Financial Pressure and Career Choices*

Sergi Quintana

Universitat Autònoma de Barcelona and Barcelona School of Economics

January 9, 2025

Job Market Paper

Latest version here

Abstract

This paper examines the impact of student debt repayment plans on educational attainment and field of study choices of U.S. youth. To this end, I develop and estimate a dynamic human capital investment model in which highly heterogeneous individuals make college enrollment, field of study, labor supply and financial choices over their early career. I fit my model to data from the National Longitudinal Survey of Youth 1997 (NLSY97) that allows me to disentangle the determinants of educational and major choices as well as college debt. I then evaluate the effect of a very generous repayment plan introduced in 2023 known as "Saving on Valuable Education" (SAVE). Results indicate that the implementation of SAVE increases graduation rates by an average of 2 percentage points, with a more pronounced effect among low-income individuals, who are more responsive to the policy. This increase primarily results from a reduction in dropout rates among financially constrained students. Additionally, the field composition of the economy shifts, particularly among low-income students, who are 50% more likely to change fields than their high-income counterparts.

*I am very grateful to Joan Llull for his continuous support and mentorship during my doctoral studies. This paper benefited greatly from my visit to the Department of Economics at Duke Universtiy, sponsored by Arnaud Maurel to whom I am especially thankful. I also want to thank Peter Arcidiacono, Olivier De Groote, Ursula Mello, Andew Proctor and Hanna Wang for insightful comments at different stages of the project. I am also thankful for the discussions of Elia Benveniste, Elliot Motte and Sijia Pan. I acknowledge funding from Spanish Ministry of Science and Innovation through FPI Grant PRE2021-099489 and support by the European Research Council (ERC) through Starting Grant n. 804989. All remaining errors are mine.

1 Introduction

Over the past few decades, college tuition in the U.S. has risen sharply, driven by increasing returns to college education and rising enrollment rates (Dynarski et al., 2023). In this context, student loans have become an essential means of financing higher education, growing to be the second-largest source of household debt, surpassed only by mortgages. The total outstanding balance is $2.1 trillion as of 2024.[1] For the 2025 fiscal year, the U.S. Department of Education estimates that new government loans will amount to $93.1 billion, comprising 69% of all new postsecondary student aid, which includes loans, grants, and work-study programs.

Although the existence of student loans allows individuals to invest in their human capital without providing any collateral, uncertainty about these investments carries risk. Repayment plans are then a crucial element in shaping this uncertainty, providing insurance against future consumption. For many years, the standard repayment option has been a fixed 10-year plan, in which students make 120 equal payments irrespective of their financial circumstances. This has led many individuals into missing payments (Looney and Yannelis, 2024) and has become a key issue in the U.S. political debate. In 2023 the "Saving on Valuable Education" (SAVE) repayment plan was introduced, designing a payment scheme that is more closely related to individual earnings. This plan allows for zero payments when income falls below a threshold and establishes a repayment plan with amounts that will never exceed a percentage of ones income. It also includes loan forgiveness on any remaining balance after a designated period post-graduation.

The risks associated with student loans are closely tied to the type of educational investment being made. While returns vary by institution (Lovenheim and Smith, 2023), significant differences also exist across fields of study. In fact, the earnings differential between some fields can be as large as the premium associated with a college degree over a high school diploma (Altonji et al., 2012). Despite these substantial differences in expected earnings, research on major selection shows that students highly value non-monetary returns from their chosen fields, such as personal fulfillment, social impact, and alignment with intrinsic interests (Patnaik et al., 2021). Understanding how repayment risk influences enrollment and field-of-study decisions—particularly how it generates trade-offs between non-monetary preferences and financial returns—remains an open question in the literature, with significant policy implications.

The goal of this paper is to investigate the role played by student loan repayment on individuals' enrollment decisions and their sorting across fields of study. To analyze these policies, I develop a human capital investment model in which individuals with heterogeneous characteristics make educational decisions based on their preferences, financial resources and expected uncertain returns from their investment. In the model, students endogenously accumulate student loans to meet their financial needs. Risk aversion and uncertainty about college investment returns will play a crucial role in shap-

[1] According to the Student Loan Debt Clock. The balance was checked in October 2024. See https://finaid.org/loans/studentloandebtclock/

ing educational decisions and debt choices. I fit the model to rich individual career choice and financial data from the National Longitudinal Survey of Youth 1997 (NLSY97). With the estimated model I simulate a baseline scenario in which students follow the standard 10-year repayment plan and compare it with a scenario where the SAVE plan is introduced, offering a more comfortable repayment option. To further explore the role of student loans in higher education decisions, I also evaluate a scenario in which borrowing is unavailable. Results show that individuals value the insurance provided by SAVE, with a positive effect on enrollment and graduation, a reduction in drop-out rates and a change in the field composition of the economy.

In my model, individuals accumulate human capital over the life cycle through both educational investments and work experience. Each academic year, individuals face decisions about enrollment and labor supply participation. I allow for a broad set of fields of study and occupations to capture the diversity in preferences, paths and returns of different careers. Although individuals are rational decision-makers, they encounter several layers of uncertainty when making choices. First, they are unsure of the financial resources available if they enroll, including potential grants, parental support and wages from work. This implies they are also uncertain about the amount of student loans they might need thought their education. Additionally, they face uncertainty regarding their likelihood of graduation and the future returns on their education, which directly impacts their willingness to accumulate debt. Student loans are chosen every period after enrollment has been made and financial resources are realized. The model implies that individuals sort into fields based on their preferences, but also considering current and future financial needs, and future debt repayment capacity.

I estimate the model using the National Longitudinal Survey of Youth of 1997 (Bureau of Labor Statistics, 1997-2021), a longitudinal panel that tracks U.S. youth as they transition from high school to the labor market, covering their college experiences. The NLSY97 allows me to construct a year-by-year panel of individual enrollment spells, with detailed information on the enrollment decision, the field of study and their financial resources available while enrolled. Additionally, it includes comprehensive labor market data, such as labor supply decisions, hourly wages, and occupations. The survey also contains a rich set of demographic information, including parental income and measures of individual cognitive ability. This allows me to account for important determinants in educational choices, and to better understand the role played by student loans.

Estimating dynamic discrete-continuous choice models is a computationally complex task, which has not often been tackled in the literature. Building on the estimation methods of Hotz and Miller, (1993), Arcidiacono and Miller, (2011) and Arcidiacono et al., (2024) I reduce the computational complexity to feasible levels. More specifically, my estimation method is based on the Conditional Choice Probability (CCP) method introduced by Hotz and Miller, (1993) and further developed by Arcidiacono and Miller, (2011) to account for unobserved heterogeneity. However, given that the complexity makes Arcidiacono and Miller, (2011) too costly, I follow a process inspired by Arcidiacono et al., (2024) using an auxiliary model expanded with measures to identify latent unobserved types with the EM-algorithm. One extra layer of complexity in my model, not common in the literature, is the burden of adding an additional student loan continuous choice, inspired by Keane and Wolpin, (2001) and Johnson, (2013). As opposed to them, I do not use interpolation methods in the estimation of the parameters. By building on the previous results I can feasible accommodate the continuous choice in a rich state space with unobserved heterogeneity and a large choice set. Finally, to improve the precision of my estimates I re-estimate the CCPs using my model, following Aguirregabiria and Mira, (2002), and repeat the estimation sequence until convergence.

Using the model as a policy laboratory, I simulate how the educational choices of students would be in a world in which they face SAVE as the repayment plan. I find that the introduction of SAVE leads to an increase on college graduation of 2 percentage points. The increase represents heterogeneous relative gains across the parental income distribution, and low-income individuals benefit the most from the policy with an 18% rise in graduation, compared with just a 5% increase for high-income individuals. Furthermore, the probability of enrollment in a four-year school rises by more than 30% for low income individuals. Using my simulations, I estimate that individuals derive a high insurance value from the repayment plan. The value differs substantially depending on the characteristics of the individual. For example, I estimate that for low-income individuals this is equivalent to a grant of $25,000 for enrolled students.

Although the introduction of SAVE increases enrollment, the increase in graduation is mainly driven by students that were previously dropping out because of financial reasons, which represent around 70% of the new graduates. This is so because the marginal students that respond to the policy by enrolling in college are more likely to drop out. In particular, I document that only around 20% of new enrolled individuals achieve graduation at four-year schools, and the ones that make it tend to have a higher schooling preferences.

The change in repayment plan also has significant consequences for the labor supply decision of students while enrolled. I find that the share of individuals that do not work while enrolled grows by 8% on average. This is so because more individuals are now willing to become indebted instead of working to finance their studies. As a consequence, there are several changes in the composition of student debt of the economy. I find that the share of individuals that graduate with some debt raises by 48%, and that the average debt at graduation also increases by 25%.

I also find that 15% of the students choose different fields to study. In particular, low-income individuals are almost 50% more likely to switch fields than high-income ones. This suggests that high-income individuals were already studying their preferred field in the first place, whereas low-income field choices were more often determined by financial reasons. Furthermore, indebted individuals are more likely to change fields. In particular, this difference is 20% for low-income students, consistent with sorting as a consequence of financial pressures. The reason to change is sometimes heterogeneous, but in most of the cases is linked with the labor supply participation decisions and the graduation probability of a field. Individuals are now willing to get indebted while studying to avoid working while enrolled. In terms of the distribution of fields, the share of students that graduate in Education majors increases by 8%, which has the lowest wage after graduation, and the share in Health majors by

10%, which takes the longest to graduate. Furthermore, there is a reduction in fields like Business, often seen as a safer financial option due to easier graduation requirements and higher potential earnings.

Finally, I compare the baseline economy to the case in which student loans are no longer offered. I find that the availability of student loans, even with the standard 10 year repayment plan, increases graduation by 40% for low income individuals. This suggests that, although the current system leaves room for improvement, student loans have an important impact on reducing higher education inequality.

