	*0.065*
Z	Aggregate labor productivity	0.608	GDP per capita 18+	*1.000*	*1.000*
β	Discount factor	0.972	Capital-to-output ratio	*3.000*	*3.000*
χ	SS replacement rate	0.187	SS expenditure, fraction of GDP	*0.048*	*0.048*

Panel B of Table 9 reports parameters that are governed by moments from the NLSY97. The college effort cost net of the consumption value of college, λ, is determined by observed college enrollment rates by age 25 (Table 17, Supplementary Appendix A.1.1). The college enrollment option shock, $q(s)$, is chosen to target enrollment rates for the top family income tercile for each skill bin (Table 18, Supplementary Appendix A.1.1). The enrollment option shock captures academic, personal, or family reasons that lead 18-year-olds to not enroll in college. By focusing on enrollment rates by skill tercile for those from high-income families, we mitigate the role of financial constraints in the data. The parameter β_c, the degree of a parent's altruism toward their

[40] In our baseline calibration, for a given cohort, total endogenous drop outs are 0.8 percent of total dropouts.

[41] Even with a value of 0 for ξ_L, the model somewhat understates the overall student loan uptake rate. In Table 37 in Supplementary Appendix C.3, we compare the student loan portfolio in the model with its data counterpart, and also consider an alternative calibration where we choose ξ_L^{pr} to target the share of students with only private loans.

child, is set so that the model matches average parent-to-child transfers (Table 21, Supplementary Appendix A.1.1), where the normalizing GDP per capita for those 18 and over covers 2016-2018 from BEA (2022, T1.1.5). The parameter that determines the labor share for low-education labor, ν, is set so that the college wage premium for the middle skill tercile matches that observed in the data as reported in Table 24 in Supplementary Appendix A.1.2. Table 32 in Supplementary Appendix C.1 shows that the resulting college wage premium in the model aligns well with its empirical counterpart for all skill terciles.

Panel C of Table 9 contains parameter governed by moments from other sources. The cost of college room and board, $\bar{c}$, is set using the average annual value for room and board, and annual tuition, κ, targets average net tuition and fees; both empirical moments are for bachelor's degree programs from 2016-2018 as reported in NCES (2019) and are normalized with GDP per capita for those 18 and over during the same period. The income exempt from garnishment in delinquency, $\bar{y}$, is set to 15.1 percent of GDP per capita for the population 18 and over, based on our calculations using results from Yannelis (2020). The parameter governing the costs of being delinquent on public loans, ξ_D, is set so that the model's delinquency rate matches the average cohort delinquency rate from 2016 to 2018 reported in FSA (2021b), where the definition of delinquency in the data is a delay in payment of 270 days or more. The delinquency cost for private loans, ξ_D^{pr}, is set so that the model matches private loan balances 90 or more days delinquent as a fraction of total private loan balances in repayment for 2016-2018 as reported in Amir, Teslow, and Borders (2021).[42] The student loan issuance cost, τ_{is}, is set so that the interest rates of federal and private student loans have the same mean, as documented in Table 29 of Supplementary Appendix A.3.[43] 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). Finally, the Social Security replacement rate, χ, targets the average ratio of total Social Security expenditure to GDP from 2016 to 2018, as measured in BEA (2022, T2.1) and BEA (2022, T1.1.5).

[42]For federal student loans, after 270 days spent in delinquency, the loan is in default. Since the model period is one year, we use the cohort default rate (270 or more days delinquent) as the empirical target for the per-period delinquency rate. For private loans, we use the available delinquency definition (90 or more days) closest to the length of a period in our model when selecting the empirical target.

[43]The loan issuance cost on federal loans affects the government budget constraint through the loan program embedded in equation (25) and the loan issuance cost for private loans affects the present value of net revenue flows for the private lender of equation (22). Both expressions are presented in Supplementary Appendix B.2.

5 Properties of the Baseline Equilibrium

This section presents properties of the initial stationary equilibrium (the "baseline") that are related to our main experiment, as well as model validation results and the model's fit of untargeted federal loan utilization rates. In Section 5.1, we examine the role of subjective beliefs in generating enrollment patterns in the baseline equilibrium.[44] In Section 5.2, we perform two validation exercises in which we compare the model's enrollment responsiveness to subjective beliefs and tuition subsidies with empirical counterparts. Finally, in Section 5.3, we examine the model's fit of federal student loan limit utilization rates in the HSLS:09. The high utilization rates we find in both the data and the model, along with our evidence on optimistic subjective beliefs, motivate the policy experiment of a federal student loan limit expansion studied in Section 6.

5.1 The effect of beliefs on enrollment patterns

Column (1) of Table 10 reports enrollment rates for each skill tercile in the NLSY97 (see Table 17 of Supplementary Appendix A.1 for details), while column (2) reports enrollment rates in the model's calibrated baseline. Moments from the model align reasonably well with the data.[45] Column (3) of Table 10 reports counterfactual enrollment rates for the same distribution of high school graduates as column (2), in a partial equilibrium when we shut off subjective beliefs by setting $\hat{p} = p_c(j, s)$ for all j and s but do not allow general equilibrium objects to adjust. Without subjective beliefs, enrollment rates decrease among low- and medium-skill 18-year-olds, whereas enrollment rates increase among the high-skill 18-year-olds. Motivated by these observations, we introduce the concept of over- and under-enrollment. An enrollee is counted as "over-enrolled" if they enroll with subjective beliefs but would not with correct beliefs. Analogously, a non-enrollee is counted as "under-enrolled" if they do not enroll with subjective beliefs but would do so with correct beliefs. Columns (4) and (5) of Table 10 report the mass of 18-year-olds who are over- and under-enrolled as a share of enrollees and non-enrollees, respectively. These model statistics highlight the impact of optimism and pessimism on college enrollment patterns in the model's initial stationary equilibrium. For college enrollees in the lowest skill bin, over-enrollment is especially high due to optimism; for non-enrollees in the highest skill bin, under-enrollment is high due to pessimism.

[44] In Supplementary Appendix C.1, we examine the role of subjective beliefs in generating borrowing behavior in the baseline.

[45] The skill-specific enrollment option shock, $q(s)$, is calibrated to match enrollment rates by skill for the top family income tercile; we chose this approach because $q(s)$ represents non-financial reasons that prevent enrollment, and at the top income tercile the role of financial constraints is minimal. We have also examined a calibration in which $q(s)$ is set to match the overall enrollment rates by skill; the main conclusions do not change.

Table 10: College enrollment statistics by skill

Statistic: Sample:	College enrollment rates High school graduates			Over-enrollment College enrollees	Under-enrollment Non-enrollees
Skill	**(1) Data**	**(2) Baseline**	**(3) Baseline, corrected beliefs**	**(4) Baseline**	**(5) Baseline**
1	22.92	13.65	6.51	56.78	0.71
2	45.57	47.52	37.59	24.98	3.70
3	77.01	85.99	91.18	0.00	37.06

Notes: Table 10 presents enrollment statistics in the data and model by skill, where skill bin is assigned with high school GPA tercile in the data and represented with s in the model. Enrollment rates are computed after high school graduation as percentages of the skill bin who enroll in a BA program. Columns (1), (2), and (3) report the enrollment rates in the data, in the initial stationary equilibrium of the model, and when $\hat{p} = p_c(j, s)$, so that there is no optimism or pessimism and consumers have correct beliefs, but general equilibrium objects are not allowed to adjust; columns (4) and (5) report the over- and under-enrollment as a share of enrollees and non-enrollees, respectively. The definitions of over- and under-enrollment are provided in the main text. All units are in percentages. Data source: NLSY97.

5.2 Model validation of enrollment responsiveness to beliefs and tuition

The first row of Table 11 compares data and model enrollment responsiveness to subjective beliefs in the cross-section (second and third column, respectively), where the model moment is computed using the coefficient on beliefs after estimating the regression model (2) of Table 1 in Section 2.1 on model output. Specifically, this exercise predicts the likelihood of college enrollment using output from the structural model, where the independent variable is reported beliefs in the model "survey," after controlling for family and individual characteristics such as the child's skill, family income, and parental education. The model performs reasonably well in matching this untargeted moment, with an estimated coefficient on beliefs of 0.56 compared to the empirical estimate of 0.47 percentage points. This result suggests that our model produces a quantitatively reasonable role for subjective beliefs in determining college enrollment choices.

The second row of Table 11 reports enrollment responsiveness to tuition subsidies using the change in the enrollment rate for an additional \$1,000 tuition subsidy (i.e., a quasi-experiment), in the data and in the model in partial equilibrium. The data estimate is from Deming and Dynarski (2009), who survey the literature on empirical estimates for enrollment responses to tuition subsidies, and conclude that the best estimates suggest a value of 4 percentage points. The model does remarkably well in matching this untargeted moment, with a response of 3.61 percentage points. This result suggests that our model produces a quantitatively reasonable role for college costs in determining college enrollment decisions.

Table 11: Model validation experiments

Experiment	Data	Model
Coefficient on subjective beliefs when predicting enrollment likelihood	0.47	0.56
Enrollment change due to additional $1,000 tuition subsidy	4.00	3.61

Notes: Table 11 reports empirical and model estimates for two model validation experiments. The first row reports the coefficient on reported beliefs as a predictor of college enrollment. The data coefficient is from model (2) of Table 1 in Section 2.1; the model coefficient results from a regression on model output in which the dependent variable is the enrollment decision (100 for an individual that enrolls and 0 otherwise) and the independent variable is the reported belief as well as individual and family characteristics (i.e., child skill, logged family income, and parental education). The second row reports changes in the enrollment rate given a $1,000 tuition subsidy increase, where the data moment is reported by Deming and Dynarski (2009) and model moment is constructed in partial equilibrium. All units are in percentage points.

5.3 Federal student loan limit utilization rates

Consistent with recent federal student loan policy, the model's federal student loan limit is enough to pay for 1.49 years of average total college costs (or 37.5 percent of annual costs). To what extent are college students using the federal loans to which they have access in the data, and how does the model perform in matching utilization rates? To measure utilization rates in the data, we turn to the HSLS:09.[46] We compute the federal loan utilization rate for college enrollees who persist for three academic years after enrollment, where the utilization rate is the ratio of the cumulative federal debt balance to cumulative borrowing limits after the first three years of college (in 2016). The results are reported in Table 12: 54 percent of those who completed their third academic year utilized more than half of their cumulative federal student loan limit, 34 percent utilized more than 90 percent, and 28 percent utilized all of their available federal loans at the end of their third academic year. Although these moments are not targeted in the calibration, Table 12 indicates that the model's baseline equilibrium also exhibits a sizable share of students using all of their available federal loans. Because we underestimate this share in the model baseline, our welfare estimates from loan expansions can be considered lower bounds.

[46]To apply for federal aid, college students submit the Free Application for Federal Student Aid (FAFSA). Students select a dependency status on the FAFSA, which determines annual borrowing limits for federal student loans. The public version of the HSLS:09 does not report which dependency status each FAFSA filer selects. We assume everyone files as dependents because most undergraduate students (and all students in the HSLS:09 in 2016, the year in which we measure their utilization rates) are less than 24 years old. Besides age, other ways to be classified as independent are to be married, enroll in a graduate program, serve on active duty in the U.S. armed forces or be a veteran, have dependent children, have deceased parents, be an emancipated minor, or be determined as an unaccompanied minor (FSA, 2022b). Most undergraduate students do not satisfy these criteria.

Table 12: Utilization rates for federal student loans

Utilization	Data	Model
$\geq 50\%$	54	44
≥ 90 %	34	19
$\geq 100\%$	28	16
Obs	1,855	

Notes: Table 12 reports utilization rates for federal student loans in the data and in the baseline model equilibrium. Data moments are estimated for students who enrolled in a BA program in the fall of 2013 and persisted to the end of their third academic year. For data and model moments, utilization rates of federal student loans are computed as percentages of the cumulative limit up to that point (the sum of annual limits for the first three academic years). Data moments use PETS-SR student records longitudinal weights. Data source: HSLS:09.

6 Main Experiment: Federal Loan Limit Expansion

This section analyzes the effects of expanding the federal student loan limit to four years worth of college tuition, net of grants, plus room and board by setting $\bar{A} = 4$; this expansion allows federal student loans to finance 100 percent of annual college costs. The welfare consequences of a federal student loan limit expansion are ex-ante ambiguous: access to more federal credit potentially worsens over-enrollment for optimistic high school graduates while also relaxing a binding constraint, with the parameterized model determining the relative magnitude of each of these forces.[47] To highlight the role of subjective beliefs, for illustrative purposes we also include results from a limit expansion in an economy without subjective beliefs that is re-calibrated to match the same set of target moments as the baseline (except for moments related to subjective beliefs).

Welfare To measure welfare, we assume that the social planner knows the true payoff of choices but internalizes that the consumer has subjective beliefs when making the college enrollment decision, the inter vivos transfer decision, and the decisions leading up to and including the age at which the inter vivos transfer is made (i.e., it is a "paternalistic government"). We compute welfare for 18-year-old consumers before they make the college enrollment decision; this age group is the one most affected by the policy change. We report two welfare statistics: first, the population share that is strictly worse off, calculated by comparing the lifetime values as computed by the social planner; and, second, consumption-equivalent variation, calculated as explained in Supplementary Appendix B.4.[48]

[47] Of course, the federal loan limit expansion also allows pessimistic students to access more federal credit, although the empirical results of Section 2.1 indicate that this group is small. We focus on the larger effects of optimism but report welfare for students across the whole distribution of subjective beliefs.

[48] Our consumption-equivalence calculations in Supplementary Appendix B.4 account for the presence of utility costs by following the method of Abbott, Gallipoli, Meghir, and Violante (2019). Specifically, in all equilibria, we compute a consumption-equivalence shifter that equalizes expected lifetime utility to that equilibrium's expected value of

6.1 Partial equilibrium analysis

The first row of Table 13 reports a key result for this paper: in partial equilibrium, one third of the low-skill and a smaller share of the medium-skill are strictly worse off when federal loan limits expand, while in a re-calibrated model without subjective beliefs, no one is strictly worse off. The distribution of subjective beliefs, which is a feature of the data that we use to inform our model, drives this result.