My paper contributes to several strands of the literature. The rapid growth of college tuition have motivated extensive research on the role of financial aid in higher education access. A significant portion of this literature has addressed how borrowing constraints and credit availability affects college enrollment decisions, with mixed findings regarding the impact of financial aid on students' willingness to invest in higher education. Early studies initially found limited evidence for the role of borrowing constraints in shaping college attendance (Kane, 1996, Cameron and Heckman, 1998, Keane and Wolpin, 2001, Keane, 2002 or Carneiro and Heckman, 2002). More recent work, however, identifies effects of financial constraints, suggesting that credit availability may influence human capital investment decisions at different stages (Belley and Lochner, 2007, Lochner and Monge-Naranjo, 2011, Johnson, 2013, Hai and Heckman, 2017, Caucutt and Lochner, 2020, Biswas, 2020). At the same time, another part of the literature has focused on the role of other financial aid, specially grants, in enrollment decisions, finding important elasticities to aid availability in enrollment (Dynarski, 2003, Avery and Hoxby, 2004, or Castleman and Long, 2016 among others, see Angrist et al., 2021 and Dynarski et al., 2023 for a detailed analysis). Finally, other studies have emphasized the role of parental transfers in supplementing financial aid, such as Abbott et al., (2019). Yet, while much attention has been paid to the availability of grants or student loans, limited research considers the role of repayment structures and their effects on students' enrollment choices and major selection.

Another strand of the literature analyses student loan repayment plans, but does not focus on how it shapes investment decisions. This is the case for Catherine et al., (2024) who also evaluates the role of the SAVE repayment plan but with a focus on re-distributional effects, without considering how it would change individual's human capital investment decisions. Similarly, Mueller and Yannelis, (2022) study income consequences of switching repayment plans, but not how it shapes choices. Among the first works that mention repayments across different majors is Lochner and Monge-Naranjo, (2015) or Beyer et al., (2015) but they just provide descriptive evidence of repayment across fields, and open the door to study those differences. Similarly, a part of the literature focus on how this repayment affects occupational choices as is the case in Rothstein and Rouse, (2011)where the authors find that loans sort individuals into higher paying occupations, but without considering how individuals decide to accumulate those loans on the first place.

This paper also builds on a large literature that studies the role of uncertainty in human capital returns in investment decisions. This uncertainty often comes from uncertain labor market returns, uncertain graduation outcomes, financial resources, or own abilities and skills (Altonji, 1993, Burland et al., 2023, Arcidiacono et al., 2024). By departing from just expected earnings and introducing expected repayment while deciding which major to study, my paper contributes to the literature by analyzing also the determinants of major choice. The literature describe as the main determinants expected earnings and non-pecuniary benefits (Altonji et al., 2012, Altonji et al., 2016 and Patnaik et al., 2021) and work finds that elasticites to expected income form fields is not as important determinant as non-pecuniary benefits (Beffy et al., 2012). My work contributes to this literature by understanding how elasticities to earnings change in the presence of repayment needs, since even if individuals do not prioritize expected earnings, repayment needs might intensify their value.

Although my paper is the first one to consider the role the insurance provided by the student loan repayment plan has on the field of study decision, there are some few exceptions that have addressed similar questions. In Arcidiacono, (2005) the author studies the role of financial aid and admission mechanisms in future earnings of black individuals considering major choice, without considering repayment plans. A work related to this one is Hendricks and Leukhina, (2017) who analyze how uncertain about college graduation and earnings shape educational decisions, and opens the door for the possibility on analyzing the insurance provided by income-contingent loans. A more recent work is Belzil et al., (2023) in which the authors study the role of different sources of financial aid on the field of study in Canada, but without considering repayment possibilities and just focusing on the loans versus grants dimension. In an early contribution on this literature Field, (2009) provides evidence of the effect of student loans on the field of study in NYU, by evaluating an RCT in the financial aid offered. In any case, any of these papers considers the role of repayment plans in providing insurance against future income risk and shaping fields of study decisions. There is some growing work that focuses on the role of student loan repayment plans in higher education decisions, without considering field of study, this is the case of Liu, (2020) and Luo and Mongey, (2024) who evaluate college access and welfare under different repayment plans, the former also considering SAVE as a possibility.

My paper builds on the framework developed by Arcidiacono et al., (2024), which examines the role of learning about individual abilities in higher education and major selection. My work departs from theirs in several key ways. First, I endogenize the student debt decision, which becomes a central element in my model. This allows me to study how debt decisions change in counterfactual scenarios. Second, I expand the set of possible fields of study and occupations allowing for a large heterogeneity in preferences. To simplify my work, I abstract from the Bayesian learning process developed in Arcidiacono et al., (2024), while still accounting for changes in preferences as college experience evolves. Finally, to improve the precision of my estimates, I use the algorithm proposed by Aguirregabiria and Mira, (2002), something that is not feasible to do in Arcidiacono et al., (2024).

The rest of the paper is structured as follows. In Section 2 I describe the data used for the paper and provide motivational evidence. Section 3 presents the structural model developed in this work. Section 4 describes the estimation of the model parameters. Section 5 describes the estimates of the model. Section 6 presents the evaluation of the SAVE repayment plan. Finally, Section 7 concludes.

2 Data and Descriptive Evidence

This study uses data from the National Longitudinal Survey of Youth 1997 (Bureau of Labor Statistics, 1997-2021). The NLSY97 contains information on 8,984 youths randomly sampled to be representative of the U.S. Individuals on the sample were born between 1980 and 1984 and are first interviewed in 1997 and followed over time until the current period (2024). The main virtuous of this data set is that it allows to create a year-by-year spell panel with information on educational choices and labor market outcomes, with very rich data on financial aid, fields and colleges, wages and occupations, complemented with family background and other demographic characteristics. Furthermore, it includes the Armed Forces Qualification Test (AFQT) which is a scored composed of Mathematical Knowledge, Arithmetic Reasoning, Word Knowledge and Paragraph Comprehension widely used in the literature as a measure of cognitive ability or IQ (Hai and Heckman, 2017). In Appendix F I provide a description of the variable creation process. Overall, I restrict my attention to High-School or GED graduates, which are the only individuals eligible to attend college. Furthermore, I only consider the first 10 academic years after an individual graduates from High-School. Although I have more data in my sample, unfortunately the computational complexity of my model does not allow to model choices after period 10. This results in a total amount of 49,820 observations for 4,982 individuals.[2] For some parts of the work I will complement also with data from the American Community Survey (ACS) pooling years 2009 to 2019.

2.1 College Enrollment and Financial Resources

In this section I document important features of the U.S. higher education system that will motivate the structural model developed in Section 3. As Table 1 shows, there is a high heterogeneity in the highest educational achievement and individual demographic characteristics. First of all, as the level of education becomes higher, the share of females increases and the share of black individuals decreases, which generates some sorting patterns. The average AFTQ percentile also increases as the education level of the individuals increases. Those that are never enrolled have an average AFQT of 39, whereas individuals that attend Graduate School have more than twice the average percentile score, with 76. A similar pattern occurs with parental income, indicating that it seems to be indeed a predictor of educational attainment. The average parental income of an individual never enrolled into school is 58,973 dollars, and it increases as the level of education becomes higher, arriving at an average of 118,081 for those individuals that attend Graduate School. This patter has been already documented by Belley and Lochner, (2007) which rise the concern about the increasing importance of ability and parental income in determining college access.

[2]Note that I am including females in my estimation sample to maximize sample size. As of now, there are not fertility decisions in my model, although this will be captured by female heterogeneous tastes for college. In the future I will include an exogenous fertility process that will shift females to Home Production with some probability to be able to better accommodate the females in my sample.

Table 1: Sample Summary Statistics: Overall Sample

	All Sample	Never Enrolled	2 Year Drop	2 Year Grad	4 Year Drop	4 Year Grad	Grad School
Demographics							
Share Female	0.50	0.40	0.53	0.55	0.51	0.58	0.63
Share Black	0.24	0.29	0.29	0.14	0.29	0.15	0.12
AFQT Score	49	31	42	50	55	69	76
Parental Income	79,673	58,873	66,343	77,393	83,038	108,885	118,081
Share ParInc Q1	0.25	0.39	0.29	0.19	0.24	0.10	0.06
Share ParInc Q4	0.25	0.12	0.18	0.24	0.26	0.42	0.49
Labor Market							
Hourly Wage Age 18	9.55	10.46	9.58	9.18	9.06	8.93	9.51
Hourly Wage Age 28	21.08	18.54	18.48	20.48	21.45	24.60	26.12
Work Part-Time Enrolled	0.43		0.37	0.42	0.39	0.47	0.24
Work Full-Time Enrolled	0.23		0.34	0.30	0.27	0.15	0.50
College Finances							
Share Grants	0.72		0.65	0.69	0.79	0.81	0.39
Grants Per Year	7,785		3,877	5,879	7,176	9,760	16,424
Share Parental Transfers	0.72		0.47	0.58	0.67	0.84	
Transfers Per Year	8,351		2,372	4,529	7,093	10,288	
Share Loans	0.57		0.32	0.55	0.66	0.67	0.45
Loans Per Year	8,000		7,448	10,956	7,932	7,847	24,580
Total Loans	21,323		11,847	20,282	18,815	25,232	53,932

Notes: This table reports the main statistics of the NLSY97 sample used for estimation. For a definition of the main variables visit Appendix F. To obtain the age of the individual, I normalize age 18 to the year in which high-school is finalized. All monetary variables are reported in 2008 dollars. The different columns represent educational categories "Never Enrolled" are those who have never attended college, "2 Year Drop" are individuals that have dropped out from a Two-Year school, "2 Year Grad" are individuals that have graduated from a Two-Year school. The same applies to Four-Year columns. "Grad School" represents individuals that attend a Graduate school.

Table A.1 shows the main college financial statistics across the parental income distribution. As we can see, there are big differences in the way individuals across the parental income distribution finance their higher education expenses. First of all, there is a negative effect on the share of individuals receiving grants as parental income increases. In both two-year and four-year schools individuals with more parental income receive less grants. Conversely, in both two-year and four-year schools individuals with more parental income receive parental transfers more often, and when receiving transfers the average amount is higher. For four-year schools an individual from the bottom quartile of the parental income distribution will receive about 5,000 dollars, compared with 14,000 dollars received from an individual from the top parental income distribution. Furthermore, because of this differences there is also heterogeneity in the share of individuals working while enrolled. In Table 1 we can see that a large share of individuals work-while enrolled. This pattern has been documented by Lovenheim and Reynolds, (2011). Furthermore, from A.1 we see that this is specially true for low-income individuals, that work full-time more often than high-income ones. This heterogeneity in preferences will make the estimation of my model crucial to disentangle the preference component from the consumption one when analyzing major choices.

Table 2: Sample Summary Statistics: Field of Study

	Business	STEM	Education	Social Sciences	Humanities	Health	Other
Demographics							
Share Female	0.5	0.38	0.82	0.61	0.62	0.79	0.54
Share Black	0.13	0.12	0.15	0.19	0.15	0.15	0.20
AFQT Score	66	77	63	65	70	69	64
Parental Income	113,279	112,613	90,555	104,447	118,009	103,534	102,095
Share ParInc Q1	0.07	0.07	0.14	0.17	0.09	0.04	0.14
Share ParInc Q4	0.41	0.44	0.29	0.38	0.48	0.47	0.37
Labor Market							
Work Part-Time Enrolled	0.42	0.42	0.45	0.45	0.45	0.46	0.45
Work Full-Time Enrolled	0.24	0.16	0.18	0.16	0.16	0.19	0.27
Wage After Graduation	22.10	23.56	17.21	19.87	18.81	23.60	18.94
College Finances							
Share Grants	0.77	0.82	0.86	0.78	0.78	0.85	0.81
Share Parental Transfers	0.84	0.87	0.80	0.83	0.88	0.89	0.75
Share Loans	0.63	0.65	0.74	0.67	0.66	0.74	0.69
Loans At Age 28	12,838	13,797	14,319	11,649	12,745	14,788	12,436

Notes: This table reports the main statistics of the sample of individuals that graduates from a specific field. Individuals that have drop out are not considered here. Loans at age 28 are computed 10 periods after the individual graduates from high-school, which is assumed to be age 18. To avoid contamination, individuals that have attended a two-year school or a graduate school are not considered for that statistic. Monetary values represent 2008 dollars. Data comes from the NLSY97 sample.

There are also well documented differences in the preferences for fields of study, as can be seen in Table 2. First of all, there are important differences in the fields that are male or female dominated. Specifically, females represent 82% in Education majors but only 38% in STEM. This fact is well documented in Speer, (2017) where the author studies what role play pre-college factors and finds that measures similars to AFQT are important determinants. This can also be seen in Table 2, since there are important differences in AFQT across fields of study. Sorting across fields of study not only depends on sex. For example, black individuals are more represented in Social Sciences than in STEM. Finally, another important factor driving major choice is parental income. In Lovenheim and Smith, (2023) the authors document that low-income individuals prefer fields with lower returns. This is clear in Table 2 since the average parental income is lower in Education. Furthermore, high-income individuals represent 44% of all STEM graduates but only 29% of Education ones.

2.2 Heterogeneous Returns to College

The labor market outcomes are different along different dimensions, as has been documented in Altonji et al., (2012) or Lovenheim and Smith, (2023) among others. This can be seen in Table 1 that reports the average wage at age 28 across educational levels. The average high school graduate that never enrolls to college will receive \$18.50 per hour compared to \$20.48 that a two-year graduate will receive or \$24.60 for four-year graduates, showing significant differences in wages. However, the heterogeneity on returns to college does not only depend on the type of institution attended, but also on the major of study. Table 3 displays the wage, unemployment and labor supply decision of individuals across different fields and points in time. As it can be seen, there are significant differences, not only on the wage that individuals have after graduation, but also on the variance of that wage and on the unemployment rates and working hours. For example the overall average wage of business graduates is very similar to the one of health graduates, but the variance of business is about 50% higher than the one in health. Furthermore, the unemployment rate of business is also twice the one in health. This suggests that different majors carry different labor market risks, and therefore different loan repayment possibilities.

3 A Human Capital Model with Financial Decisions

In this section, I develop a dynamic model of human capital investment decisions in the spirit of Keane and Wolpin, (1997) that incorporates endogenous student debt accumulation, where individuals are forward-looking. The model builds on the frameworks of Johnson, (2013) and Arcidiacono et al., (2024), with the goal of capturing key features of the U.S. higher education system. The primary trade-offs in the model involve whether to pursue further education, what field of study to select in the event of enrollment, and what type of job to take if entering the labor market. A crucial component is that individuals enrolled in education must determine how much student debt to take on.

I will now proceed with an overview of the overall functioning of the model, to later describe the different ingredients that will govern the choices of the individuals.

Table 3: Sample Summary Statistics: Labor Market for Fields

		Business	STEM	Education	Social Sciences	Humanities	Health	Other
Average Wage	Overall	32.57	35.09	21.38	28.25	25.70	31.01	24.54
		(28.47)	(26.80)	(17.58)	(26.34)	(23.98)	(19.88)	(20.65)
	Age 25	19.70	21.50	15.72	17.45	16.14	21.81	16.29
		(13.72)	(15.54)	(10.53)	(13.84)	(12.76)	(14.17)	(12.62)
	Age 40	34.05	36.75	21.29	30.97	28.46	31.92	26.25
		(24.48)	(26.07)	(16.06)	(27.26)	(25.23)	(20.12)	(20.19)
Unemployment	Overall	0.04	0.04	0.03	0.04	0.05	0.02	0.03
	Age 25	0.04	0.04	0.03	0.05	0.06	0.03	0.04
	Age 40	0.03	0.03	0.02	0.04	0.05	0.02	0.03
Share Full-Time	Overall	0.80	0.81	0.67	0.73	0.70	0.70	0.78
	Age 25	0.77	0.70	0.66	0.64	0.62	0.69	0.69
	Age 40	0.81	0.84	0.68	0.75	0.71	0.70	0.81

Notes: This table reports the main labor market statistics of the different fields of study used in this work. The table uses data from ACS pooling years 2009 to 2019. For the wage I report the average hourly wage and the standard deviation in parenthesis. The "Overall" category includes individuals from ages 25 to 65. Only individuals in the labor market are considered. Monetary amounts represent 2008 dollars.

3.1 Overview and Structure

The model aims to capture the central trade-offs individuals face when transitioning from high school to the labor market, focusing on how these choices are affected by the financing of higher education and taste towards careers.[3] The model starts when individuals complete high school and operates in discrete time fashion over 10 academic years, which is sufficient to account for the completion of key educational stages.[4] In each of these discrete time periods, agents face different education and occupation decisions that will endogenously build their human capital.

In terms of job market opportunities, individuals can choose whether to work and, if so, whether to work full-time or part-time. The available occupations include Business, STEM, Education, Social Sciences, Humanities and Arts, Health, Sales and Office, and Production, reflecting the heterogeneity of preferences within the labor market.[5] If individuals choose to work while enrolled the occupation

[3] The notion of career refers as fields of study or occupations.

[4] The model is limited to the first ten periods after high school due to the computational burden that increases with each additional period. In Appendix D, I discuss the growth of the state space over time. Restricting the education phase to the first 10 periods is reasonable, since most individuals in the sample have already done their education decisions after 10 years from high-school.

[5] Appendix D provides a detailed description of the available choices and specifies which occupations can be selected depending on the individual's level of human capital. In Appendix E I explain the occupation classification.

is considered unimportant. I will denote O the set of all possible occupations, including undefined for individuals that work while enrolled and no occupation if the individual is not working.

For the education decisions, individuals can choose whether to enroll in higher education, which field of study to pursue, whether to work while enrolled, and how much student debt to take on if they decide to enroll. Individuals without a graduate degree can choose to attend either a two-year institution to earn an Associate Degree or a four-year institution to pursue a Bachelor's Degree. If enrolling in a four-year program, individuals must also select a field of study. The available fields of study include Business, STEM, Education, Social Sciences, Humanities and Arts, Health, or Other, capturing the heterogeneity in individual preferences. I will refer to F as the set of all possible fields. Those that have an Associate Degree can only decide whether to enroll in a four-year institution, and those individuals that have a Bachelor Degree can only choose either to work or to enroll in graduate school. Finally, individuals that hold a Graduate Degree cannot make any education decision. In order to achieve graduation individuals should be enrolled at least two years in a two-year school, four-years in a four-year school or one year in a graduate school. Graduation is probabilistic and depends on individuals characteristics. This means that individuals are allowed to switch from one field to another at each period before graduation to capture the dynamics of major of choice of the U.S. higher education system.

Finally, a key aspect of the model is the student debt decision. After individuals make their education and labor market choices, they must decide how much student debt to take on. This decision depends on the financial resources available at that time, such as labor market earnings, grants, and parental transfers, as well as the expected returns from their education. However, individuals face uncertainty regarding the actual returns from their education, and this uncertainty, combined with their aversion to risk, will influence their educational and debt decisions.

Individuals may also choose not to enroll in school or participate in the labor market. In such cases, they are assumed to engage in home production, with their utility level normalized as explained in the next section.

To make decisions, individuals must form expectations about their future utility based on each possible choice. I assume that individuals are rational and have perfect information about the elements that influence the economy. This implies that they have complete knowledge of wage distributions, grants, parental transfers, and the probabilities of receiving grants and parental transfers, as well as the likelihood of graduating at each point in time and the distribution of the idiosyncratic shocks.

In the rest of this section I will describe the different elements that govern the economy, to later explain how individuals value the different alternatives and make choices.

3.2 Labor Market Occupation Decision

Individual i at period $t \in T$ will choose which occupation $o \in O$ to work at and will consequently receive a log hourly wage w_{oit}. The individual will only choose an occupation if decides to participate in the labor market. The labor supply is $l \in \{0, P, F\}$ where 0 refers to not-working, P refers to work part-time and F refers to work full-time.

The wage will be determined by a set of characteristics X_{oit} that are known to the individual and includes: time invariant characteristics (parental income, afqt, sex and race) and endogenously acquired labor market experience, education level and field of graduation.[6] The log hourly wage equation is defined as:

$$w_{oit} = \gamma_{0o} + \gamma_{1o} X_{iot} + \varepsilon_{oit} \tag{1}$$

Where ε_{oit} is an idiosyncratic shock assumed to be distributed $N(0, \sigma_o^2)$ and uncorrelated across individuals, occupations, time and other covariates. The different elements of the log wage equation allow to capture permanent skills and endogenously acquired ones, as well as human capital occupation complementarities, since the field of graduation will differently affect different occupations. A key element is the variance of each occupation wage, since it will affect the level of uncertainty of the returns to education.