In particular, these partial-equilibrium welfare losses arise because the limit expansion increases over-enrollment: young people who did not enroll in the baseline now do so entirely because of their optimistic beliefs. In Supplementary Appendix C.2.1, we show that, if we shut off parental altruism, transitioning from being a non-enrollee to being an over-enrolleed college student after the loan limit expansion is both sufficient and necessary to suffer welfare losses after the policy change. With altruism, we establish that this mechanism is quantitatively the main driver of partial-equilibrium welfare losses in our calibrated baseline model with subjective beliefs. Specifically, among those that are strictly worse off, 100 percent are non-enrollees in the baseline equilibrium economy that transition to being over-enrolled in the equilibrium with higher federal loan limits; among those that transitioned from non-enrollees in the baseline equilibrium to being over-enrollees, 99.38 percent are strictly worse off.

Table 13: Share of 18-year-olds that are strictly worse off

	(I) Baseline				**(II) No subjective beliefs**			
		Skill				**Skill**		
Equilibrium	**All**	**Low**	**Medium**	**High**	**All**	**Low**	**Medium**	**High**
Partial	12	32	3	0	0	0	0	0
General	34	34	36	30	18	0	9	35

Notes: Table 13 reports the share of 18-year-olds that are strictly worse off, overall and for each skill endowment, after the federal loan limit expansion in our model with subjective beliefs ("Baseline" columns) and in an alternative recalibrated framework without subjective beliefs ("No subjective beliefs" columns). Rows determine the equilibrium concept being applied: "Partial" refers to 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; "General" refers to general equilibrium. For the welfare comparison in general equilibrium, the final steady state distribution of 18-year-olds was used. The share of the population that is strictly better off is the reciprocal of those that suffer losses (that is, no 18-year-old is indifferent).

non-enrollment; we use the non-enrollment value function as a within-equilibrium baseline because it is free from utility costs. When reporting changes for a given skill level and family income tercile, we use the expected lifetime utility and expected value of non-enrollment conditioning on those attributes to compare with the within-equilibrium baseline. We then take the difference in these equilibrium-specific shifters across equilibria. The result is our consumption-equivalent estimates of welfare changes relative to the baseline stationary steady-state, which have the property that positive values indicate gains and negative values indicate losses.

In columns (3), (4), and (5) of Table 14, we quantify the magnitudes of welfare changes using consumption-equivalent variation in partial equilibrium. Here, welfare estimates are reported by the skill, family income, and the subjective beliefs bin of the 18-year-old. The largest welfare losses, 1.15 percent of lifetime consumption, are experienced by low-skill young adults from poor families who have high expectations about their own likelihood of BA attainment. This subgroup is not small: they account for 6.52 percent of all 18-year-olds. The largest welfare gains, between 4.88 and 5.45 percent of lifetime consumption, are experienced by high-skill young adults from poor families who have moderate to high expectations about their own likelihood of earning a BA. This subgroup is not small either: they account for 8.58 percent of all 18-year-olds. The pattern of welfare implications for young people from higher-income families is qualitatively similar to those from low-income families, although the magnitudes are smaller.[49]

To summarize, in partial equilibrium the presence of subjective beliefs (which exhibit optimism for many young people) leads to welfare losses for roughly one third of low-skill 18-year-olds after a limit expansion. By contrast, no one is worse off after such a policy change in a model environment without subjective beliefs. Furthermore, the magnitudes of the welfare losses can be larger than 1 percent of lifetime consumption for some subgroups (e.g., those with low skill those from low-income families who have high subjective expectations). In the next section, we analyze how general equilibrium adjustments affect these findings.

Table 14: Consumption-equivalent variation for 18-year-olds by skill, family income, and beliefs

		(I) Partial equilibrium			**(II) General equilibrium**		
			Skill			**Skill**	
Family inc. tercile	**Exp. prob. BA**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**
1	*0 to 39*	0.10	0.20	0.37	0.68	0.72	0.92
	40 to 79	-0.13	0.65	5.45	0.68	0.74	3.93
	80 to 100	-1.15	2.03	4.88	-1.10	1.28	3.61
2	*0 to 39*	0.10	0.13	0.23	0.68	0.70	0.78
	40 to 79	-0.14	0.47	2.69	0.69	0.45	1.08
	80 to 100	-0.66	1.04	2.43	-0.75	0.06	0.90
3	*0 to 39*	0.09	0.09	0.11	0.73	0.66	-0.04
	40 to 79	-0.02	0.17	0.34	0.62	-0.18	-0.91
	80 to 100	-0.01	0.19	0.27	-0.04	-0.57	-1.09

Notes: Table 14 reports consumption-equivalent variation estimates in percentage points for 18-year-olds by skill, family income, and subjective beliefs of BA attainment likelihood in the baseline in partial and general equilibrium. In partial equilibrium, the income tax rate, prices, bequests, Social Security transfers, and the 18-year-old distribution are fixed at their initial steady state values. In general equilibrium, the aforementioned objects are allowed to adjust, and we compare the initial steady state value to the corresponding final steady state value in each skill, family income, and subjective beliefs bin.

[49] Our results also highlight that the gains from a loan limit expansion can be small for those with high skill from low-income families, provided that they have low expectations about the likelihood of BA attainment.

6.2 General equilibrium analysis

General equilibrium adjustments act to dampen the increase in the value of college from a limit expansion and increase the value of not going to college—Table 35 of Appendix C.2.2 reports details on the resulting changes to changes to education and skill statistics, macroeconomic aggregates, and general equilibrium objects. In particular, the wage rate and Social Security transfers for a low-education worker increase, whereas they decrease for a high-education worker; the risk-free rate interest rate on savings increases, which increases the interest rate for federal student loans; the income tax rate increases, which increases the marginal tax rate for a high earner; parental transfers for college decrease; and accidental bequests increase. In light of these changes, we next examine how the welfare effects of a limit expansion are affected by general equilibrium adjustments, with a focus on the implications of widespread optimism due to mistaken subjective beliefs.

The second row of Table 13 reports the share of 18-year-olds that are strictly worse off after a limit expansion in general equilibrium, both overall and for each skill endowment. As with the partial equilibrium analysis, we contrast this statistic for our model with subjective beliefs to one computed in a re-calibrated model without subjective beliefs. In the baseline general equilibrium economy, roughly one third of the population is strictly worse off across all skill endowment levels, whereas in partial equilibrium, as discussed above, those strictly worse off were almost entirely from the lowest skill level. A rise in the share of the high-skilled who suffer welfare losses in general equilibrium is also seen in the framework with correct beliefs. This common pattern across the two model environments is due to a decrease in the value of a high-education worker (which is the most likely outcome for an 18-year-old with high skill) relative to the initial equilibrium. In Supplementary Appendix C.2.3 we show that, among the general equilibrium objects discussed in the preceding paragraph, the fall in the wage rate for high-education workers is the primary driver of welfare losses for the consumers with the highest skill endowment level. An additional insight from Table 13 is that the presence of subjective beliefs almost doubles the share of all 18-year-olds that are strictly worse off in general equilibrium after the loan limit expansion, from 18 to 34 percent. This result highlights that optimism is a quantitatively important rationale for why a policy maker may not want to increase federal loan limits to fully fund college for everyone.

In Table 14, columns (6) to (8) show the magnitudes of welfare changes for 18-year-olds in general equilibrium by skill, family income, and subjective expectation bin. The magnitudes of these general-equilibrium changes tend to be lower in absolute value than their partial-equilibrium counterparts, although the changes are usually not large. For example, those from low-income families with low skill and high expectations about their likelihood of BA attainment experience losses worth 1.10 percent of lifetime consumption, slightly smaller than the partial equilibrium estimate of 1.15 percent. In Appendix C.2.4, we analyze welfare implications for these subgroups along the transition path. The main takeaways are unchanged.

In order to highlight the role that subjective beliefs play in driving the general-equilibrium welfare changes of Table 14, in Table 15 we compare welfare changes for the group with the largest losses in general equilibrium with subjective beliefs to welfare changes for the analogous group in an environment with correct beliefs. In our baseline model with subjective beliefs, a low-skill 18-year-old from a low-income family with high expectations about the likelihood of earning a BA experiences losses worth 1.10 percent of lifetime consumption. Without subjective beliefs, low-skill 18-year-olds from low-income families see welfare gains of 0.95 percent. For this sort of 18-year-old, ignoring subjective beliefs leads one to overestimate gains by 2.02 percentage points.

Table 15: Consumption-equivalent variation: lowest skill and income, highest expectations

Baseline model	**No subjective beliefs**	**Difference**
-1.10	0.95	2.02

Notes: Table 15 reports consumption-equivalent variation estimates in units of percentage points for 18-year-olds from the lowest skill bin and lowest family income tercile with the highest expectations about BA attainment probability (80 to 100) in the baseline model economy, as well as the same welfare statistic for 18-year-olds from the lowest skill bin and lowest family income tercile in a re-calibrated economy without subjective beliefs. Note that, without subjective beliefs, within a skill bin all beliefs are the same. To compute welfare, we compare the initial steady state value to the corresponding final steady state value conditioning on skill, family income, and beliefs bin.

7 Conclusion

In this paper, we document empirically that both young adults and their parents exhibit subjective beliefs about the likelihood of earning a bachelor's degree that positively predict college enrollment and that are often optimistic. We build a structural model of college choice that features these subjective beliefs, and also places discipline on key sources of financing college, especially private student loans, to examine the welfare effects of a federal student loan limit expansion. We find welfare losses for low-skill young adults from poor families who exhibit a high degree of optimism, because access to more federal student loans worsens over-enrollment in college (that is, enrollment due to optimistic beliefs) for these consumers. Many questions, including the implications of subjective beliefs for the design of loan repayment policies, remain for future research. We hope that the empirical findings and quantitative analysis we present here will be useful for researchers seeking to evaluate and improve the design of college financial aid.

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. (accessed 05.08.2022).

Bell, F. C. and M. L. Miller (2020). Life Tables for the United States Social Security Area, 1900-2100, Actuarial Study No. 6, Table 6. (accessed 06.01.2022).

Board of Governors of the Federal Reserve System (2019). The Survey of Consumer Finances: About. (accessed 07.24.2022).

Bureau of Economic Analysis, U.S. Department of Commerce (2022a). Table 2.1. (accessed 05.15.2022).

Bureau of Economic Analysis, U.S. Department of Commerce (2022b). Table 1.1.5. (accessed 05.15.2022).

Bureau of Economic Analysis, U.S. Department of Commerce (2022c). Table 3.1. (accessed 05.15.2022).

Bureau of Labor Statistics, U.S. Department of Labor (2019). National Longitudinal Survey of Youth 1997 cohort, 1997-2019 (rounds 1-19). Produced and distributed by the Center for Human Resource Research (CHRR), The Ohio State University. Columbus, OH.

Card, D. and T. Lemieux (2001). Can falling supply explain the rising return to colleg for younger men? A cohort-based analysis. *The Quarterly Journal of Economics 116*(2), 705–746.

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.

Caucutt, E. M. and K. B. Kumar (2003). Higher education subsidies and heterogeneity: a dynamic analysis. *Journal of Economic Dynamics and Control 27*(3), 1459–1502.

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.

Chief Operating Officer for Federal Student Aid (2021). Annual Notice of Interest Rates for Fixed-Rate Federal Student Loans Made Under the William D. Ford Federal Direct Loan Program. *Federal Register 86*(152), 44003–44005. (accessed 07.05.2022).

Colas, M., S. Findeisen, and D. Sachs (2021). Optimal need-based financial aid. *Journal of Political Economy 129*(2), 492–533.

Deming, D. and S. Dynarski (2009). Into college, out of poverty? Policies to increase the postsecondary attainment of the poor. *NBER Working Paper 15387*.

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. (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*.

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

Funding U., Inc. (2022). FAQs: Am I eligible for a no-cosigner student loan? (accessed 07.08.2022).

Goldin, C. and L. F. Katz (2007). The race between education and technology: The evolution of U.S. educational wage differentials, 1890 to 2005. *NBER Working Paper 12984*.

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

Holhoski, A., C. Clark, and C. Beresford (2022). 16 Best Private Student Loans. (accessed 07.08.2022).

Hoxby, C. and C. Avery (2013). The Missing "One-Offs": Hidden Supply of High-Achieving, Low-Income Students. *Brooking's Papers on Economic Activity*.

Hoxby, C. M. and S. Turner (2015). What High-Achieving Low-Income Students Know About College. *American Economic Review: Papers and Proceedings 105*(5), 514–517.

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. 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.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity 47*, 2025–2047.

Kydland, F. E. and E. C. Prescott (1982). Time to Build and Aggregate Fluctuations. *Econometrica 50(6)*, 1345–1370.

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). Assets and job choice: Student debt, wages and amenities. *Unpublished manuscript*.

Matsuda, K. (2020). Optimal timing of college subsidies: Enrollment, graduation, and the skill premium. *European Economic Review 129*, 103549.

Matsuda, K. (2022). Progressive Taxation versus College Subsidies with College Dropout. *Journal of Money, Credit and Banking* (available online).

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, U.S. Department of Education (2019). Table 330.10. Produced and distributed by the Institute of Education Sciences, National Center for Education Statistics (NCES). (accessed 01.04.2022).

National Center for Education Statistics, U.S. Department of Education (2020a). High School Longitudinal Study of 2009 (public use data file). Produced and distributed by the Institute of Education Sciences, National Center for Education Statistics (NCES). (accessed 10.11.2022).

National Center for Education Statistics, U.S. Department of Education (2020b). Table 331.35. Produced and distributed by the Institute of Education Sciences, National Center for Education Statistics (NCES) (accessed 01.04.2022).