3.3 Educational Market

In terms of education, the outcomes of interest are those governing the financial part of the model, which includes grants and parental transfers, as well as the graduation probabilities of agents.

Grants are assumed to be obtained with a logistic probability, and will depend on the vector X_{git} which includes individual time invariant characteristics. The functions is:

$$P(grants = 1 | X_{git}) = \frac{\exp(X_{git}\Gamma_{ge})}{1 + \exp(X_{git}\Gamma_{ge})} \tag{2}$$

where the parameters that govern the probability of obtaining grants, Γ_{ge}, are allowed to vary across educational decisions $e \in \{2y, 4y, G\}$.[7] In a similar way, parental transfers will be also obtained with a logistic probability dependent on the vector X_{pit} which includes individual observable characteristics and the education decision e. The functions is:

$$P(transfers = 1 | X_{pit}) = \frac{\exp(X_{pit}\Gamma_p)}{1 + \exp(X_{pit}\Gamma_p)} \tag{3}$$

Individual log parental transfers p_{eit} and log grants g_{eit} amounts are obtained respectively:

$$p_{eit} = X_{pit}\lambda_p + \varepsilon_{iet}^p \tag{4}$$

[6]These characteristics depend on the state space of the individual, that is described in Appendix D.

[7]The different decisions correspond to two-year schools, four-year school or graduate school. For completeness, the set of all possible education decisions can also include 0, which implies not education.

$$g_{eit} = X_{git}\lambda_{ge} + \varepsilon^{g}_{iet} \tag{5}$$

where $\varepsilon^{p}_{iet} \sim N(0, \sigma^2_p)$ and $\varepsilon^{g}_{iet} \sim N(0, \sigma^2_g)$. Importantly, because of the lack of data availability individuals will not receive parental transfers while enrolled in a graduate school.[8]

Finally the graduation probabilities are assumed to follow a logistic distribution that depends on X_{Git} and includes individual time invariant characteristics, but also field of study and dummies for current years of education, to account for a flexible function on years of education.

$$P(grad = 1|X_{Git}) = \frac{\exp(X_{Git}\phi)}{1 + \exp(X_{Git}\phi)} \tag{6}$$

Overall, these functions capture in a very simple way the dynamics of the financial aid set up of the U.S. higher education system. In particular, grants are known to be decreasing in parental income and increasing in the individuals ability, and parental transfers will be determined by parental income. The coefficient of ethnicity is supposed to capture affirmative action policies towards minorities and the coefficient on female will try to capture either positive or negative discrimination towards females.

3.4 Flow Payoff

At every period t, agents face a discrete choice which is a combination of education, labor supply, occupation and field of study decision. A choice is defined as $d_{it} = (e, l, o, f)$. Following Arcidiacono et al., (2024) the base alternative is home production, with $d_{it} = (0, 0, 0, 0)$.

The main trade off individuals will face each period is leveraging preferences for specific fields against the potential income and therefore consumption C_{ieloft} that can be obtained at each different alternative. For this reason, the payoff function will be composed of two components, a CRRA utility from consumption, and non-consumption component $u_{elof}(Z_{1it})$ that will capture preferences or abilities towards specific fields of study or occupations. I define $Z_{it} = (Z_{0it}, Z_{1it})$ to be the vector of characteristics that will affect the individuals flow payoff, Z_{oit} include the characteristics that will affect the utility only through consumption and Z_{1it} includes the characteristics that affect the utility of the choice. [9] Therefore, the period specific payoff function is defined in Equation (7) as:

$$U_{elof}(Z_{it}, b_{it}, \varepsilon_{ieloft}) = u_{elof}(Z_{1it}) + \mathbb{E}\left[\frac{C^{1-\sigma}_{ieloft}}{1-\sigma}\right] + \varepsilon_{ieloft} \tag{7}$$

where b_{it} is the individual current student loan level and ε_{ieloft} is an unobserved i.i.d preference shock assumed distributed Type 1 Generalized Extreme Value (GEV). For identification purposes I will normalize the function to $u_h(Z_{1it}) = 0$ in the event that the individual chooses home production.

[8] Note, however, that the effect of potential parental transfers will be captured by the distribution of the budget shock. This means that in reality individuals might still be receiving parental transfers, but I am not modeling how those are obtained.

[9] The vector Z_{it} is composed by the individual's state space, that is defined in Appendix D. The elements that only affect consumption is the labor market experience. The element that only affect schooling preferences is the latent schooling type.

Therefore, the interpretation of all the coefficients should be made with respect to the home production base category.

3.4.1 Consumption and Individuals Budget

The level of consumption at each discrete alternative will be determined by the financial resources of the agent once the discrete choice has been made. Since those are unknown at the moment of making the discrete choice decision, individuals will make the decision based on expected utility from consumption. Following Arcidiacono et al., (2024) the budget constraint will be a modified version of Johnson, (2013) in the sense that it binds in each period. Individuals will face a slightly different budget constraint depending on whether they are enrolled in school or not, since enrolled individuals could receive parental support or grants. The latent consumption is defined in Equation (8) as following:

$$C^*_{ieloft} = \begin{cases} W_{ilot} + G_{iet} + P_{iet} - \tau(e) + b_{elofit+1} - (1+r)b_{it} + \xi_{ieloft}, & \text{if } e \in \{2y, 4y\}, l \neq 0 \\ G_{iet} + P_{iet} - \tau(e) + b_{elofit+1} - (1+r)b_{it} + \xi_{ieloft}, & \text{if } e \in \{2y, 4y\}, l = 0 \\ W_{ilot} + G_{iet} - \tau(e) + b_{elofit+1} - (1+r)b_{it} + \xi_{ieloft}, & \text{if } e = G, l \neq 0 \\ G_{iet} - \tau(e) + b_{elofit+1} - (1+r)b_{it} + \xi_{ieloft}, & \text{if } e = G, l = 0 \\ W_{ilot} - R(b_{it}, W_{ilot}), & \text{if } e = 0, l \neq 0 \\ 0 & \text{if } e = 0, l = 0 \end{cases} \tag{8}$$

If an individual is not enrolled to school, consumption will depend in the yearly wage W_{ilot} [10] and the repayment scheme followed $R(b_{it}, W_{ilot})$. If the individual decides to enroll into school it will face two more sources of income which are grants G_{iet} and parental transfers P_{iet}.[11] It will also need to pay the school tuition $\tau(e)$ which depends on the educational level that the individual is attending. Furthermore, because the individual will not know with certainty the total amount of financial help received it will also face a budget shock $\xi_{ieloft} \sim N(0, \sigma^2_{\xi})$. Finally, individuals will have to make student debt decisions $b_{elofit+1}$ once the budget is realized. Individuals will be then facing two sources of uncertainty regarding their budget constraint at each choice d_{it}. The first source of uncertainty is uncertainty about possible wages, which is common across all choices. Furthermore, if the individual is also enrolled into school it will face another layer of uncertainty coming from the fact that before enrolling into school he does not know with certainty the financial resources available. Notice that if the individual decides to engage in home production there will be no uncertainty in the budget constraint.[12]

Following Arcidiacono et al., (2024) to reflect the fact that individuals will always have a non

[10] In this case $W_{ilot} = \exp(w_{iot} + \varepsilon_{ijt}) \times 52 \times 40$ if the individual is working full time, $W_{ilot} = \exp(w_{ilot} + \varepsilon_{ilot}) \times 52 \times 20$ if working part-time and zero otherwise.

[11] Defined in a similar way as wages, $G_{iet} = \exp(g_{iet} + \varepsilon^e_{iet})$ and , $P_{iet} = \exp(p_{iet} + \varepsilon^p_{iet})$.

[12] Because the different layers of uncertainty increase the computational complexity of the model, some simplifications will be done, as explained in Appendix C

negative value of consumption, realized consumption is assumed to be:

$$C_{ieloft} = max\{C^*_{ieloft}, \underline{C}\} \tag{9}$$

Consumption is evaluated in 2008 dollars and the minimum consumption level is calibrated to 3,842 following Hai and Heckman, (2017).[13]

An important element of the budget constraint is the student debt decision $b_{elofit+1}$. This determines how much debt the agent decides to move for the next period. Notice that this will be either a decision in case the agent is enrolled into school, or it will be determined by a repayment rule $R(b_{it}, X_{it})$ in case the individual is not enrolled.

3.4.2 Loans Repayment

Individuals will pay back the students loans under the 10-year repayment plan.[14] In that sense, loans will evolve period to period as:

$$b_{it+1} = (1+r)b_{it} - R(W_{it}, b_{it})$$

Following Catherine et al., (2024) and Liu, (2020) the yearly payment under this scheme is:

$$R^{10}_{it} = b_{it}\left[\frac{r(1+r)^{P^{10}_{it}}}{(1+r)^{P^{10}_{it}} - 1}\right] \tag{10}$$

where P^{10}_{it} is the amount of remaining repayment periods within the repayment plan, and evolves as $P^{10}_{it+1} = \max\{0, P^{10}_{it} - 1\}$. However, the actual payment individuals make will depend on their available financial resources. For this reason, the actual student debt payment will be:

$$R(W_{it}, b_{it}) = \begin{cases} 0, & \text{if } W_{it} - R^{10}_{it} < \underline{C} \\ W_{it} - \underline{C}, & \text{if } 0 < W_{it} - \underline{C} < R^{10}_{it} \\ R^{10}_{it}, & \text{if } 0 < W_{it} - R^{10}_{it} - \underline{C} \end{cases} \tag{11}$$

The idea of Equation (11) is that individuals will only make student loans payments if their income is above the minimum consumption level of the economy.