Office of Federal Student Aid, U.S. Department of Education (2022). Dependency Status. (accessed 12.01.2022).

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).

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

Survey Research Center, Institute for Social Research (2021). Panel Study of Income Dynamics. Public use dataset, University of Michigan, Ann Arbor , MI.

U.S. Congressional Budget Office (2019). The Distribution of Household Income, 2016.

U.S. Congressional Budget Office (2020). The Distribution of Household Income, 2017.

U.S. Congressional Budget Office (2021). The Distribution of Household Income, 2018.

U.S. Department of Education, Office of Federal Student Aid (2021). National Default Rate Briefing for FY 2018 Official Cohort Default Rates. (accessed 05.15.2022).

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

Online Appendix - not for publication
Supplements: "Optimism About Graduation and College Financial Aid"

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

Contents

A Data Appendix

A.1 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 (Bureau of Labor Statistics, U.S. Department of Labor, 2019). This survey collected data annually from 1997 to 2011 and biannually from 2011 to the present (rounds 1 through 19).

A.1.1 Model parameters and robustness exercises

The distribution of expected graduation probabilities How are beliefs about the likelihood of earning a BA distributed among high school graduates? Table 16 shows the distribution of beliefs within each skill tercile, where the tercile is assigned using the distribution of high school GPA among high school graduates. Specifically, for the sample of high school graduates, Panel A reports the fraction of a given skill tercile that responded with an expected probability within a given range; the skill tercile is assigned a row, and the expected probability range is shown in the column header. Each row of Panel A sums to one. In all terciles, the plurality of respondents give values between 80 and 100, although the lowest skill tercile also has a large mass reporting a likelihood between 40 and 59. However, note that no skill level has a mass of 0 in any column. Additionally, reported values within a given interval are not uniformly distributed; this is shown in Panel B, which demonstrates that the average value for a given skill tercile is not the midpoint of the column's interval. In particular, for responses between 80 and 100 percent, the average value is very close to 100 percent and increasing in the skill bin, while for responses between 0 and 19 percent the average probability is closer to the lower bound of that interval.

College enrollment rates Table 17 reports enrollment rates by age 25 and by age 30 in the NLSY97 for each skill tercile, assigned using the distribution of high school GPA among high school graduates. These enrollment rates are very similar; most enrollment happens before age 25. We use enrollment by age 30 to compare true graduation rates with expectations, because this aligns with the wording of the expectations question in the NLSY97 questionnaire. For the enrollment rates used as calibration targets, enrollment by age 30 is not an intuitive mapping to the one-time enrollment choice consumers make at age 18 in the model. Since the model allows this choice to be made once immediately after high school graduation, but in reality young people may wait a few years after high school before enrolling in college, using enrollment by age 18 in the data is not satisfactory either. We therefore use enrollment by age 25, between these two ages, as the calibration target.

Table 16: Discretized distribution of beliefs among high school graduates

Panel A:		**Expected probability of earning BA**				
Distribution	**Skill**	*0 to 19*	*20 to 39*	*40 to 59*	*60 to 79*	*80 to 100*
	1	0.083	0.079	0.206	0.138	0.494
	2	0.064	0.038	0.122	0.149	0.627
	3	0.020	0.019	0.053	0.098	0.810
Panel B:		**Expected probability of earning BA**				
Mean values	**Skill**	*0 to 19*	*20 to 39*	*40 to 59*	*60 to 79*	*80 to 100*
	1	4.060	24.578	49.386	71.820	95.243
	2	5.000	22.903	49.384	72.471	95.884
	3	2.867	24.286	48.900	73.027	97.140
Obs	2,367					

Notes: Panel A of Table 16 reports the fraction of each skill bin (rows) with reported beliefs in a given interval (columns); the values in each row sum to 1. Panel B reports, for the row's skill tercile, the average belief for responses within each column's interval in units of percentages. Source: NLSY97.

Table 17: Bachelor's degree program enrollment rates by skill tercile and overall

Skill	**Obs**	**Enrolled by age 25**	**Enrolled by age 30**
1	807	22.92	27.51
2	812	45.57	48.65
3	748	77.01	78.48
Total	2,367	47.78	50.87

Notes: Table 17 shows enrollment rates in a 4-year degree program by age 25 and by age 30, for each skill tercile. Skill terciles are assigned using the distribution among high school graduates. Enrollment rates computed for the same sample. Source: NLSY97.

Table 18 shows enrollment rates by age 25 broken down by family income tercile in addition to skill tercile. Family income terciles are assigned using the distribution of high school graduates; note that the sample with valid family income observations is smaller than the main high school graduates sample.

Educational attainment outcomes versus expectations Table 19 reports the difference between student and parent expected probabilities of obtaining a BA, within the same family, when both expectations are reported (parent beliefs are only reported with valid responses for a subset of the student beliefs sample). The results are reported seperately by whether the child later enrolled in a BA (Panel A) or not (Panel B). Regardless of enrollment outcome, the average expected probabilities of parents and children in the same family agree within a few percentage points of each; the median difference is 0. Percentiles of the distribution of differences other than the median (p50) are also reported in the table and indicate the the distribution is largely symmetric around 0. These results support our modeling assumption that parents have the same subjective beliefs as

Table 18: Bachelor's degree program enrollment rates by skill and family income terciles

Income:	**1**		**2**		**3**	
Skill	**Enr. rate**	**Obs**	**Enr. rate**	**Obs**	**Enr. rate**	**Obs**
1	19	242	22	190	32	148
2	27	204	48	198	66	173
3	64	116	72	196	88	248
Obs	1,715					

Notes: Table 18 reports the enrollment rate in 4-year program by age 25, by skill tercile (rows) and family income tercile (columns). Enrollment rates are in percentages. Sample is high school graduates for whom family income is also observed. Source: NLSY97.

their child.

Table 19: Moments of the distribution of within-family difference in beliefs

Panel A: College enrollees	**Skill**	**Obs**	**mean**	**p10**	**p25**	**p50**	**p75**	**p90**
	1	166	0.99	-40	-10	0	20	40
	2	297	2.09	-25	-1	0	15	28
	3	429	0.31	-15	-5	0	0	20
	Obs	892						
Panel B: Non-enrollees	**Skill**	**Obs**	**mean**	**p10**	**p25**	**p50**	**p75**	**p90**
	1	423	3.79	-40	-10	0	25	50
	2	286	3.80	-35	-10	0	25	50
	3	116	-0.03	-31	-10	0	15	40
	Obs	825						

Notes: Table 19 shows statistics on the distribution of within-family differences between parent and child expected probabilities of the child earning a BA. Samples: Panel A, students who enrolled in a BA program before age 30, whose parents responded to the beliefs question; Panel B, students who did not enroll in a BA program before age 30, whose parents responded to the beliefs question. Source: NLSY97.

In Table 20 we report the within-skill-tercile average expected graduation rate, realized graduation rate, and the difference between these (the extent of optimism, which takes a negative value for pessimistic beliefs) by gender and student skill tercile (Panel A) and by parental education and student skill tercile (Panel B). In Panel A, we see that the difference across genders within each skill bin is small. In Panel B, we see that parental education is more predictive of optimism than gender (note that parental education is defined at the family level where having at least one parent with a BA or more is "High"; otherwise, the family is a "Low" education family). Within a skill bin, low education families tend to be more optimistic than high education families. Nevertheless, within a skill bin, we see more similarity across education categories than across skill bins within an education category.

Inter vivos transfers In order to estimate average inter vivos transfers from parents to their college-aged children in the NLSY97, we proceed as follows. We use the cleaned data from

Table 20: Subjective beliefs about BA attainment among college enrollees: breakdowns

Panel A: by student gender and skill	**Gender**	**Skill**	**Obs**	(a) **Expected graduation prob.**	(b) **Realized graduation rate**	**Difference** $(a)-(b)$
	Male	1	127	81.67	30.71	50.96
		2	168	83.88	55.95	27.93
		3	226	91.94	79.20	12.74
	Female	1	95	81.93	33.68	48.24
		2	227	90.04	55.95	34.10
		3	361	94.57	77.56	17.00
	Obs	1,204				

Panel B: by parental education and skill	**Gender**	**Skill**	**Obs**	(a) **Expected graduation prob.**	(b) **Realized graduation rate**	**Difference** $(a)-(b)$
	Low	1	156	80.18	28.85	51.33
		2	292	87.66	51.71	35.95
		3	350	92.64	73.14	19.50
	High	1	56	85.57	42.86	42.71
		2	80	89.65	75.00	14.65
		3	214	95.67	86.45	9.22
	Obs	1,148				

Notes: Table 20 compares expectations and outcomes across skill terciles by student gender and parental education level. Panel A is students who enrolled in a BA program before age 30, and Panel B is students who enrolled in a BA program before age 30 and for whom parental education is observed. Source: NLSY97.

the earnings process estimation, described in Section A.1.2 below. Next, we restrict attention to observations where sample members were independents between the ages of 18 and 23 during the years from 1997 to 2003.[50] To account for an implicit transfer from parents to their children who cohabit with them and do not pay rent, we flag those cohabiting with their parents and paying no monthly rent, then impute the average monthly rent paid by sample members with the same family income tercile, college enrollment status, and observation year who are not cohabiting. Next, we transform monthly rent to yearly rent, and add it to yearly net income received from parents (if both parents are present) or from both the mother and father (if both parents are not present). We also add any yearly allowances received. The resulting quantity is the yearly nominal transfers from parents to their child. Within each year, we then multiply the quantity by 6 and divide by nominal GDP per capita in that year (for those over 18) to find a unitless implied ratio of transfers received to per capita income for each individual while they are young adults of college age. We then average this ratio across invididuals and years to find the ratio reported in the first row of Table 21. The average real values of the components of transfers are also reported. To convert these to real values in 2000 U.S. dollars, we use the Consumer Price Index (CPI).

[50]We keep observations that are enrolled in post-secondary education, which broadens the sample relative to the earnings estimation in any given year. For independence criteria used in the inter vivos transfers estimation, see National Longitudinal Surveys, https://www.nlsinfo.org/content/cohorts/nlsy97/topical-guide/Income. The NLS criteria for dependency status are not the same as those used in the FAFSA (FSA, 2022b).

Table 21: Inter vivos transfers

Variable	Mean
Transfer ratio	0.578
Transfers	4,706
Transfers not allowance	539
Allowance	138
Imputed rent	4,671
Obs	8,114
Individuals	2,991

Notes: Table 21 reports average transfers for the sample used to estimate inter vivos transfers. Sample: independents between 18 and 23 observed during 1997-2003. Units for transfer amounts: year 2000 USD. Data are at the individual-year level. Source: NLSY97.

A.1.2 Earnings process estimation and the college wage premium

The earnings process we use in our structural model realizes a quantity of efficiency units at each age j. This quantity has a deterministic component, $\epsilon_{j,e,s}$, and a stochastic component, η_j. The deterministic component depends on the consumer's age, j, their education, e, and their skill endowment (high school GPA), s:

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

The stochastic component is an AR(1) process where the persistence parameter depends on the consumer's educational attainment, as does the Normal distribution from which the error term is drawn:

$$\begin{aligned} \eta_j &= \rho_{\eta,e}\eta_{j-1} + \nu_{e,j} \\ \nu_{e,j} &\sim \mathbb{N}\left(0, \sigma_{\nu,e}\right) \end{aligned}$$

To estimate the earnings process for each education category e, we implement a modification of the approach described in Abbott, Gallipoli, Meghir, and Violante (2019).[51] First, we use the Panel Study of Income Dynamics (PSID) to estimate how logged real wages depend on a third-order polynomial of age for a given education group, $e = \ell$ (HS or some college) or $e = h$ (BA or higher). This identifies $\beta^A_{e,1}$, $\beta^A_{e,2}$, and $\beta^A_{e,3}$ for each education group e. We use the PSID to estimate the age polynomial because it allows us to see a more complete life cycle of earnings than is visible in the NLSY97 due to the latter survey's shorter panel dimension. Next, we take logged hourly real wages in the NLSY97, clean them of age effects with the PSID estimation results, and regress the resulting age-free log hourly real wages on indicators for skill terciles. The coefficients on skill

[51] That paper includes gender as a type, while we do not have that kind of heterogeneity. This necessitated re-estimating the earnings profiles so that they are compatible with our model specification.

tercile indicators are the factor loadings on skill s for a given education e, $\beta^s_{e,s}$. Finally, using the residuals from the NLSY97 regression, we jointly estimate $\rho_{\eta,e}$ and $\sigma_{\nu,e}$ for each education group. Point estimates are reported in Table 23.

Estimating age profiles in the PSID The PSID collects data on the household head and, if present, their resident spouse (Survey Research Center, Institute for Social Research, 2021). We use information on the educational attainment of the household head and resident spouse (if any), as well as each individual's sex, total income, total income from transfers, total labor earnings, labor component of business income, 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 the labor component of business income when necessary) by total hours worked for the individual.[52] We correct all income and wage variables for inflation using the CPI and thereafter use real dollar values in our analysis. We then reshape the data into an individual-level panel where each male or female adult in the household is followed over time.