3.4.3 Elements of $u_{elof}(Z_{1it})$

A key aspect of the flow-payoff is the elements that are included inside of $u_{elof}(Z_{1it})$, which can be interpreted as controls for the non-monetary drivers of choices. Here I will include different families of parameters:

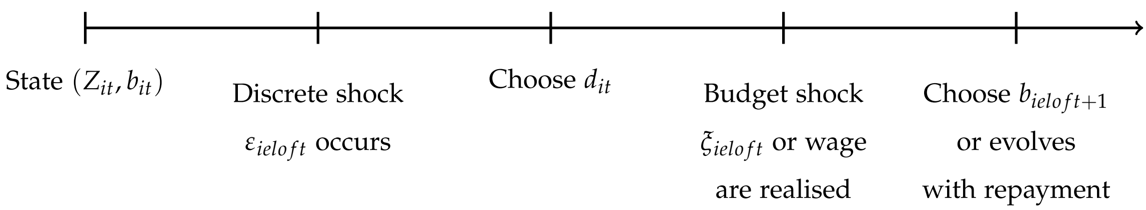
[13] I adjusted to 2008 dollars the value of their work.

[14] This was the main repayment plan during the period of study of the sample.

- Indicator variables for the different possible time-invariant state space characteristics and unobserved type effects. Those will vary at the e, o level.
- Switching costs to capture persistence in choices.
- Indicator variable to capture the fact that it might be the first time an individual is enrolled into a school. This will allow to capture persistence.
- Indicator variables for whether the individual works part-time or full-time during an academic spell, or for working full-time during an occupation spell. This will capture work dislike or preferences, depending on the sign of the coefficient.
- Time effects on preferences for working while enrolled. It is common to see that individuals have a higher taste for consumption as they grow up, and to capture that need to work I will include time-effects on preferences for working.
- Switch in preference effects. As documented in Arcidiacono et al., (2024) learn about their ability as they enroll in different fields. Since I am not explicitly modeling learning about owns abilities or preferences, I will capture this changes with experience effects on the different fields. In that sense, an individual with two years of college might have different preferences that one that is on the first year.

3.5 Timing of the model

The timing of the model is such that the agents make the discrete choice observing the discrete-choice-specific shock ε_{ieloft}, but without observing the continuous-choice-specific shocks included in the wage and budget shock ξ_{ieloft}. Therefore, at each period we have:



The nature of dynamic-discrete-continuous-choice models give rise to selection challenges as described in Bruneel-Zupanc, (2022). This is primarily due to the fact that the discrete and the continuous choices could be endogenous to each other. To solve these problems the timing of the model is crucial. Suppose an scenario in which choices are assumed to occur simultaneously, it becomes clear that the continuous-choice-specific shock will affect the discrete choice. This generates a problem of selection on unobservable variables where the shocks are observed by the agent and yet remain unobservable to the econometrician. In line with the methodology proposed by Kennan and Walker, (2011) or Iskhakov et al., (2017) I will address this problem by assuming a sequential decision process. Agents make the discrete choice once the discrete-choice-specific shock has realized but prior to the realization of the continuous-choice-specific shock. This framework implies that the agents base their discrete choice on anticipations regarding the future continuous debt choice. This also solves an observational problem, since the continuous choice could only be observed at the chosen discrete alternative. While, as noted in Bruneel-Zupanc, (2022), this assumption could lead to biased estimates if choices were made simultaneously, it is justifiable within the context of this model to assume that agents first determine their enrollment in university, based on an expected level of debt, and subsequently make adjustments to this debt level in response to a budget constraint shock. Indeed, it is a desirable assumption to make since this will allow to capture by the model risk-aversion towards possible unwanted debt scenarios.

3.6 The optimization problem

Agents are assumed to be forward looking and to choose the sequence of discrete choices d_{it} together with student debt level decisions $b_{ieloft+1}$ to sequentially maximize the discounted sum of payoffs:

$$\mathbb{E}\left[\sum_{t=1}^{T}\beta^{t-1}\sum_{e}^{E}\sum_{l}^{L}\sum_{o}^{O(e)}\sum_{f}^{F(e)}\left(u_{elof}(Z_{1it})+\frac{C_{ieloft}^{1-\sigma}}{1-\sigma}+\varepsilon_{ieloft}\right)\mathbb{1}\left[d_{it}=(e,l,o,f)\right]\right] \tag{12}$$

where β is the time discount factor. It is important to mention that the expectation is taken with respect to all future idiosyncratic preference shocks, budget shocks, wage realizations and graduation probabilities.

Define $V_t(Z_{it}, b_{it})$ as the ex-ante value function which captures the expected discounted sum of future payoffs just before the idiosyncratic shock is revealed. I can apply the Bellman's optimality principle to express the conditional value function $v_{eloft}(Z_{it}, b_{it})$ as:

$$v_{eloft}(Z_{it},b_{it})=\begin{cases}u_{elof}(Z_{1it})+\mathbb{E}\left[\max_{b_{ieloft+1}}\left\{\frac{C_{ieloft}^{1-\sigma}}{1-\sigma}+\beta V_{t+1}(Z_{it+1},b_{ieloft+1})\right\}\right], & \text{if } e\neq 0\\ \\ u_{elof}(Z_{1it})+\mathbb{E}\left[\frac{C_{ieloft}^{1-\sigma}}{1-\sigma}+\beta V_{t+1}(Z_{it+1},R(b_{ieloft},W_{iolt}))\right], & \text{if } e=0\end{cases} \tag{13}$$

where the expectation denotes uncertainty in financial resources and student debt of alternative (e, l, o, f) but also about future idiosyncratic preference shocks and graduation probabilities. Notice that for choices that do not imply an enrollment decision the conditional value function simplifies since the next period student debt level will be determined by the repayment scheme.

Given that idiosyncratic preference shocks are assumed to be Type 1 GEV distributed, the ex-ante value function can be expressed as:

$$V_t(Z_{it},b_{it})=\ln\left(\sum_{j}\exp\{v_{eloft}(Z_{it},b_{it})\}\right)+\gamma \tag{14}$$

where γ is the Euler's constant. This result is very convenient since it provides a closed form expression of the continuation value and it will be a key element in the solution of the model, because it allows to link the future with the present.

3.6.1 Terminal Continuation Value

The estimation of the parameters of the model requires solving the individual's problem. Given that the model is a finite horizon one, it will be solved backwards starting at the last period. I will assume that after the first 10 academic years upon high-school graduation, individuals will receive a terminal continuation value that captures the present discount value of the utility value of all future wages minus potential debt repayment scenarios. It is then defined as:

$$V_{T+1}(Z_{iT+1}, b_{iT+1}) = \sum_{l=35}^{60} \beta^{l-35} \frac{\left(\frac{w_l(Z_{iT+1})-R(b_{iT+l})}{(1+r)^l}\right)^{(1-\sigma)}}{1-\sigma} \tag{15}$$

where the different wages $w_l(Z_{iT+1})$ will be estimated using ACS data. In a nutshell, I estimate the wage distribution for each field of study and occupation for individuals at each specific age, sex and race. This allows to compute expected future wage trajectories for individuals arriving at period T in any state Z_{iT+1} and with any student debt level.

4 Estimation

In this section I will describe the estimation of the parameters of the model and how the assumptions shape this process. Importantly, as I will now show, the model can be estimated sequentially which implies that the estimation can be performed in stages, as described in Arcidiacono et al., (2024). This means that I can first estimate the parameters governing the wage, parental transfers, grants and graduation probability functions, and taking those estimates as given and with empirical estimates of the conditional choice probabilities (CCPs) I can then estimate the parameters of the flow payoff u_{elof}. Sequential estimation will be crucial to address the computational complexity of the model and all the estimation algorithm will be focused in achieving this property.

Consider the contribution to the likelihood function of an individual with unobserved schooling type k, given the observed sequence of choices d_{it}, wages w_{iot}, grants g_{iet}, parental transfers p_{iet}, graduation outcomes Gr_{iet}, and student debt decisions $b_{ieloft+1}$. For simplicity, I will omit the dependence on observable characteristics. The individual likelihood function conditional on k, $l_i(k)$, is given by:

$$\begin{aligned} l_i(k) &= L(d_{i1}, ..., d_{iT_i}, w_{oi1}, ..., w_{oiT_i}, g_{ei1}, ..., g_{eiT_i}, p_{ei1}, ..., p_{eiT_i}, Gr_{ei1}, ..., Gr_{eiT_i}, b_{i1}, ..., b_{iT_i} | k) \\ &= L_{dit}(k) L_{woit}(k) L_{geit}(k) L_{peit}(k) L_{Geit} L_{bit+1}(k) \end{aligned} \tag{16}$$

This implies that the likelihood for an individual can be expressed as the product of the likelihoods associated with each observed component: the likelihood of choices, L_{dit}, wages, L_{woit}, grants, L_{geit}, parental transfers, L_{peit}, graduation outcomes, L_{Greit}, and student debt decisions, L_{bit+1}.[15] However, the individual's schooling type k is unobserved by the econometrician, which necessitates its integration

[15]This follows from the law of successive conditioning under the assumption that error terms are serially uncorrelated and independent over time, as well as the fact that the state evolution depends only on the current state and choices, thereby eliminating dependence on previous states.

out using a finite-mixture model, as developed by Heckman and Singer, (1984). Thus, the unconditional individual log-likelihood, $\ln l_i$, is given by:

$$lnl_i = \ln\left(\sum_{k=1}^{K=2} \pi_k L_{dit}(k) L_{woit}(k) L_{geit}(k) L_{peit}(k) L_{Greit} L_{bit+1}(k)\right) \tag{17}$$

There are several important considerations to address in estimating the model's parameters. First, to account for the finite-mixture likelihood I will use the Expectation-Maximization (EM) algorithm. Additionally, the likelihood function is not additive separable across its components, which adds significant complexity to the estimation process. In this section, I will briefly outline how these challenges are addressed, with a more detailed explanation provided in Appendix B.