We exclude observations from the SEO census sample and drop observations for whom we do not observe state of residence, marital status, or sex of the household head. We then count the number of times an individual is observed and drop individuals observed fewer than eight times. We compute yearly real wage growth and drop observations with growth higher than 4 percent or less than -2 percent, or where the level of real wages exceeds 400. We then 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 21 if they have a BA or more. Next, we drop those who are self-employed. We define those with a high school education as individuals who have between 12 and 15 years of education ("high school"); those with a college education are individuals with 16 years or more of education ("BA"). These definitions mean that those with only an associate's degre and dropouts from 4-year bachelor's program are assigned to the high school graduates group in our estimation procedure. The estimation sample has 85,898 individual-year observations for the high school group, and 65,042 for the BA group. Using this estimation sample, we proceed in two stages to account for selection into working within each education category. In the first stage, we regress an indicator for working positive hours on an age polynomial and a set of standard controls (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. In addition to the standard controls, X (where Xincludes a constant), in the first stage we also control for Z, which is unearned real

[52]The labor component of business income is not included in labor earnings for some years of the PSID. For years when it is not included, we manually add it to reported labor earnings.

income. This first-stage regression can be written as

$$\mathbb{I}_{hrs>0} \;=\; \gamma_{e,Z} Z + \alpha_e X + \epsilon$$

where ϵ is the residual. This first-stage 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 inverse Mills ratio, IM, from the first stage. In this second stage regression, the dependent variable is the log of the real wage, w, and we use an Ordinary Least Squares (OLS) estimator. This regression estimated on a given education group can be written as

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

where u is the i.i.d. residual. The age profile of education e is given by $\beta^A_{e,1}$, $\beta^A_{e,2}$, and $\beta^A_{e,3}$.[53] As a check on our model specification, we also estimate the effect of some college or an associate's degree, relative to only a high school degree, on the age profile of earnings by running the same regression augmented with the interaction of a flag for some college, $\mathbb{I}_{\text{SC}}$, with the age polynomial. Results for the interaction terms of this estimation are presented in Table 22; these coefficients are statistically insignificant.

Table 22: Log wages as a function of age: robustness on pooling assumption

Controls	$\log(wage)$
$\mathbb{I}_{\text{SC}} \times$ age	0.0130
	(0.0138)
$\mathbb{I}_{\text{SC}} \times$ age^2	-0.0000750
	(0.000351)
$\mathbb{I}_{\text{SC}} \times$ age^3	-0.000000944
	(0.00000285)
$\mathbb{I}_{\text{SC}}$	-0.167
	(0.174)
R^2	0.119
Obs	85,898

Notes: Table 22 reports regression results. Not shown but included: uninteracted age polynomial, state and year FE, flag for married, inverse Mills ratio, constant. Source: PSID.

Estimating skill loadings in the NLSY97 Our sample keeps only observations where we observe high school GPA, wage, educational attainment, and completion of high school. We correct for inflation using the CPI and drop observations with real wages in dollar units above 400 and

[53] Because the average rejected wage offer is likely lower than the average accepted wage offer, the expected sign of the inverse Mills ratio coefficient in the second stage, $\gamma_{e,IM}$, is positive. In our estimation, this coefficient has the expected sign for both education groups.

below 1 or wage growth above 4 percent or below -2 percent. We drop those with either some high school or with a GED, and those currently enrolled in a BA program, and restrict ages to be above 24 and below 39 so that each age bin has at least 100 observations. We group observations as either "high school" meaning those with a high school degree or some college, or "BA" meaning those with a BA degree or more. Since the NLSY97 records information at the individual level, we reshape the data to be a panel at the individual-year level. We estimate the factor loadings on skill using these remaining observations in the resulting panel data: there are 14,961 observations for the high school group and 8,545 for the BA group.

Using the estimated age contributions to log wages from the PSID, we log real wages in the NLSY97 and, using the observation's associated age, clean logged real wages of their estimated age component. The resulting "age-free" log wages, w_{AF}, are then regressed on dummies for high school GPA terciles, as well as a set of controls X that include 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 NLSY97. Standard errors in this regression are clustered at the individual level. The estimation equation can be written as

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

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

Estimating the stochastic component of earnings After estimating the skill loadings in the NLSY97, we use the residuals of that regression as inputs to estimate a shock process for each education category. 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.

Summary of earnings process estimation results Table 23 presents the results of the earnings process estimation. We find that earnings increase at a decreasing rate over the life cycle and the college wage premium is lower for those with lower skill endowments. We also find that the stochastic component of the earnings process is more persistent for those with more education, although random-shock variances are similar.

College wage premium by skill tercile Table 24 reports the median wage within each skill tercile by education group. The last column of the table is the college wage premium within each skill tercile, which is the ratio of the two medians. The sample used in Table 24 is at the individual-year level and is the same as what is used at in the earnings process estimation for skill loadings.

Table 23: Earnings process estimation results

Parameter	Description	Value $e = \ell$	$e = h$
$\beta^A_{e,1}$	Age third-order polynomial	0.105	0.182
$\beta^A_{e,2}$		-0.00174	-0.00309
$\beta^A_{e,3}$		0.00000874	0.0000165
$\beta^s_{e,1}$	Skill endowment shifter	-0.0426	-0.180
$\beta^s_{e,2}$		-0.0362	-0.132
$\rho_{\eta\|e}$	Persistence AR(1)	0.855946	0.879158
$\sigma^2_{\nu\|e}$	Variance AR(1)	0.082112	0.078444

Notes: Table 23 summarizes the results from the earnings process estimation. Sources: PSID and NLSY97.

The wage premiums reported in Table 24 are compared with their untargeted model counterparts in Table 32 of Subsection C.1 of this appendix.

Table 24: Bachelor's degree wage premium by skill tercile: ratio of median wages

	High school		Bachelor's degree		
Skill	**Wage**	**Obs**	**Wage**	**Obs**	**Wage premium**
1	10.64	6,902	14.18	880	1.33
2	11.06	5,382	15.56	2,369	1.41
3	11.23	2,677	17.58	5,296	1.57

Notes: Table 24 tabulates the median wage within each high school GPA tercile for those with a high school degree but less than a bachelor's degree ("High school") and those with a bachelor's degree or higher ("Bachelor's degree"), for those not currently enrolled in post-secondary education. The last column is the ratio of median wages in the two educational attainment categories. Source: NLSY97.

A.2 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 (National Center for Education Statistics, U.S. Department of Education, 2020a). We use the public version of the HSLS:09, where this information is reported up to and including the 2015-2016 academic year (Duprey et al., 2020).

The structure of the HSLS:09 is complex.[54] Waves of the study occur in the fall of 2009, in the spring of 2012 (first follow-up), in the summer of 2013 (2013 update), and in the spring of 2016 (second follow-up). High school transcripts are collected during the 2013-2014 academic year, and post-secondary transcripts (as well as student records) are collected in the 2015-2016 aca-

[54] Questionnaires are available here: National Center for Education Statistics, https://nces.ed.gov/surveys/hsls09/questionnaires.asp.

demic year (after potentially three full years of academic enrollment in post-secondary education). The second follow-up in the spring of 2016 includes information from students who are currently enrolled in post-secondary education, as well as those who are not enrolled but used to be, and those who did not pursue post-secondary education. If sample members begin a four-year degree program in the fall after high school graduation (the fall of 2013) and do not take any time off from school, then they complete the second follow-up questionnaire in the spring of their third year of college and student records are available up to and including the 2015-2016 academic year. Regardless of persistence status, survey information about the focal sample member includes their high school GPA, as well as any financial aid and private loans they took out to pay for post-secondary education. Information on federal financial aid (loans and grants) and private loans are also collected from institutions themselves in the post-secondary transcripts and student records data collection wave. Our estimations use variables based on student record information, when available.

A.2.1 Calibration targets and model primitives

Table 25 reports moments computed by skill tercile in the HSLS:09 used to discipline our quantitative model. The table includes three categories of moments, indexed with roman numerals: child skill by parental education, tuition and grant aid, and persistence rates. Category I shows that, among students who have graduated from high school, parents with higher education tend to have children in higher skill (high school GPA) terciles. Category II reports the average tuition paid by each skill tercile of fall 2013 college enrollees. The fact that tuition does not vary greatly across skill terciles is why the model of the main text includes a pre-subsidy tuition level set to the same value for all college students. The second column in Category II is the ratio of aggregate grants to aggregate tuition and fees within each skill tercile during the first academic year of enrollment. This ratio is used to compute the subsidy rate from public and private grants reported in Table 7 of the main text. Finally, Category III reports moments used to discipline the true probability of completing the third academic year of a BA program, conditional on being enrolled their first year and on being enrolled in their second year in the first and second column, respectively.

Table 26 reports moments describing average labor supply among college enrollees and reasons for not enrolling in post-secondary education. The average time spent working per week, for fall 2013 enrollees who persist through their third year, is expressed as a fraction of full-time work (40 hours). The last three rows of this table report suggestive evidence for why students never enroll in post- secondary education to motivate the introduction of the enrollment option shock, $q(s)$, in the quantitiative model. This evidence uses responses to the question “why did you never enroll in college?” (in this survey question, unlike in the main text and this appendix generally, “college”

Table 25: Statistics by skill tercile

		(I) Child skill by parental education		(II) Tuition and grant aid		(III) Percentage persisted (Y3)	
		$\pi(s_c \mid e = \ell)$	$\pi(s_c \mid e = h)$	**Tuition + Fees**	$\frac{\text{Agg Grants}}{\text{Agg Tuition + Fees}}$	**if Enr. Y1**	**if Enr. Y2**
Skill	1	42.64	17.63	17,139	0.407	47.57	77.40
	2	34.08	31.12	17,694	0.462	71.08	86.26
	3	23.28	51.24	19,959	0.520	82.90	91.26

Notes: Table 25 shows statistics by skill tercile for three categories of variables. Category I reports the conditional distribution over high school GPA terciles among high school graduates given parental education (where $e = h$ denotes at least one parent with BA or more); Category II reports tuition and fees in dollar amounts and total grants as a fraction of tuition and fees during the first academic year for fall 2013 (Y1) enrollees; Category III reports conditional persistence probabilities given enrollment in year 1 (first column) and given enrollment in year 1 and year 2 (second column). Samples vary across categories. Weights are Second Follow-up longitudinal weights for Category I and PETS-SR longitudinal weights for Categories II and III. Source: HSLS:09.

refers to *any* post-secondary education). Respondents are only asked this question if they say that they never enrolled in post-secondary education, so those who never enroll in a four-year degree program are frequently not asked this question because they may have enrolled in another type of post-secondary program instead. Even conditioning on being asked, non-response rates are high. Nevertheless, when presented with a menu of possible reasons for not enrolling, many respondents indicate that factors such as academics, family, or other reasons that do not include financial or work factors led to them not enrolling in post-secondary education.

Table 26: Labor supply and dependency status, and reasons for never enrolling

Category	**Variable**	**Value**	**Sample obs**
Labor supply junior year	$\frac{\text{Average weekly hours worked}}{40}$	0.347	1,855
Reason never enrolled in post-secondary ed. (answered "yes" for a given reason)	Academic, personal/family, other	0.244	5,393
	Financial	0.193	
	Work, military, career	0.150	

Notes: Table 26 reports labor supply and reasons for never enrolling in a post-secondary program. Samples: first row is students who enrolled in a 4-year program in the fall of 2013 and persisted through their third academic year; remaining rows are sample members who graduated from high school in 2013 and either did not enroll or enrolled in a 4-year degree in the fall of 2013; enrollees are counted as answering 'No' for each possible reason; values are frequencies of answering "Yes' for a given reason. Weights are PETS-SR longitudinal weights for the first row and 2013 Update longitudinal weights for the remaining moments. Source: HSLS:09.

Table 27 presents regression results for an exercise in which we regress an indicator for persisting to the next academic year on various attributes of the student in the current year. We use an OLS estimator with the dependent variable being indicator for persisting from year 1 to year 2 (model 1 in the table) and from year 2 to year 3 (model 2). The results indicate that high school GPA plays a statistically significant role in predicting persistence early in one's college career in both the first and second academic years, even controlling for parent attributes (household income and education) and student attributes (debt, hours worked, gender) and institution attributes (tuition and fees in first institution attended). Other than GPA, no other control is statistically significant at the 5 percent level; parent education is slightly significant in the first academic year, but this fades in the second academic year. These results, along with the conditional persistence probabilities presented in Table 25, motivate our model specification linking the probability of being allowed to continue in college, $p_c(j, s)$ to student skill, s, and year of enrollment, j.

Table 27: Predicting enrollment persistence

	(1) Y2 ∣ Y1	(2) Y3 ∣ Y2
High school GPA	0.10122	0.06686
	(0.02553)	(0.02226)
Log(HH income)	-0.00950	-0.00151
	(0.02142)	(0.01631)
Log(SL debt)	0.01771	0.05529
	(0.05360)	(0.03755)
Hours worked per week	-0.00143	-0.00147
	(0.00147)	(0.00120)
Log(tuition and fees Y1)	0.02090	0.00035
	(0.02139)	(0.02190)
Flag: no SL debt	0.18505	0.52812
	(0.46425)	(0.34355)
Flag: parents BA+	0.05243	0.04187
	(0.02943)	(0.02728)
Flag: female	-0.00155	0.03369
	(0.02527)	(0.02521)
Constant	0.24260	0.10548
	(0.52389)	(0.44690)
R^2	0.065	0.036
Obs	2,356	2,097

Notes: Table 27 reports results from regressing an indicator for persisting to the next academic year on various controls measured in the current academic year using an OLS estimator. Sample: students who enrolled in a four-year program in the fall of 2013 (Y1); the second column additionally conditions on being enrolled in the 2014-2015 academic year (Y2). Bootstrap standard errors are in parentheses; weights are PETS-SR student records longitudinal weights. Source: HSLS:09.

A.2.2 Educational attainment expectations versus outcomes

In the first follow-up wave (spring of junior year of high school) the HSLS:09 asks interviewees about their expected educational attainment. Unlike the phrasing of a similar question in the NLSY97, the phrasing of the question in the HSLS:09 on expected educational attainment is not probabilistic: the specific wording of the question when posed to students is “As things stand now, how far in school do you think you will actually get [in your education]?” The survey also asks the same question of the student’s parent about their child’s prospects. The possible answers range from 1 (“Less than high school completion”) to 12 (“Complete a PhD”), with 13 “Don’t know” as an optional response. To flag those who expect to complete a four-year BA program, an indicator is created that is set to 0 for responses between 1 and 13 (“Don’t Know” is a valid response) and replaced with a 1 if the response x is such that $8 \leq x < 13$, that is expect to complete a BA or higher. An indicator for those who expect to enroll in a master's degree or higher is constructed a similar way, but with the lower bound starting at 10 ("Start a Master's degree"). Subsequently, we are able to verify whether the sample members enroll in a four-year BA program after high school 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).

Panel A of Table 28 presents, by high school GPA tercile, the percentage of each skill bin that expected to complete a BA program and the percentage of the bin that complete their third academic year of a 4-year BA degree.[55] In particular, Panel A of Table 28 demonstrates that the sample of students who enroll in a four-year program in 2013 tend to overestimate their educational attainment, given their skill. This is especially the case for those in the lowest skill tercile.

A concern with the findings reported in Panel A of Table 28 is that responents claim they will get a BA to avoid a utility cost, which may generate a "social desirability bias" in the survey responses. To address this concern, in Panel B we show a tabulation restricting to those who expect to attend a master's (MA) degree or higher. Note that, by implication, in this group everyone expects to get a BA. This eliminates students who are fibbing in their responses that they expect to earn a BA or more because of stigma costs, by dropping those right on the threshold of admitting they won't get a BA. It seems less likely that stating you expect to begin an MA or more, relative to a BA, is driven by fear of stigma costs. The tabulation demonstrates that the percentage who persist in each tercile still remains well below the expected graduation rate from college, especially for the lowest skill tercile.

Finally, in Panel C of Table 28, we tabulate the parent responses to what they expect their child's educational attainment will be.[56] Parents tend to overestimate the likelihood of college graduation for their children, especially when their child belongs in a lower skill tercile.

A.3 The 2019 Survey of Consumer Finances

The 2019 SCF is a nationally representative cross-sectional survey of families that is conducted every three years. It is sponsored by the Federal Reserve Board of Governors and the U.S. Depart-

[55] Note that, because of the short panel dimension of the HSLS:09, we cannot definitively say if they permanently drop out of college or fail to ever enroll during the course of their life. For this reason, we use terms such as "persistence" and "non-persistence" when discussing findings from the HSLS:09, as opposed to more definitive terms like "dropping out" and "graduating", respectively.

[56] 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 percent of families in the sample.

Table 28: Educational attainment expectations versus outcomes

Panel	Sample	Skill	Sample obs	Group obs	% Persisted BA	% Expect BA	Difference
A	Fall 2013 enrollees	1	2,356	155	76	48	29
		2		659	80	71	9
		3		1,542	93	83	10
B	Expect MA+	1	1,356	57	100	44	56
		2		310	100	70	30
		3		989	100	83	17
C	Parent expectations	1	1,021	62	76	38	38
		2		277	92	71	21
		3		682	94	81	13

Notes: Table 28 compares realized and expected bachelor's degree attainment. Samples vary across panels. Weights are PETS-SR student records longitudinal weights. Source: HSLS:09.

ment of the Treasury (Board of Governors of the Federal Reserve System, 2019). The SCF reports interest rates for federal and private student loans for which respondents still owe a positive amount when the survey is conducted. Together with findings on private student loans from the HSLS:09, we use interest rates by loan type from the SCF to discipline model attributes of the private student loan market.

A.3.1 Interest rates by loan type

Along with demographic information, for each family the 2019 Survey of Consumer Finances records information on up to six student loans, including the interest rate, repayment status, and type of loan (federal or private). We separate student loans into federal or private loans and report the mean and median of interest rates within each loan type in Table 29, both overall and by the borrower outcome groupings of income, education, and delinquency status. These three statistics are very similar across the two loan types (second column). The third column breaks down interest rates by income tercile, while the remaining four columns break down interest rates by graduate status (that is, education outcome) for all families and additionally restricting to families who are delinquent on their loans. Along all of these margins, the difference between federal and private student loans in the mean or median interest rate is small. Additionally, within each loan type, the interest rate shows little variation across borrower attribute groupings. The fact that federal loans exhibit this pattern in the SCF, which one would expect because they are set by statute as a common add-on, lends credence to the findings about private loan interest rates from the same dataset.

Table 29: Student loan interest rates

		All families					Delinquent families	
		Income tercile			Graduate status		Graduate status	
Loan type	**All families**	**1**	**2**	**3**	**Yes**	**No**	**Yes**	**No**
Federal								
Mean	5.97	5.95	6.08	6.26	5.92	6.29	6.02	6.88
Median	5.50	5.50	5.32	5.96	5.50	5.60	6.00	6.00
Obs	3,841	592	1,647	1,602	2,658	675	202	194
Private								
Mean	5.85	5.65	5.95	6.78	5.86	6.07	6.18	6.90
Median	5.84	6.00	4.85	6.38	5.84	5.40	6.70	6.00
Obs	779	85	253	441	554	144	52	30

Notes: Moving from left to right, Table 29 reports moments (indicated in the first column) for interest rates of federal and private student loans for all families (second column), by income tercile within all families (third through fifth columns), by educational attainment within all families (sixth and seventh columns), and by educational attainment within delinquent families (eighth and ninth columns). Graduate families (for whom the graduate status is "Yes") have completed at least one of the programs for which they took out their education loans. Delinquent families have at least one education loan for which they are late making payments. All moments use survey weights. Source: 2019 SCF.

A.4 The Congressional Budget Office's "The Distribution of Household Income"

In order to estimate the degree of income tax progressivity, τ_p, we use aggregate data on the distribution of household income published by the Congressional Budget Office (CBO) for 2016, 2017, and 2018 (U.S. Congressional Budget Office, 2019, 2020, 2021); specifically, we apply the robustness method of Heathcote, Storesletten, and Violante (2017) to data underlying Figures 1, 3, and 4 of those publications.

A.4.1 Selected data underlying figures

We use the data underlying figures from the CBO report to find the baseline federal tax rate (column 1 in Table 30), as well as the transfer rate from Temporary Assistance to Needy Families (TANF), Supplemental Nutrition Assistance Program (SNAP), and Supplemental Security Income (SSI) shown in columns (2), (3), and (4) of Table 30, respectively. We compute the empirical equivalent of the net tax rate for our model as the federal tax rate (which includes refundable credits as reported in column 1) minus the transfer rates from TANF, SNAP, and SSI and report this net tax rate in column (5). Average pretax income in column (6) is logged in column (7) and logged after-tax income reported in column (8) is computed by taking the log of the net tax rate in column (5) applied to the pretax income of column (6). The specific figures within each CBO report whose underlying data provides the empirical moments for the corresponding year are: for column (1), Figure 4; for columns (2)-(4), Figure 3; and, for column (6), Figure 1.

Table 30: Estimating income tax progressivity using CBO data: estimation data

	Percentiles		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Year	Min	Max	Fed. tax	TANF	SNAP	SSI	Net tax	Ave. Y	$\log(Y)$	$\log(Y_{AT})$
2016	99	100	33.3				33.3	1789	0.25	0.08
	96	99	26.8				26.8	360	-0.44	-0.58
	91	95	23.6				23.6	218	-0.66	-0.78
	81	90	21.2				21.2	160	-0.80	-0.90
	60	80	17.9				17.9	110	-0.96	-1.04
	40	60	13.9	0.5			13.4	72	-1.14	-1.21
	20	40	9.4	2.0	1.2	0.9	5.3	45	-1.35	-1.37
	0	20	1.7	10.1	8.4	6.4	-23.2	21	-1.68	-1.59
2017	99	100	31.6				31.6	1,960	0.29	0.13
	96	99	26.5				26.5	380	-0.42	-0.55
	91	95	23.4				23.4	230	-0.64	-0.76
	81	90	21.3				21.3	170	-0.78	-0.88
	60	80	17.9				17.9	110	-0.95	-1.03
	40	60	14.0	0.5			13.5	80	-1.12	-1.19
	20	40	9.2	2.0	1.1	0.9	5.2	50	-1.34	-1.36
	0	20	1.3	9.7	8.1	5.9	-22.4	20	-1.68	-1.59
2018	99	100	30.2				30.2	2,000	0.30	0.14
	96	99	24.2				24.2	400	-0.40	-0.52
	91	95	21.9				21.9	240	-0.62	-0.73
	81	90	20.0				20.0	170	-0.77	-0.87
	60	80	16.7				16.7	120	-0.92	-1.00
	40	60	12.8				12.8	80	-1.10	-1.16
	20	40	8.1	1.6	0.9	0.8	4.8	50	-1.30	-1.32
	0	20	0.05	9.2	6.9	5.9	-21.95	20	-1.70	-1.61

Notes: Table 30 reports the components for the estimation of the income tax progressivity parameter τ_y. Data is from 2016, 2017, and 2018, and dollar values in column (6) are in millions of current USD. After-tax income is defined as $Y_{AT} \equiv (1 - \text{Net tax})$, where the net tax rate is defined as $(5) \equiv (1) - (2) - (3) - (4)$.

A.4.2 Estimation of income tax progressivity parameter

To estimate τ_p, we derive the estimation equation from the relationship $Y_{AT} = \lambda Y^{1-\tau_p}$. Taking the log of both sides yields $\log(Y_{AT}) = \log(\lambda) + (1-\tau_p)\log(Y)$. This yields the estimation equation, $\log(Y_{AT}) = \beta_0 + \beta_1 \log(Y)$, where $\beta_1 = 1 - \tau_p$. We therefore regress column (8) from Table 30 on column (7), using population shares for each row as weights (which are implied by percentiles in that row). The results are presented in Table 31. The average estimated value for τ_p is 0.177.

Table 31: Income tax progressivity estimation results

Coefficient	$\log(Y_{AT})$ 2016	2017	2017
β_1	0.815 (0.0277)	0.822 (0.0267)	0.833 (0.0231)
β_0	-0.253 (0.0335)	-0.243 (0.0323)	-0.224 (0.0275)
Implied $\hat{\tau}_p$	0.185 (0.0277)	0.178 (0.0267)	0.167 (0.0231)
Average 2016-2018 $\hat{\tau}_p$	0.177		

Notes: Table 31 reports estimation results. Standard errors are in parentheses (for $\hat{\tau}_p$ in each year, these are computed using the delta method); coefficients are significant at the 0.1 percent significance level.

B Model Appendix

B.1 Value functions

The subjective value of college for $j = 4$ is given by

$$\hat{V}(j,h,s,\eta,a,x,\hat{p}) = \max_{\hat{c}\geq 0,\hat{a}',\hat{x}'} U(c,j,h) - \xi_L \mathbb{I}_{a\geq 0 \text{ and } x=0 \text{ and } (\hat{a}'<0 \text{ or } \hat{x}'>0)} - \xi_L^{pr} \mathbb{I}_{x=0 \text{ and } \hat{x}'>0} \tag{15}$$
$$+ \beta\psi_j[\hat{p}E_{\eta'|h,\eta}V(j+1,h,s,\eta',\hat{a}',\hat{x}') + (1-\hat{p})E_{\eta'|\ell,\eta}V(j+1,\ell,s,\eta',\hat{a}',\hat{x}')]$$
$$s.t.$$
$$(1+\tau_c)\hat{c} + \hat{a}' + (1-\theta(s)-\theta^{pr}(s))\kappa = y_{j,h,s,\eta,a} + a + Tr_j - T(y_{j,h,s,\eta,a}) + (\hat{x}'-x)$$
$$\hat{a}' \geq -\bar{A}\left(\tfrac{j}{4}\right)[(1-\theta(s)-\theta^{pr}(s))\kappa + \bar{c}]$$
$$\hat{a}' \leq a \;\text{ if }\; a \leq 0$$
$$\hat{x}' - x \in \left[0, [(1-\theta(s)-\theta^{pr}(s))\kappa + \bar{c}] - [\max(-\hat{a}',0) - \max(-a,0)]\right]$$

The idiosyncratic state of a consumer while $j > 4$ and $j \neq j_f + j_a$ is given by the tuple (j,e,s,η,a,x). The consumer's value function is given by

$$V(j,e,s,\eta,a,x) = \max_{d_f,d_x}(1-d_f)(1-d_x)V^R(j,e,s,\eta,a,x) + \tag{16}$$
$$d_f(1-d_x)V^{D_f}(j,e,s,\eta,a,x) + (1-d_f)d_xV^{D_x}(j,e,s,\eta,a,x) + d_fd_xV^D(j,e,s,\eta,a,x)$$

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

$$V^R(j,e,s,\eta,a,x) = \max_{c\geq 0,a'} U(c,j,e) + \beta\psi_j E_{\eta'|e,\eta} V(j+1,e,s,\eta',a',x') \tag{17}$$
$$s.t.$$
$$(1+\tau_c)c + a' = y_{j,e,s,\eta,a} + a + \mathbb{I}_{\{a<0\}} r_{SL} a + Tr_j - T(y_{j,e,s,\eta,a}) - \rho_R^{pr}(j,x)$$
$$a' \begin{cases} = (1+r_{SL})a + \rho_R(j,a) & \text{if } \; a < 0 \\ \geq 0 & \text{if } \; a \geq 0 \; \text{ and } \; x = 0 \\ = 0 & \text{otherwise } (a \geq 0 \; \text{ and } \; x > 0) \end{cases}$$
$$x' = (1+r_{SL}^{pr})x - \rho_R^{pr}(j,x)$$

Alternatively, these consumers can choose delinquency on either type of loan or on both loans. If a consumer chooses delinquency on only federal loans, their value function for $j > 4$ and $j \neq j_f + j_a$ is given by

$$V^{D_f}(j,e,s,\eta,a,x) = U(c,j,e) - \xi_D + \beta\psi_j E_{\eta'|e,\eta} V(j+1,e,s,\eta',a',x') \tag{18}$$
$$s.t.$$
$$(1+\tau_c)c = y_{j,e,s,\eta,a} + Tr_j - T(y_{j,e,s,\eta,a}) - \rho_D(j,a,y_{j,e,s,\eta,a}) - \rho_R^{pr}(j,x)$$
$$a' = (1+r_{SL})a + \rho_D(j,a,y_{j,e,s,\eta,a}) - \phi_D[\rho_R(j,a) - \rho_D(j,a,y_{j,e,s,\eta,a})]$$
$$x' = (1+r_{SL}^{pr})x - \rho_R^{pr}(j,x)$$

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 wage garnished to make a partial payment of $\rho_D(j,a,y_{j,e,s,\eta,a})$. 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 in Section 3.3, 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 on only private loans, their value function for $j > 4$ and $j \neq j_f + j_a$ is given by

$$
\begin{aligned}
&V^{D_x}(j, e, s, \eta, a, x) = U(c, j, e) - \xi_D^{pr} + \beta\psi_j E_{\eta'|e,\eta} V(j+1, e, s, \eta', a', x') \qquad (19)\\
&s.t.\\
&(1+\tau_c)c + a' = y_{j,e,s,\eta,a} + a + \mathbb{I}_{\{a<0\}} r_{SL} a + Tr_j - T(y_{j,e,s,\eta,a}) - \rho_D^{pr}(j, x, y_{j,e,s,\eta,a})\\
&a' = \mathbb{I}_{a<0}(1 + r_{SL})a + \rho_R(j, a)\\
&x' = (1 + r_{SL}^{pr})x - \rho_D^{pr}(j, x, y_{j,e,s,\eta,a}) + \phi_D[\rho_R^{pr}(j, x) - \rho_D^{pr}(j, x, y_{j,e,s,\eta,a})]
\end{aligned}
$$

where ξ_D^{pr} 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 are subject to wage garnishment because of delinquency on private loans. The garnishment amount is denoted by $\rho_D^{pr}(j, x, y_{j,e,s,\eta,a})$, as described in Section 3.3. Similar to the case of delinquency on federal loans, the consumer faces a collection fee, which is equal to a fraction ϕ_D multiplied by the difference between full payment and partial payment on private loans.