4.1 Estimation Algorithm Summary

I will first explain how estimation works without using the nested pseudo-likelihood (NPL) algorithm proposed by Aguirregabiria and Mira, (2002), and later I will explain how the algorithm works. As described in detail in Appendix B, the implementation of the EM-algorithm estimates the model parameters by iterating between the expectation (E-step) and the maximization (M-step) . In the E-step, the type probabilities distributions π_r and the posterior type probabilities q_{ir} are updated. Using those estimates, the model parameters (θ) are updated at the M-step maximization problem and this procedure is repeated until convergence. The M-step objective function is the following:

$$\hat{\theta} = \arg\max \sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{t=1}^{T} \hat{q}_{ik} \ln\left[L_{dit}(k) L_{woit}(k) L_{geit}(k) L_{peit}(k) L_{Greit} L_{bit+1}(k)\right] \tag{18}$$

The EM-algorthm provides a way of estimating the model parameters, but because of the computational complexity of the model it becomes unfeasible in this case. For this reason, I will build on the algorithm proposed by Arcidiacono et al., (2024) in which an auxiliary model with measures is used in estimation to get an estimate of the posterior distribution of unobserved types $\hat{q}_{ik}$ as well as the type distribution $\hat{\pi}_r$. Once this is obtained, the model parameters can be estimated using only the M-step of the algorithm, without the need of re-updating the types distribution. Importantly, as described by Arcidiacono and Jones, (2003), in the M-step additive separability of the likelihood is re-introduced, allowing again for a sequential estimation. This can be seen in Equation (18) since it is additive on logs. The estimation procedure is as follows:

1. Estimate an auxiliary model with measures. This model is a simplified version of the full structural model that will allow to obtain the posterior distributions for the unobserved type and the estimates for the CCPs. The auxiliary model accommodates choices, wages, grants and parental transfers together with measures that will help in the identification and interpretation of the unobserved types as noted in Carneiro et al., (2003).

2. Using the posterior probabilities as weights on the M-step, additive separability is reintroduced

allowing for a sequential estimation of the model parameters. I will first estimate the parameters of the wages, grants, transfers and graduation probabilities and together with the CCPs that I have previously estimated this will allow to construct the necessary object to estimate the parameters of the flow-payoff function.

4.1.1 Estimation of the production functions

Because the additive separability properties are re-introduced in the M-step, estimation of the wages, grants, parental transfers and graduation probabilities is trivial.

Estimation of wages, grants and transfers. Once additive separability has been reintroduced the estimation of the wage, grants and parental transfer parameters can be done by a normal maximum likelihood, and in this case it is consistently estimated by OLS.

Estimation of the grants, parental transfers and graduation probabilities. Under the logistic assumption of those distributions, the parameters will be consistently estimated under the maximum likelihood estimation of a logistic distribution.

4.1.2 Estimation of the budget shock distribution

Before proceeding with the estimation of the flow-payoff parameters, it is necessary to first estimate the distribution of the budget shock, which will allow for the computation of expected consumption for individuals based on their respective choices. Two primary challenges complicate the identification of the budget shock. First, there is no closed-form expression for the student debt decision, which is the variable used to identify the budget shock distribution. Second, the empirical distribution of student debt is censored at zero, as individuals with larger budgets do not incur student debt, complicating the identification of the underlying budget shock. To address these challenges, the estimation will be conducted using the simulated method of moments. In this approach, I will target two key moments: the average student debt among indebted individuals and the proportion of non-indebted individuals, to ensure that the budget shock aligns with empirical student debt decisions. Since the continuation values can be constructed using conditional choice probabilities (CCPs), the estimates of the budget shock distribution will remain consistent even before the flow-payoff parameters are estimated. However, to improve the precision I will perform Aguirregabiria and Mira, (2002).

4.1.3 Estimation of the Flow-payoff parameters

The estimation of the flow-payoff parameters will be done taking as given the estimates from the previous stages. Given the Type 1 GEV assumption of the idiosyncratic preference shock, the likelihood of the choices becomes a multinomial logit:

$$\sum_{i=1}^{N}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{e}^{E}\sum_{l}^{L}\sum_{o}^{O(e)}\sum_{f}^{F(e)}\hat{q}_{ik}d_{ioelft}\ln\left(\frac{\exp(v_{eloft}(Z_{it},b_{it})-v_{ht}(Z_{it},b_{it}))}{1+\exp(v_{eloft}(Z_{it},b_{it})-v_{ht}(Z_{it},b_{it}))}\right) \tag{19}$$

And since identification is only obtained up to a reference category, I will set home production as the reference alternative normalizing $u_h = 0$, with slightly abuse of notation. The main estimation challenge is that the conditional value function depends in part on the continuation value, which is a computationally costly object to construct.[16]

To address the computational problem I will estimate the model using the empirical conditional choice probabilities (CCPs) to avoid having to solve for the continuation value many times. As opposed to many studies, the benefits of the usage of CCPs in this study are lower than expected because of the use of the Nested Pseudolikelihood Algorithm proposed by Aguirregabiria and Mira, (2002). In a nuthsell, to improve the precision of my estimates, I will re-update the CCPs using model estimates until convergence, which means that I still need to solve the model. For this reason, the benefits from CCPs in this model comes from the fact that they will act as a control function, allowing to control for the differences in continuation values:

$$v_{eloft}(Z_{it},b_t)-v_{ht}(Z_{it},b_t)=\underbrace{u_{elof}(Z_{1it})}_{\substack{\text{Flow-payoff function}\\ \text{to be estimated}}}+\underbrace{\delta(Z_{it},b_{it},d_{it})}_{\substack{\text{Function controlling}\\ \text{differences in continuation}\\ \text{values}}} \tag{20}$$

where $\delta(Z_{it},b_t,d_{it})$ is a function that captures differences in the sequence of conditional choice probabilities and terminal continuation values obtained by alternative (e,l,o,f) and the base category h.[17] The key element of Equation (20) is that the parameters to be estimated enter linearly in the equation only through $u_{elof}(Z_{1it})$, and once $\delta(Z_{it},b_{it},d_{it})$ has been computed, it can be included in the estimation as a nuisance term that acts as a control function for the difference in future payoffs across the two alternatives. To improve the precision of my estimation once I obtain the parameters of the flow-payoff I re-update the CCPs with the model estimates and I repeat this process until convergence, with corresponds to the NPL algorithm presented by Aguirregabiria and Mira, (2002).[18]

5 Results

In this section I present the estimated parameters and discuss the model fit. To ensure identification some parameters are calibrated, as reported in Table 4.

[16] The construction of the continuation value requires solving the conditional value functions for each possible state space backwards since the last period of the model. In Appendix D a discussion can be found about the model state space.

[17] See Appendix B.4 for the derivation of this term

[18] See Appendix B for a detail explanation of the functioning of the NPL algorithm and the modifications done in this work to accommodate the current estimation.

Table 4: Calibrated Parameters

Parameter	Value	Source
σ	0.4	Arcidiacono et al., (2024)
r	0.05	Department of Education
β	0.98	Literature
Tuition Two-Year	\$6,490	Average NLSY97
Tuition Four-Year	\$17,089	Average NLSY97
Tuition Grad School	\$23,889	Average NLSY97

Notes: The tuition for each educational alternative is obtained by computing the average educational expenditure at each education institution by individuals of the sample. Monetary amounts are expressed in 2008 U.S. dollars.

5.1 Estimated Parameters

5.1.1 Wage Parameters

Table A.6 reports the estimates of the wage equations for the time invariant individual characteristics and the level of labor market experience. Overall, there is a positive gradient in wage returns in parental income and ability, which means that initial characteristics are important in explaining hourly wages. Indeed, for some occupations like Social Sciences, being from the highest parental income quartile implies a return equivalent to almost 10 years of labor market experience more than being from the bottom quartile of the parental income distribution. The value of experience is also quite different across occupations, with STEM being the occupation at which experience is more valuable. As well documented in the literature, females earn less than their male counterparts in all occupations, and only on social sciences the effect seems unclear. In some of the cases, like in education professions, females earn the equivalent of three years of experience less than their male counterparts. Similarly, black individuals tend to earn less than non-black ones in most of the occupations except Humanities. Also, the effect seems to be not negative in Education or STEM occupations. The highest penalty is obtained in Social Sciences. The effect is greater in Social Sciences, in which a black individual makes the equivalent of four years of experience less than a non-black individual.

In Table A.7 I report the major-occupation complementarities. This table reveals interesting patterns. STEM seems to be the most valuable major, with one of the highest effects in almost every occupation. Actually, for Business, STEM and Production occupations, a STEM graduate receives the highest compensation compared with other fields of study. On the other hand, the fields that have less value are Education or Humanities. If we take a look at the natural occupation of each field (that is, the one that corresponds with the field), Humanities has the lowest return with a gain just equivalent to 1.5 years of experience. Actually, the highest value of a humanities graduate is at a Business occupation. The most valuable combination is a Health major at a Health occupation, probably dominated by doctors. Business has the highest return in Sales and Office occupations, although the starting wage was higher at a Business occupation. Education has the highest return at an Education occupation. For Associate Degrees, the return is very high at a Health occupation, and negative but insignificant for Humanities. For a master's degree, the highest return is for a Health field, which complemented with the return if the individual has a Health degree, makes Health occupations the ones with a higher possible gain from education.

However, the value from a field is not only measured on the monetary effect on occupations, but also on which occupations allows to access.[19] In Table D.4 we can see that having specific majors open the door to particular occupations. Notice that occupations like Sales and Office or Production are dominated by individuals without any degree, which implies that there less educational barriers to entry. On the other hand, Occupations like STEM, Education or Social Sciences only have around 35% of individuals without a degree, which implies that the value of a major here is not only the monetary return, but also the access to those fields.

5.1.2 Financial Parameters

The results from the financial functions are reported in Tables A.3 and A.4 and match with the dynamics of the U.S. financial aid system. Regarding parental financial support, individuals with more parental income or with a higher measure of ability will receive transfers more often and of higher amount. Also, the amount of the transfer will be higher when attending a four-year institution.

For grants, individuals with a higher parental income are less likely to receive grants and to receive less in the event of reviving some, specially at four-year institutions. On the other hand, ability seems to not have an effect on grants for two-year schools, but a very positive gradient on four-year schools, consistent with merit based grants. Finally, both females and black individuals are more likely to receive grants at two and four-year schools, capturing the dynamics of affirmative action. However, when receiving grants, the effect on females is almost negligible and black individuals are positive discriminated in total amounts.

5.1.3 Graduation Probabilities

The graduation probabilities are reported in Table A.5. Individuals with more parental income are more likely to graduate. The highest effect is the one by the highest quartile of the ability distribution, which shows that individuals are much more likely to graduate than their counterparts, conditional on the same years of education. The effect is not the highest at graduate school, but since I am not considering experience there, it could indicate that they take longer master degrees.