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

$$
\begin{aligned}
&V^{D}(j, e, s, \eta, a, x) = U(c, j, e) - \xi_D - \xi_D^{pr} + \beta\psi_j E_{\eta'|e,\eta} V(j+1, e, s, \eta', a', x') \qquad (20)\\
&s.t.\\
&(1+\tau_c)c = y_{j,e,s,\eta,a} + Tr_j - T(y_{j,e,s,\eta,a}) - \rho_D(j, a, y_{j,e,s,\eta,a}) - \rho_D^{pr}(j, x, y_{j,e,s,\eta,a})\\
&a' = (1 + r_{SL})a + \rho_D(j, a, y_{j,e,s,\eta,a}) - \phi_D[\rho_R(j, a) - \rho_D(j, a, y_{j,e,s,\eta,a})]\\
&x' = (1 + r_{SL}^{pr})x - \rho_D^{pr}(j, x, y_{j,e,s,\eta,a}) + \phi_D[\rho_R^{pr}(j, x) - \rho_D^{pr}(j, e, x, y_{j,e,s,\eta,a})]
\end{aligned}
$$

A consumer who chooses this outcome is subject to stigma cost, wage garnishment, and a collection fee (analogous to the previous two cases) from both the federal student loan program and the private lender; their consumption for the current period and outstanding loan balances for the next period follow from the same set of delinquency rules described above. When $j = j_f + j_a$ and the consumer chooses delinquency, we assume those consumers cannot make an inter vivos transfer to their child in order to be consistent with our assumption that consumers cannot save until they have paid off their student loans. Therefore, the value functions for delinquency are largely the same as in equations (18)-(20), with the difference that the parent has a term reflecting altruistic utility toward their child, represented by the addition of $\beta_c E_{\eta'|\ell} \hat{W}(s_c, \eta', b = 0, \hat{p})$ to the objective function.

B.2 Definition of equilibrium

To define the equilibrium, we must first discuss notation, define the Social Security transfer function, and present the zero expected profit condition that pins down the private student loan interest rate. Let $\overrightarrow{\omega}$ denote the idiosyncratic state of a consumer. This state depends on age and enrollment status in the following way:

$$\overrightarrow{\omega} = \begin{cases} (s, \eta, a, \hat{p}) & \text{for 18-year-olds, before making the college entrance decision} \\ (j, h, s, \eta, a, x, \hat{p}) & \text{for consumers in college} \\ (j, e, s, \eta, a, x) & \text{for consumers not enrolled, dropouts, or graduates, if } j \neq j_f + j_a \\ (j, e, s, \eta, a, x, s_c, \hat{p}) & \text{if } j = j_f + j_a \end{cases} \tag{21}$$

Furthermore, let $\hat{d}_{d,t}(\overrightarrow{\omega})$ and $d_{d,t}(\overrightarrow{\omega})$ denote the dropout decisions that solve the endogenous discrete dropout problems in the continuation values of equations (3) and (4), respectively.

Private loan interest rate: $r^{pr}_{SL,t}$ is such that the lender makes zero expected profits in pooling each cohort of 18-year-old-consumers. The zero expected profit condition is given by

$$\begin{aligned} &\sum_{i=1}^{4}(\beta)^{i-1}\int((1+\tau_{is})x'_{t+i-1}(\overrightarrow{\omega}) - x)\Omega_{t+i-1}d(\overrightarrow{\omega}|j=i) = \\ &\sum_{i=5}^{J}(\beta)^{i-1}\int\Big[(1-d_{x,t+i-1}(\overrightarrow{\omega}))\rho^{pr}_{R}(j,x)+ \\ &d_{x,t+i-1}(\overrightarrow{\omega})\big[\rho^{pr}_{D}(j,x,y_{j,e,s,\eta,a}) - \phi_D[\rho^{pr}_{R}(j,x) - \rho^{pr}_{D}(j,x,y_{j,e,s,\eta,a})]\big]\Big]\Omega_{t+i-1}d(\overrightarrow{\omega}|j=i), \end{aligned} \tag{22}$$

where β is the lender's discount factor and τ_{is} is a student loan issuance cost.

Social Security transfer function: Social Security transfers replace a fraction χ of the average labor earnings for the 30 years before retirement conditional on education and skill plus the average unconditional labor earnings for the 30 years before retirement, divided by two. The transfer function is given by

$$ss_{e,s} = \frac{\chi}{2}\left[\frac{\int w_e\eta\epsilon_{j,e,s}\Omega_t d(\overrightarrow{\omega}|18\leq j<j_r, e, s)}{\int\Omega_t d(\overrightarrow{\omega}|18\leq j<j_r, e, s)} + \frac{\int w_e\eta\epsilon_{j,e,s}\Omega_t d(\overrightarrow{\omega}|18\leq j<j_r)}{\int\Omega_t d(\overrightarrow{\omega}|18\leq j<j_r)}\right] \tag{23}$$

Definition Given an initial level of capital stock K_0 and an initial distribution over idiosyncratic states $\Omega_0\,(\overrightarrow{\omega})$, a competitive equilibrium consists sequences of household value functions $\{\hat{W}_t(\overrightarrow{\omega}), V_t(\overrightarrow{\omega}), \hat{V}_t(\overrightarrow{\omega}), V^R_t(\overrightarrow{\omega}), V^D_t(\overrightarrow{\omega}), V^{D_f}_t(\overrightarrow{\omega}), V^{D_x}_t(\overrightarrow{\omega})\}$, household college entrance and dropout policy functions $\{\hat{d}_{e,t}(\overrightarrow{\omega}),\ \hat{d}_{d,t}(\overrightarrow{\omega}), d_{d,t}(\overrightarrow{\omega})\}$, 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 delinquency policy functions $\{d_{f,t}(\overrightarrow{\omega}), d_{x,t}(\overrightarrow{\omega})\}$, household inter vivos transfer policy function $\{b_t(\overrightarrow{\omega})\}$, production plans $\{Y_t, K_t, L_t, L_{\ell,t}, L_{h,t}\}$, tax policies $\{\gamma_t\}$, prices $\{r_t, w_{\ell,t}, w_{h,t}, r^{pr}_{SL,t}\}$, Social Security transfers $\{ss_{t,e,s}\}$, accidental bequests $\{Tr_{t,j}\}$, and 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 (1)-(6) and (15)-(20);
(ii) The saving interest rate and wage rates satisfy equations firm first order conditions;
(iii) The private student loan interest rate satisfies equation (22);
(iv) Social Security transfers satisfy equation (23);
(v) Accidental bequests are transferred to households between ages 50 and 60 $(33 \leq j \leq 43)$ after deducting expenditure on private education subsidies[57]

$$Tr_{t+1,j} = \frac{\int (1-\psi_j) a'_t(\overrightarrow{\omega})\Omega_t d(\overrightarrow{\omega}) - \kappa \int \theta^{pr}(s)\mathbb{I}_{e=h \text{ and } j\in\{1,2,3,4\}}\Omega_{t+1} d(\overrightarrow{\omega})}{\sum_{j=33}^{43} N_{t+1,j}} \tag{24}$$

where $N_{t,j}$ denotes the mass of population of age j at time t;
(vi) Government budget constraint balances as follows, by adjusting γ:

$$\int [\tau_c c_t(\overrightarrow{\omega}) + T(y_{t,j,e,s,\eta,a})]\Omega_t d(\overrightarrow{\omega}) = G_t + E_t + D_t + SS_t \tag{25}$$

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;
(vii) Labor, capital, and goods markets clear in every period t; and
(viii) $\Omega_{t+1} = \Pi_t\Big(\Omega_t\Big)$, where Π_t is the law of motion that is consistent with consumer household policy functions and the exogenous processes for population, labor productivities, skill, subjective beliefs, and the true probabilities of being allowed to continue college for each skill endowment bin and academic year.

B.3 Computational algorithm for the stationary equilibrium

1. Guess interest rate r_{guess}, wage rates $w_{\ell,\text{guess}}$ and $w_{h,\text{guess}}$, private student loan interest rate $r^{pr}_{SL,\text{guess}}$, the level parameter for the income tax rate γ_{guess}, accidental bequests $Tr_{j,\text{guess}}$, and Social Security transfers $ss_{e,s,\text{guess}}$

[57] In our baseline calibration and in 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 exceed private subsidies, then bequests would be negative, which is equivalent to a lump-sum tax.

2. Use backward induction to solve consumer problem: $j = j_f + j_a + 1, \ldots, J$ (equations (16)-(20))

3. Guess subjective value function before college, $\hat{W}_{\text{guess}}(s, \eta, a, \hat{p})$ (equation (1))

4. Use backward induction to solve consumer problem: $j = 1, \ldots, j_f + j_a$ (equations (1)-(6) and (15))

- In solving consumer problem at $j = j_f + j_a$, use $\hat{W}_{\text{guess}}(s, \eta, a, \hat{p})$ for altruistic term
- For consumers before college graduation age, not in college, and without loans, $(j \leq 4, e = \ell, a \geq 0, x = 0)$, and for consumers after college graduation age and without loans, $(j > 4, a \geq 0, x = 0)$, use golden-section search to solve consumption-savings problem. Continuous optimization is possible as these consumers will not choose delinquency
- For consumers before college graduation age and, in college or with loans $(j \leq 4, e = h$ or $a < 0$ or $x > 0)$ and, for consumers after college graduation age with loans $(j > 4, a < 0$ or $x > 0)$, use discrete grid search for optimization as these consumers may choose delinquency

5. Use new value before college to update $\hat{W}_{\text{guess}}(s, \eta, a, \hat{p})$; repeat 4.-5. until convergence

6. Guess initial distribution of 18-year-old consumers $\Omega(j = 1, s, \eta, a, \hat{p})_{\text{guess}}$

7. Simulate and solve for distribution of Ω for $j = 2, \ldots, J$

8. Use distribution of Ω for $j = j_f + j_a$ and inter vivos transfers policy function to compute new estimates for distribution of initial 18-year-old consumers $\Omega(j = 1, s, \eta, a, \hat{p})$

9. Update $\Omega(j = 1, s, \eta, a, \hat{p})_{\text{guess}}$ and repeat 7.-9. until convergence

10. Given the stationary distribution of Ω for $j = 1, \ldots, J$, solve for new guesses:

- Compute interest and wage rates from the firm's first order conditions
- Compute private loan interest rate using zero-expected-profit condition (equation (22))
- Compute the level parameter for the income tax rate using the government budget constraint (equation (25))
- Compute accidental bequests and Social Security transfers (equations (24) and (23))

11. Update guesses in 1., and repeat steps 2.-11. until convergence

Solving for the transition path is analogous, except there are time subscripts for all value functions, policy functions, prices, taxes, transfers, and distributions.

B.4 Measuring welfare

Let value functions with a tilde denote expected lifetime utilities computed by the planner. For $j = j_f + j_a + 1, \ldots, J$, the values computed by the planner are equal to that of the consumer (i.e., $\tilde{V}(\vec{\omega}) = V(\vec{\omega})$). They are equal because subjective beliefs about being allowed to continue in college only affects the college enrollment decision, the inter vivos transfer decision, and the decisions leading up to and including the age at which the inter vivos transfer decision is made $(j_f + j_a)$. For $j = j_f + j_a$, the age at which the consumer makes the inter vivos transfer decision, the planner's value function is given by

$$\tilde{V}(j, e, s, \eta, a, x) = \sum_{s_c} \pi_{s_c}(s_c|e) \sum_{\hat{p}} \pi_{\hat{p}}(\hat{p}|s_c)[(1 - d_f)(1 - d_x)\tilde{V}^R(j, e, s, \eta, a, x, s_c, \hat{p})+ \tag{26}$$
$$d_f(1 - d_x)\tilde{V}^{D_f}(j, e, s, \eta, a, x, s_c, \hat{p}) + (1 - d_f)d_x\tilde{V}^{D_x}(j, e, s, \eta, a, x, s_c, \hat{p}) + d_f d_x\tilde{V}^D(j, e, s, \eta, a, x, s_c, \hat{p})]$$

In computing $\tilde{V}(\cdot)$, the planner takes as given the delinquency decisions $d_f(\cdot)$ and $d_x(\cdot)$, which solve equation (5). The values for $\tilde{V}^R(\cdot)$, $\tilde{V}^{D_f}(\cdot)$, $\tilde{V}^{D_x}(\cdot)$, and $\tilde{V}^D(\cdot)$ are given by

$$\tilde{V}^R(j, e, s, \eta, a, x, s_c, \hat{p}) = U(c, j, e) + \beta\psi_j E_{\eta'|e,\eta}\tilde{V}(j + 1, e, s, \eta', a', x') + \beta_c E_{\eta'|\ell}\tilde{W}(s_c, \eta', b, \hat{p})$$
$$\tilde{V}^{D_f}(j, e, s, \eta, a, x, s_c, \hat{p}) = U(c, j, e) - \xi_D + \beta\psi_j E_{\eta'|e,\eta}\tilde{V}(j + 1, e, s, \eta', a', x') + \beta_c E_{\eta'|l}\tilde{W}(s_c, \eta', b, \hat{p})$$
$$\tilde{V}^{D_x}(j, e, s, \eta, a, x, s_c, \hat{p}) = U(c, j, e) - \xi_D^{pr} + \beta\psi_j E_{\eta'|e,\eta}\tilde{V}(j + 1, e, s, \eta', a', x') + \beta_c E_{\eta'|l}\tilde{W}(s_c, \eta', b, \hat{p})$$
$$\tilde{V}^D(j, e, e, \eta, a, s, x, s_c, \hat{p}) = U(c, j, e) - \xi_D - \xi_D^{pr} + \beta\psi_j E_{\eta'|e,\eta}\tilde{V}(j + 1, e, s, \eta', a', x') + \beta_c E_{\eta'|l}\tilde{W}(s_c, \eta', b, \hat{p})$$

where $\tilde{W}(\cdot)$ is the value before college computed by the planner (given below) and policy functions $\{c(\cdot), a'(\cdot), b(\cdot)\}$, taken as given, solve equation (6) and the parent's delinquency value functions at age $j = j_f + j_a$. These value functions are the first of the two instances in which the planner's computation differs from that of the consumer with subjective beliefs. Note that the planner uses $\tilde{W}(\cdot)$, whereas the consumer with subjective beliefs uses $\hat{W}(\cdot)$. For $j = 5, \ldots, j_f + j_a - 1$, the planner's value function is computed analogously. For $j = 4$, the planner's value of college is given by

$$\tilde{V}(j, h, s, \eta, a, x, \hat{p}) = U(c, j, h) - \xi_L \mathbb{I}_{a \geq 0 \text{ and } x=0 \text{ and } (a'<0 \text{ or } x'>0)} - \xi_L^{pr}\mathbb{I}_{x=0 \text{ and } x'>0} \tag{27}$$
$$+ \beta\psi_j[p_c(j, s)E_{\eta'|h,\eta}\tilde{V}(j + 1, h, s, \eta', a', x') + (1 - p_c(j, s))E_{\eta'|\ell,\eta}\tilde{V}(j + 1, \ell, s, \eta', a', x')]$$

The planner's value of college for $j = 1, 2, 3$ and the planner's value of not going to college (as well as the value of dropping out) for $j \leq 4$ are computed analogously. Finally, the planner's value before college is given by

$$\tilde{W}(s, \eta, a, \hat{p}) = q(s)[(1 - \hat{d}_e)\tilde{V}(1, \ell, s, \eta, a, x = 0) + \hat{d}_e\tilde{V}(1, h, s, \eta, a, x = 0, \hat{p})] \tag{28}$$
$$+ (1 - q(s))\tilde{V}(1, \ell, s, \eta, a, x = 0)$$

where the planner takes as given the enrollment decision $\hat{d}_e(\cdot)$, which solves equation (1). This value function is the second of the two instances in which the planner's computation differs from that of the consumer with subjective beliefs. The planner uses $\tilde{V}(\cdot)$, which uses the true probability $p_c(j,s)$ for the likelihood of being allowed to continue college, whereas the consumer with subjective beliefs uses $\hat{V}(\cdot)$, which uses the subjective belief probability $\hat{p}$ for the likelihood of being allowed to continue in college.

To measure welfare changes for the 18-year-old consumer, we use two statistics: (1) the share of the population that is strictly worse off and (2) consumption-equivalent variation. Following Abbott, Gallipoli, Meghir, and Violante (2019), we measure consumption equivalence units relative to the value of not going to college in the initial stationary equilibrium. We do this because the value of not going to college does not include any utility (psychic) fixed costs. For the average 18-year-old in period t of the transition to the new stationary steady state, the consumption equivalent variation, $g_{c,t}$, is computed using the following equation

$$(1+g_{c,t})^{1-\sigma}\int \tilde{V}_{\text{initial}}(1,\ell,s,\eta,a,x=0,\hat{p})\Omega_{\text{initial}}d(\overrightarrow{\omega}) = \int \tilde{W}_t(s,\eta,a)\Omega_t d(\overrightarrow{\omega}) \tag{29}$$

where on the left-hand side of the equation, "initial" refers to the initial stationary equilibrium. To compute the resulting gains or losses from a policy change in consumption equivalent units, we report the difference between period t and the initial stationary equilibrium: $100\times(g_{c,t}-g_{c,\text{initial}})$. When measuring welfare holding the distribution of 18-year-old consumers fixed to that from the initial stationary equilibrium, we use distribution Ω_{initial} instead of Ω_{t} for the right-hand side of equation (29).

C Results Appendix

C.1 Baseline initial steady state: additional model validation

College wage premiums by skill Table 32 reports the college wage premium by skill tercile in the data and the baseline model. Data moments are from the NLSY97, as reported in Table 24 of Appendix A.1.2. The college wage premium in the model is the median earnings for an individual with a four year college degree divided by the median earnings for an individual without a four year college degree for workers in the age group from 25 to 39 given their skill level (ages are chosen to match the NLSY97 sample). While the wage premium for the middle skill tercile was targeted in our calibration, the model does remarkably well in explaining college wage premiums for all skill endowment bins. Specifically, the college wage premium is increasing in skill. As indicated by the enrollment rates reported in Table 10, the enrollment rate is increasing in skill in the baseline equilibrium, implying that the marginal returns to college are lower than the average returns.[58]

Table 32: College wage premiums by skill endowment

Skill	Data	Model
1	1.33	1.37
2	*1.41*	*1.41*
3	1.57	1.56

Notes: Table 32 reports the college wage premium in the NLSY97 and in the baseline model.

Subjective beliefs by enrollment status and skill Table 33 reports subjective beliefs in the baseline calibration by enrollment status and skill bin. The difference between the reported mean expectations about BA attainment and the realized graduation rate in the model matches that observed in the data (from Tables 2 and 3) although these moments were not directly targeted in the calibration.

Table 33: Subjective beliefs by enrollment status and skill endowment

		Model			Data
	Skill	(a) **Expected graduation prob.**	(b) **Realized graduation rate**	**Difference** $(a)-(b)$	**Difference** $(a)-(b)$
Panel A: Enrollees	1	87.92	36.97	50.95	49.80
	2	90.85	58.23	32.62	31.47
	3	92.74	70.31	22.44	15.36
Panel A: Non-enrollees	1	66.49	36.97	29.52	32.74
	2	66.62	58.23	8.38	13.36
	3	65.68	70.31	-4.62	-6.02

Notes: Table 33 reports subjective beliefs about college graduation likelihood by skill endowment bin from the model survey on expectations about BA attainment by enrollment status and skill bin, along with the realized graduation rate of the skill bin for those who enroll in college. The difference refers to the difference between the reported mean expectations and the realized graduation rate. The estimated differences in the NLSY97 data are also included for comparison (see Tables 2 and 3). Expectations, graduation rates, and differences are all in units of percentages.

Student loan incidence by persistence status Table 34 reports loan uptake by persistence status for a given cohort of enrollees in the data (Panel A, from Table 4 in Section 2.2), in the model baseline (Panel B), and in a partial equilibrium counterfactual in which we shut off subjective beliefs by setting $\hat{p} = p_c(j, s)$ for all j and s but do not allow general equilibrium objects to adjust (Panel C). Although the data moments are untargeted in our calibration, the baseline model does reasonably well in accounting for aggregate balance shares in column (2) and the magnitude of loan

[58] Alternatively, note that in the main text we perform a quasi-experimental study in our baseline calibration in which we increase the tuition subsidy by 1,000 dollars. In this exercise we observe a decline in the average college wage premium, indicating that the marginal returns to college are lower than the average returns in the baseline initial economy.

balances among student debtors in columns (4) and (5). However, the model does not perform well in capturing the share of non-persisters with any student debt in column (3). We attribute this to fewer dropouts with small loan balances in the model as compared to the data. A comparison of Panels B and C in Table 34 indicates that student loan statistics by persistence status barely change when beliefs are corrected. These statistics indicate that subjective beliefs do not affect borrowing behavior conditional on enrollment in college.

Despite the similarity in loan statistics across Panels B and C in Table 34, one should not infer that the intrinsic riskiness of college as an investment is the sole driver of total debt held by dropouts in our baseline model, with subjective beliefs playing no role. In fact, although enrollment statistics are not shown in Table 34, when beliefs are corrected (the equilibrium of Panel C), in comparison to the baseline (the equilibrium of Panel B), the total mass of enrollees decreases leading to a fall in the total mass of dropouts. Consequently, the total mass of dropouts with a student loan and the total amount of debt held by dropouts decreases by 25 and 21 percent, respectively.

Table 34: Student loans by persistence status

Panel and Source	Persistence status	(1) % of enrollees	(2) % of SL $	(3) % with SL	(4) Average $	(5) Median $
A: Data	Did not persist	24	19	78	15,270	12,238
	Persisted	76	81	65	24,648	19,500
B: Baseline	Did not persist	25	8	20	20,687	16,849
	Persisted	75	92	58	25,293	12,169
C: Baseline, corrected beliefs	Did not persist	22	7	18	21,891	16,849
	Persisted	78	93	56	25,819	12,169

Notes: Table 34 reports loan uptake patterns by persistence status to the third academic year for a given cohort of enrollees. Panels A, B, and C contain moments from the HSLS:09, as reported in Table 4, the model baseline equilibrium, and when $\hat{p} = p_c(j, s)$, so that there is no optimism or pessimism and consumers have correct beliefs, but general equilibrium objects are not allowed to adjust.

C.2 Main experiment: additional results

C.2.1 Proof of Proposition

Proposition. *In a partial equilibrium economy without parental altruism, transitioning from non-enrollee to an over-enrolled college student is both sufficient and necessary to suffer welfare losses after the loan limit expansion.*

Proof. Let $\hat{V}_{0,h}$, $V_{0,h}$, $\hat{V}_{0,\ell}$, and , $V_{0,\ell}$ denote, in the status quo economy, the subjective value of college, the value of college with correct beliefs, the subjective value of not going to college, and the value of not going to college with correct beliefs, respectively. Let $\hat{V}_{1,h}$, $V_{1,h}$, $\hat{V}_{1,\ell}$, and , $V_{1,\ell}$

denote the analogous values in an economy with a higher federal student loan limit (post-policy economy).

Suppose individuals are optimistic about graduation such that $\hat{V}_{0,h} > V_{0,h}$ and $\hat{V}_{1,h} > V_{1,h}$. Without an altruistic motive to make a transfer to a child in the future, $\hat{V}_{0,\ell} = V_{0,\ell}$ and $\hat{V}_{1,\ell} = V_{1,\ell}$ because subjective beliefs do not affect the value of not going to college. Furthermore, in partial equilibrium without an altruistic motive to make transfers to future children, $V_{0,\ell} = V_{1,\ell}$.

For an 18-year-old that chooses non-enrollment in the status quo economy, it must be that $\hat{V}_{0,h} < \hat{V}_{0,\ell}$. Their realized value is $V_{0,\ell}$.

If this individual chooses non-enrollment in the post-policy economy, they do not experience a welfare gain or loss because the post-policy realized value is $V_{0,\ell} = V_{1,\ell}$.

If this individual chooses enrollment in the post-policy economy, it must be that $\hat{V}_{1,h} > \hat{V}_{1,\ell}$. The realized value in the post-policy economy for this individual is $V_{1,h}$. This individual is over-enrolled if $V_{1,h} < V_{1,\ell}$. This individual is strictly worse off if $V_{1,h} < V_{0,\ell}$. Because $V_{0,\ell} = V_{1,\ell}$ in a partial equilibrium without altruism, the criteria for being strictly worse off and for being an over-enrollee are the same.

Furthermore, it is straightforward to establish that an individual that enrolls in the pre-policy economy is never strictly worse off with a limit expansion. Therefore, a non-enrollee in the status quo economy becoming an over-enrolleed college student in the post-policy economy is both a sufficient and necessary condition for being strictly worse off.

C.2.2 Discussion of general equilibrium adjustments

The effects of expanding the federal loan limit to $\bar{A} = 4$ on the baseline model's steady state equilibrium are shown in Table 35. The effects on the model economy are summarized by changes in education and skill statistics (Panel A), macroeconomic aggregates (Panel B), and prices, income tax rate, and transfers (Panel C).

The first row of Panel A reports changes in the enrollment rate by skill. The expansion in the federal loan limit increases enrollment especially for the low and medium skill endowment bins. Enrollment increases because young adults previously constrained in their access to federal credit, which has a lower uptake cost compared to private loans, can now access more of it. The next row of Panel A indicates that the expansion in enrollment leads to a lower graduation rate overall. This is because the average college student now has lower skill and is therefore less likely to graduate. Nevertheless, higher enrollment also increases the share of college graduates in the population.