[19] In the current version of this paper there are no frictions in the labor market, so all occupations can be accessed. However, for future versions I would like to measure heterogeneous arrival rates of offers from different occupations given your field of study.

5.1.4 Flow Payoff Parameters

The coefficients of the flow payoff utility are separated between the utility of education alternatives, the utility of occupation alternatives and the education-occupation complementarities. Table A.8 reports the coefficients for the educational alternatives. Overall, the dynamics are very similar with the ones presented in the descriptive evidence section. As expected, individuals dislike working while enrolled. The effect is almost three times as negative for individuals working full-time than part-time. The dislike parameter is lower when enrolled in an associate-degree and also lower when enrolled in graduate school. In terms of switching cost, all the fields have negative costs associated from switching in from a different field. Education is the field with the higher switching cost, but this could be capturing that individuals don't like to switch into education, followed by health and STEM, which are known to be harder to study (see Ahn et al., 2019 or Arcidiacono et al., 2024).

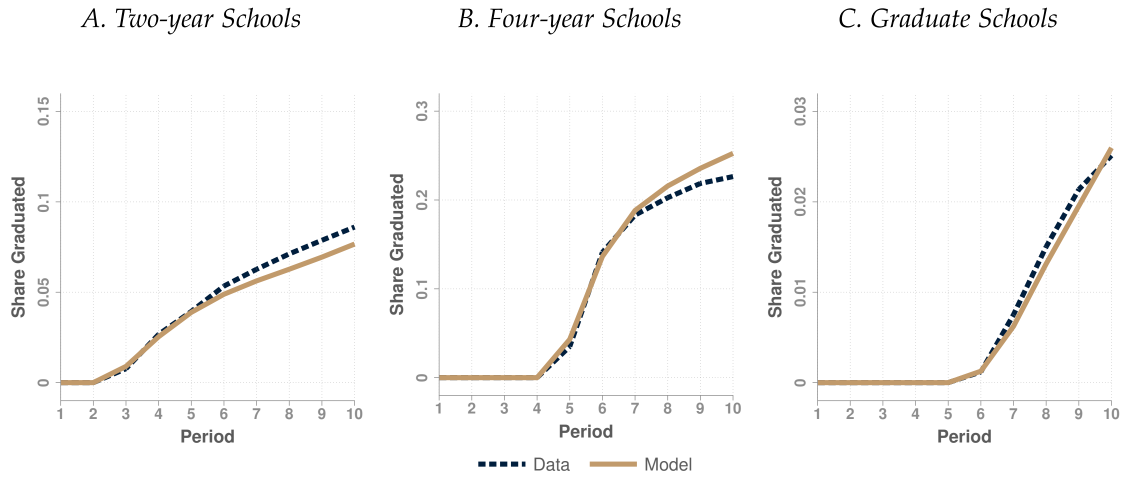
In terms of occupation alternatives, the story is similar. Table A.9 reports estimated coefficients. As with fields of study, there is an associated switching cost across occupations. The occupations harder to switch in are STEM and Social Sciences, but also Humanities probably by the lack of jobs. The occupations easier to switch in are Production and Sales and Office. Interestingly, the parameter of working full-time is this time positive. In this case, the base alternative would be working part-time, and the positive sign seems to capture than individuals prefer to work full-time than part-time, although they dislike both options reflected in a very negative constant. This suggest that the work-dislike effects are non-linear, and that once having to work, individuals tend to prefer working full-time.

Finally, Table A.10 reports the coefficients of the education-occupation complementarities. All the coefficients have a positive sign, which indicates that at any occupation individuals will have higher utilities if they have some education relative to non-educated individuals. This result reflect non-monetary returns to college, which are indeed important to understand education decisions. For each occupation, individuals have higher utility if they actually have a degree that is closely related to that specific occupation. For example, within a Business occupation the higher utility comes from Business graduates. Although this pattern is true for almost all occupations, there are some exceptions. For example, for Social Sciences occupations, the individuals with the higher non-consumption taste are the ones with an "Other" graduate degree. This could reflect that the return for that degree in that occupation is very little, and therefore only non-consumption preferences could explain this choice.

5.2 Model Fit

To evaluate the fit of the model I simulate choices by individuals of the sample under the estimated parameters. To do so, I simulate 30 times each individual of the sample and compute its choices along 10 periods based on the estimated structural parameters and estimated distributions. To obtain future terms I first solve the model backwards starting at the terminal period for each possible future state, and resolve backwards.

Figure 1: **Distribution of graduates over time**



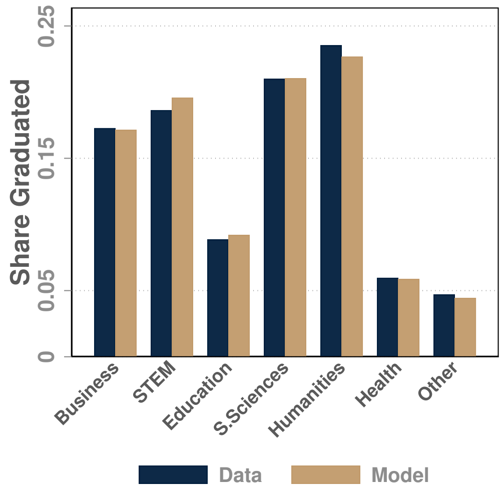
Notes: This figure reports the share of individuals graduated at each educational level for each academic year after high-school. The model data is computed by simulating 30 times each individual of the sample, and simulating their educational choices over 10 academic years. Graduation shares are obtained by aggregating the total amount of individuals graduated across different fields of study.

Table A.11 reports the model fit in terms of the main choices pooled by periods and by fields or occupations. As it can be seen the model does an excellent job in matching the share of individuals that make each choice, which is not surprising given that the whole estimation algorithm is built to predict those choices. A more complicated job is to generate the model dynamics. Figure 1 reports the dynamics of graduations across periods and institution types, both in the data and in the model. It can be seen that the model does a very good job, although in the last periods it tens to slightly over-predict graduation at four-year schools and under-predict graduation at two-year schools. In terms of graduation at graduate-schools the fit seems perfect. Figure 2 reports the field distribution of the economy during the last period, and it can be seen that the model does an excellent job not only at explaining four-year graduation, but also at matching the distribution of fields of the economy.

Finally, Table 5 reports the fit of student debt across the AFQT and parental income distributions. As we can see, the model matches quite well the shares and the gradient of indebted individuals. For the average student debt, the model does a good job for all individuals except for high-income ones, in which the average debt at graduation is lower than it should.[20]

[20] I am currently working on a refinement of the model that will make sure all the moments match. This will be done by introducing unobserved heterogeneity in the financial resources across parental income.

Figure 2: **Graduation distribution by Field**



Notes: This figure reports the overall distribution of fields of graduation 10 years after high-school. The figure is computed by simulating 30 times each individual of the sample, and simulating their educational choices over 10 academic years. The shares are with respect to the share of individuals that obtain a bachelor degree.

Table 5: Student Debt Fit by Ability and Parental Income

	AFQT		Parental Income	
	Data	Model	Data	Model
Q1 *Share*	0.71	0.65	0.78	0.70
Avg	26,502	20,190	20,774	22,163
Q2 *Share*	0.71	0.66	0.76	0.71
Avg	25,395	20,317	24,992	23,060
Q3 *Share*	0.69	0.66	0.74	0.72
Avg	22,665	22,625	24,255	20,741
Q4 *Share*	0.58	0.53	0.49	0.43
Avg	23,781	15,499	23,853	12,182

Notes: This table is created by simulating 30 times each individual of the sample, and simulating their educational choices over 10 academic years. It reports the share of indebted individuals and the average debt at the moment of graduating at a four-year school. Monetary amounts represent 2008 dollars.

6 Saving on Valuable Education (SAVE)

In 2023 the U.S. government introduced what is considered to be the most affordable repayment plan in history, known as Saving on Valuable Education (SAVE) repayment plan. Although it is not the first income-driven plan, and others have already been used, it is the most generous at promoting an insurance against labor market uncertainty at the moment of making student loans decisions.

The plan has mainly three changes compared with the standard 10-year repayment plan, as described in Catherine et al., (2024): (1) the introduction of an income exception at 225% of the poverty line, allowing for zero monthly payments [21]; (2) payments are capped at 5% of an individual discretionary income (or 10% for graduate borrowers)[22]; (3) there is a forgiveness on outstanding balances after 10 years of repayment for balances below $12,000, and for every additional $1,000 the forgiveness is delayed one extra year, until arriving at 20 years (or 25 years for graduate borrowers).

To evaluate the effects on enrollment and field of study of the SAVE repayment plan, I will now simulate a counterfactual economy in which this is the repayment system for loans.

6.1 The Enrollment and Graduation Effect

The introduction of the SAVE repayment plan has significant effects on different dimensions as reported on Table 6. Focusing on enrollment, Table A.12 reports the effect on the probability of choosing enrollment in a four-year school across different dimensions, both for individuals with low or high latent schooling taste. On average, the introduction of the policy increases the enrollment probabilities by about 30% for low-income individuals and 5% for high income ones. The effect is stronger for individuals with low skills, since at the bottom of the AFQT distribution the increase on enrollment implies an increase of almost 40%. To evaluate the value of labor market insurance, I compute what would be the equivalent grant that implies the same increase in enrollment probabilities. The results reported in Table A.12 suggest that just to achieve the same enrollment effects in period 1 (without considering other periods), low-income individuals would have to receive on average a grant of $25,000. The grant equivalence would just be of about $9,000 if the individual is from the top of the parental income distribution. Overall, those who feel more attracted by the policy are individuals with higher drop-out chances or with higher financial risk possibilities during their labor market trajectory. This is consistent with the findings of Hendricks and Leukhina, (2017) who establishes that low skill individuals are aware of their high drop-out possibilities. On the other hand, high skilled individuals know their likely graduation outcomes, and for this reason they are less concerned about uninsured labor market risk.

In terms of graduation the effects are also positive but slightly lower. The results are reported in

[21] The definition of poverty line depends on family structure, which is not modeled in the paper. For this reason, I will assume the most conservative family structure which treats individuals as single families.