Moving to Panel B, the increase in the mass of college graduates increases the total efficiency units

Table 35: Steady state changes

Panel	Variable	Changes from initial equilibrium
A: Education and skill statistics	College enrollment rate by s	(13.62,11.78,1.33)
Units: percentage point change	Graduation rate	-2.31
	Population share college graduates	4.89
B: Macroeconomic aggregates	Low-education labor (efficiency units)	-6.66
Units: percentage change	High-education labor (efficiency units)	14.71
	Labor	1.94
	Capital	-1.25
	Output	0.78
	Consumption	0.41
C: Prices, income tax rate, transfers	Risk-free savings interest rate	0.25
Units: percentage point/percentage change	Wage rate for low-education	0.64
	Wage rate for high-education	-3.42
	Private student loan interest rate	-
	Income tax rate | Baseline mean income	0.22
	Inter vivos transfers	-21.09
	Accidental bequests	1.62
	$\mathrm{ss}_{\ell,s}$ by s	(0.98,0.99,0.97)
	$\mathrm{ss}_{h,s}$ by s	(-1.57,-1.53,-1.57)

Notes: Table 35 provides results from a steady state comparison of an expansion in the federal student loan limit expansion to fund four years of college tuition plus room and board net of grants (i.e., $\bar{A} = 4$) in the baseline economy. Panels A, B, and C report changes in education and skill statistics, macroeconomic aggregates, and prices, income tax rate, and transfers, respectively. Statistics that vary over s are presented as a tuple in the order (s_1, s_2, s_3).

of high-education labor, which outweighs the fall in the total efficiency units of low-skill labor, leading to an increase in aggregate labor. Aggregate capital declines because the limit expansion decreases the incentive to save for inter vivos transfers for children, and also because there are more dropouts with student debt who would have had higher savings in the pre-policy economy: in the new equilibrium, total assets among consumers until the age of the inter vivos transfer is 8.2 percent lower, but older consumers are richer in assets. The increase in aggregate labor outweighs the decline in aggregate capital, which increases output and consumption.

In Panel C, the risk-free interest rate on savings rises because aggregate labor increases and aggregate capital declines. With fewer low-education workers and more high-education workers, the wage rate for low-education workers increases and the wage rate for high-education workers decreases. By construction, the private student loan market completely shuts down when students can use federal loans to pay for all college costs, because the borrowing limit on private loans is set as the residual of what can be financed with financial aid. The average income tax rate increases slightly because government expenditure on the federal student loan program, public grants, and Social Security transfers increases. Accidental bequests rise because older consumers are richer in assets: retirees in the new equilibrium are 1.5 percent richer in wealth. The signs of Social Security transfers reflect the signs of the wage rate of the respective education groups: the transfers increase for low-education retirees and fall for high-education retirees.

C.2.3 Isolating general equilibrium effects on welfare

In Table 36, we isolate the impact of general equilibrium objects on the population that is strictly worse off from a limit expansion in the baseline model. The table establishes that the decline in the wage rate for high-education workers is the primary driver of welfare losses for the high-skill. For the the low-skill, we observe that roughly one third of the population is always strictly worse off, which further emphasizes that welfare losses for that group is primarily driven by optimism, and not general equilibrium effects.

Table 36: 18-year-olds strictly worse off by skill: Decomposing general equilibrium effects

		Skill		
Equilibrium	**Total**	**Low**	**Medium**	**High**
Partial (all)	12	32	3	0
General	34	34	36	30
$w_{t,\ell} = w_{initial,\ell}$	39	44	42	32
$w_{t,h} = w_{initial,h}$	13	36	2	0
$r_t = r_{initial}$	41	51	41	31
$\gamma_t = \gamma_{initial}$	28	33	19	30
$Tr_{j,t} = Tr_{j,initial}$	35	36	40	30
$ss_{e,s,t} = ss_{e,s,initial}$	32	35	30	30

Notes: Table 36 reports the share of 18-year-olds that are strictly worse off in total and by skill in the baseline in the following cases: partial equilibrium where all general equilibrium objects are held fixed (that is, income tax rate, prices, bequests, Social Security transfers, and the 18-year-old distribution are fixed at their initial steady state values); general equilibrium; wage rate for low-education workers, $w_{t,\ell}$, fixed at its initial level; wage rate for high-education workers, $w_{t,h}$, fixed at its initial level; risk-free savings rate, r_t fixed at its initial level; income tax level parameter, γ_t, fixed at its initial level; accidental bequests, $Tr_{j,t}$, fixed at its initial level; and Social Security transfers, $ss_{e,s,t}$, fixed at their initial level. For each partial equilibrium case in which an individual general equilibrium object is held fixed, while the relevant variable is fixed at its initial level, the other variables change as they do in general equilibrium.

C.2.4 Welfare implications along the transition path

Figure 3 plots consumption-variation estimates in each period of the transition for 18-year-old consumers from the lowest family income tercile with high expectations about BA attainment (80-100 percent), and have either low or high skill.[59] The figure shows that the welfare estimates for these consumers do not change drastically from their values in the first few periods of the transition path as the economy transitions to the new steady state.

C.3 Sensitivity analyses

In this section, we perform several sensitivity analyses by considering alternative variations of the model with and without subjective beliefs. In each case, the model variation is re-calibrated

[59] When we compute transition dynamics, 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.

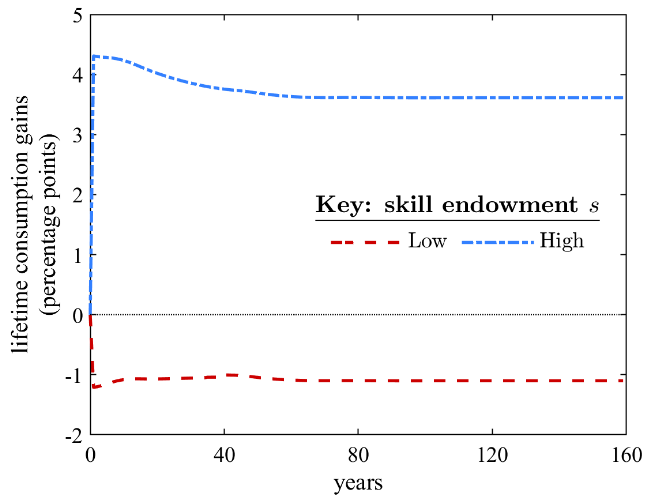

Figure 3: CEV for low and high skill endowment bins, for low family income and high expectations

Notes: Figure 3 plots consumption-equivalent variation (CEV) estimates in percentage points for 18-year-old consumers with low or high skill who are from the lowest family income tercile and have high expectations about their likelihood of earning a BA (80-100 percent), in each period of the transition path in general equilibrium.

to target the same set of moments as the baseline calibration to the extent possible. The welfare implications from the limit expansion are shown in Tables 38 and 39.

No learning about subjective beliefs In this sensitivity analysis, we consider the case in which students never learn their true probabilities of being allowed to continue in college and continue to maintain their subjective beliefs for the whole duration of college. A comparison of the baseline with this model variation in Tables 38 and 39 shows that the welfare implications of the limit expansion barely change. The assumption about learning does not matter because most drop outs happen between the first and second year of college.

Higher add-on for federal student loans In the baseline model, we abstracted from unsubsidized loans and loan fees, which meant the baseline model underestimated the cost of borrowing from the federal student loan program. In this sensitivity analysis, we consider the case in which students pay a higher add-on to the federal student loan interest rate by increasing τ_{SL} from 0.0205 to 0.0305. Tables 38 and 39 show that the welfare implications do not change much in this case as well. The small impact of raising the add-on to federal student loan interest rates suggests that, in the baseline specification, students are not highly responsive to small changes in the cost of borrowing.

College tuition that depends on skill In our baseline calibration, college tuition κ does not depend on skill. In reality, high skill students are more likely to attend higher quality colleges that cost more.[60] In this sensitivity analysis, we consider the case where college tuition κ depends on skill. We use average tuition estimates by skill reported in Table 25 as target moments. Tables 38 and 39 show that the key welfare insights from the main experiment do not change.

Lower private loan uptake cost In the baseline model, we calibrated the private loan uptake cost, ξ_L^{pr}, to target the total loan uptake of private students loans. This cost generates a pecking order in the model consistent with the data, where students borrow from the federal student loan program before turning to private lenders. The first two rows of Table 37 compare the student loan portfolio composition in the HSLS:09 data (repeated here from Table 6 in Section 2.3) and the baseline model. While the baseline model does remarkably well in explaining the student portfolio composition observed in the data, one could argue that in the data, the uptake of "Only private" loans is 2 percent, whereas that statistic in the baseline model is 0. In this sensitivity analysis, we calibrate the private loan uptake cost to target the "Only private" loan uptake of 2 percent instead of the "Any private" loan uptake of 22 percent. In this calibration, ξ_L^{pr} is equal to 0.692, as opposed to 3.146 in the baseline calibration, so the cost is almost 80 percent lower.[61] Tables 38 and 39 show that the key qualitative welfare insights from the main experiment do not change.

Table 37: Student loan portfolio composition in data and model

Case	Either	Only federal	Only private	Both	Any private
Data: HSLS:09	65	44	2	20	22
Model: baseline	58	36	0	22	22
Model: lower private loan uptake cost	59	9	2	48	50

Notes: Table 37 reports the share of students who owe money for either, only federal, only private, both types, or any private student loans three years after enrollment in the data, the baseline model, and a variation of the model that is re-calibrated with a lower private loan uptake cost. Numbers in italics in the model rows are calibration targets to discipline the loan uptake costs. Percentages are rounded to the nearest percentage point, so the sum of the last three columns may not exactly equal the value in the first column.

Skill does not depend on parental education In our baseline calibration, the child's skill depends on parental education; our estimates presented in see Table 25 indicate that high education parents are more likely to have children with higher skill. In this sensitivity analysis, we consider the case where the child's skill does not depend on parental education. We do this by setting $\pi(s_c|e) = 1/3$ for all s_c and e. Tables 38 and 39 show that the key takeaways from the main experiments do not change.

[60] The higher benefits of college for higher skill students is captured through the higher college wage premium in our model.

[61] A lower cost for ξ_L^{pr} generates a positive "Any private" loan uptake for the following reason. In our baseline model framework, the only benefit of private student loans over federal student loans is that college enrollees can save when they borrow from private lenders, but cannot save when they borrow from the federal student loan program.

Lower substitutability between low- and high-education labor In this sensitive to analysis, we allow for lower substitutability between low- and high-education labor in comparison to the estimate used in the baseline. We consider setting $\iota = 1 - \frac{1}{3.32} = 0.70$, where 3.32 is the elasticity of substitution and represents the average of the estimate of Goldin and Katz (2007) and the midpoint of the range of 4 to 6 reported in Card and Lemieux (2001); this average is the value used for the analogous parameter to ι in Abbott et al. (2019). The key insights about welfare from the main experiment do not change.

Higher (perfect) substitutability between low- and high-education labor In this sensitivity analysis, we allow for perfect substitutability between low- and high-education labor. The key insights about welfare from the main experiment do not change.

Table 38: 18-year-olds strictly worse off by skill

	(I) Subjective beliefs				(II) No subjective beliefs			
		Skill				Skill		
Exercise	**Total**	**Low**	**Medium**	**High**	**Total**	**Low**	**Medium**	**High**
Baseline								
Partial	12	32	3	0	0	0	0	0
General	34	34	36	30	18	0	19	35
No learning								
Partial	13	33	2	1	0	0	0	0
General	34	35	36	31	18	0	19	35
Higher add-on								
Partial	12	31	3	0	0	0	0	0
General	38	40	42	32	4	0	3	11
Tuition | Skill								
Partial	12	34	0	0	0	0	0	0
General	66	91	73	32	41	53	34	35
Lower private loan uptake cost								
Partial	50	92	45	8	0	0	0	0
General	29	13	36	39	12	0	0	35
Skill independent of parents								
Partial	12	29	8	0	0	0	0	0
General	36	34	51	23	30	0	37	53
Lower substitutability between low- and high-education labor								
Partial	12	31	3	0	0	0	0	0
General	33	25	42	31	17	0	15	35
Higher (perfect) substitutability between low- and high-education labor								
Partial	14	34	7	0	0	0	0	0
General	60	89	62	30	36	63	18	26

Notes: Table 38 reports the share of 18-year-olds that are strictly worse off in total and by skill in the limit expansion in the model with subjective beliefs and in an alternative without subjective beliefs. "Partial" refers to 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. "General" refers to general equilibrium. For the welfare comparison in general equilibrium, the final steady state distribution of 18-year-olds was used. Welfare is reported for the following cases: baseline, students do not update subjective beliefs for the whole duration of college, higher add-on for the federal student loan interest rate, college tuition depends on skill, lower uptake cost for private loans, skill does not depend on parental education, lower substitutability between low- and high-education labor, and higher (perfect) substitutability between low- and high-education labor. Each variation of the model is re-calibrated.

Table 39: CEV for low income, low skill, and high belief 18-year-olds

Exercise	**(I) Subjective beliefs**	**(II) No subjective beliefs**	**Difference**
Baseline	-1.10	0.95	2.05
No learning	-1.06	0.95	2.01
Higher add-on	-1.52	0.81	2.33
Tuition | Skill	-1.75	0.07	1.82
Lower private loan uptake cost	-0.40	0.89	1.29
Skill independent of parents	-1.26	0.86	2.12
Lower substitutability: Low- and high-education labor	-0.28	1.36	1.64
Higher (perfect) substitutability: Low- and high-education labor	-1.89	0.05	1.94

Notes: Table 39 reports consumption-equivalent variation estimates in percentage points for an 18-year-old from a family in the lowest income tercile who has low skill and has the highest expectations (80 to 100 percent) about BA attainment in the economy with subjective beliefs and for an 18-year-old from a family in the lowest income tercile who has low skill in the economy without subjective beliefs in the following cases: baseline, students do not update subjective beliefs for the whole duration of college, higher add-on for the federal student loan interest rate, college tuition depends on skill, lower uptake cost for private loans, skill does not depend on parental education, lower substitutability between low- and high-education labor, and higher (perfect) substitutability between low- and high-education labor. To compute welfare, we compare the initial steady state value to the corresponding final steady state value in each skill, family income, and beliefs bin. Each variation of the model is re-calibrated.