[22] Discretionary income is the income above 225% of federal poverty line

Table 6: Summary of Effects

Outcome	Baseline	SAVE	Change	
Graduation				
Two-Year	0.08	0.09	+1.00ppt	(**+11.81 %**)
Four-Year	0.25	0.27	+2.27ppt	(**+9.10%**)
Graduate School	0.03	0.03	+0.47ppt	(**+18.84%**)
Four-Year Schools				
Enrollment Share	0.43	0.46	+ 3.32ppt	(**+7.71%**)
Drop Out Share	0.46	0.45	-0.96ppt	(**-2.11%**)
Average Student Loans	18,358	22,925	+ 4,567	(**+24,87%**)
Share Indebted	0.58	0.86	+27.80ppt	(**+48.28%**)
Share Not Working Enrolled	0.34	0.37	+2.27ppt	(**+8.19%**)

Notes: This table reports the main results of the counterfactual scenario in comparison with the baseline scenario, with a special focus on four-year schools. Graduation and enrollment shares are obtained with respect to the overall population. Drop out shares are computed only with respect to enrolled individuals. The share and average of student loans is obtained at the moment of graduation at a four-year school.

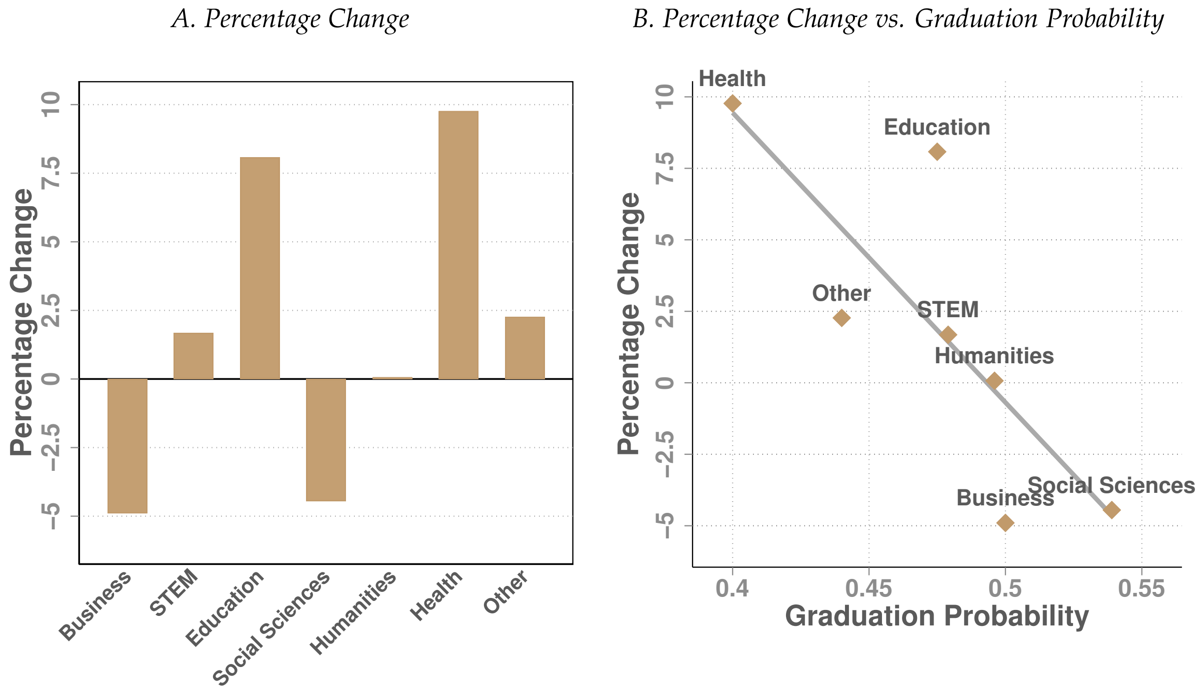
Table A.13 which show that the policy increased graduation for low-income individuals by 18% in comparison to just a 5% increase for high income ones. The same Table also reports the decomposition of the new graduates based on their origin in the baseline economy. As it can be seen, about 70% of the new graduates were previously dropping out from school. And only between 7 to 11% are actually new enrolled individuals. There is also a significant fraction of individuals (between 14 to 22%) that where previously graduating from a two-year school and now have decided to enroll into a four-year one. This fact is consistent with the one being reported on Table A.14 that shows that only about 20% of new enrolled individuals achieve graduation, and those who do it tend to have a higher unobserved schooling preference taste. Similar, the results are in the line of Belzil et al., (2023) who found that a tuition reduction would attract individuals with lower graduation probabilities and much higher drop out rates.

6.2 The Labor Supply and Field of Study Effect

The change in repayment plan not only changes the enrollment and graduation, but also the trajectories of individuals that were already enrolled. In the light of the new insurance provided by the new repayment plan, individuals are now willing to change fields of study and also to reduce their labor supply participation decision while enrolled.

Table A.16 reports the change in the labor supply decision of individuals. As we can see, they now prefer to reduce their labor supply while enrolled by about 8%. The effect is stronger for indebted

Figure 3: **Change in the distribution of Fields**



Notes: This figure reports the change in the distribution of fields of the economy. Panel (A) reports the percentage change in the share that each field represents over the total amount of graduates. Panel (B) represents the relationship between the percentage change and the probability of graduating after four years at each of the different fields. The fitted regression has an estimated slope of -0.006 (0.002).

individuals, which have an increase of 16% in the share that does not work. Those individuals will very likely finance education with debt instead of with work, and for this reason once they arrive to graduate school they will be more likely to work while enrolled. Consistent with this story Table A.17 reports the change in student debt patterns. As we can see, there is an increase of about 48% in the share of indebted individuals, and the average debt at graduation increases by about 24%, consistent with the fact that debt is now more attractive. The results are in the line with the ones in Catherine et al., (2024) which suggest that the introduction of SAVE promotes both more students to borrow, and also borrow larger amounts. The results are also consistent with those found in Black et al., (2023) in which after increasing loans limits individuals improve graduation and reduce labor supply while enrolled.

In terms of fields of study individuals are now willing to change their field. In particular this is the case for indebted individuals, as reported in Table A.18. Notice that low-income individuals are almost 50% more likely to switch fields than high-income ones, and also indebted individuals are 24% more likely to switch fields than non-indebted ones.

Figure 3 reports the percentage change in the final distribution of fields. As we can see, Business and Social Sciences represents a lower share of graduates after SAVE was introduced. On the other hand, Education and Humanities are the majors with a larger relative gain. Panel (B) reports the correlation between the change and the probability of graduation at each field, and shows that individuals are moving from easy to more difficult fields. Furthermore, expected earnings also matter, and individuals are sorting into fields with lower earnings, such as Education. To put some light on the changes of the distribution Table A.23 decomposes the changes comparing individuals from the top and the bottom of the parental income distribution. Overall, all fields gain individuals, but only Education and Health have positive effects among switchers. As we can see, the patterns are similar along both distributions, but are intensified in individuals form the bottom of the distribution since they are more sensitive to the policy. Overall, all the fields of the economy increase, but with special focus to Education and Health. This is consistent with the results reported in Table A.24 where I decompose final distribution of fields of the economy across individual characteristics. The share of individuals that graduates from Education increases by 8% (15% for low-income individuals) and the share of individuals that graduate from Health increase by 10% (22% for low-income individuals).

Finally, Table A.26 reports the average of individuals that have an increase in non-consumption preferences (that is $u_{ieloft} + \epsilon_{ieloft}$) after switching fields, conditional on either the field that they are leaving or the field that they arrive. What we see is that on average, individuals that leave a field to work less with have a higher increase than those that leave a field to work more. Furthermore, individuals that leave STEM or Business tend to like the new field more than those who leave Education or Health. This is consistent with the fact that individuals working in STEM or Business are there for financial reasons more than non-consumption preferences.

7 Conclusion

This paper examines how student debt repayment affects higher education enrollment and field of study choices, focusing on the introduction of the SAVE repayment plan—an income-driven repayment option that provides substantial insurance against labor market uncertainties. Specifically, I document the effects of SAVE on enrollment, graduation rates, and field-of-study choices, highlighting its differential impact on low-income individuals. To do so, I develop and estimate a dynamic human capital investment model with rich career possibilities and student debt decisions. The estimated model allows to compare human capital investment choices of a scenario with a flat repayment plan with one in which SAVE is the repayment option.

I highlight four main findings. First, the introduction of the SAVE repayment plan has an effect on enrollment that is coming from the insurance value it provides for students. Individuals now are more willing to enroll in school, less likely to drop out, and less likely to work during their education. The effects vary substantially by socioeconomic status. For low-income individuals there is an 18% increase in graduation rates, compared with 5% increase for high-income ones.

Second, despite attracting new students into college, only about 20% of the new enrolled individuals graduate. This is because the marginal students that enroll after the policy have high chances of dropping out. On the other hand, the availability of a more comfortable repayment reduces drop out rates among previously enrolled individuals that were dropping because of financial reasons.

Third, there is an increase in the use of student loans. The share of indebted individuals rises by 48% and the average loans at graduation by 25%. This is a consequence of students reducing their labor supply while studying, and increasing their consumption while enrolled, given the higher insurance against uncertain future income.

Finally, there is a change in the field composition of the economy, mainly coming from low-income individuals whose choices where more influenced by financial reasons. The level of insurance provided by SAVE switches individuals towards majors in Education, with lower earnings upon graduation, or Health, that require more years to achieve graduation. At the same time, individuals leave majors in Business, considered the save option in financial terms, with easier graduation prospects and high earnings upon graduation.

Overall, this paper highlights a few important considerations. On the one hand, it seems clear that individuals are concerned about uninsured labor market risk, and providing some level of insurance will improve their willingness to make educational investments. Since this is particularly true for low-income individuals, the policy helps in reducing higher-education inequalities. The policy is specially helpful in reducing drop out rates among financially constrained students. However, I also document that many of the students that now are willing to enroll into school will never achieve graduation. For this reason, policies trying to reduce higher education access for low-income individuals should also be accompanied by policies targeted at reducing their non-financial difficulties while enrolled. Finally, because the repayment plan changes the field composition of the economy it is then clear that policy makers can have an influence in the field individuals decide to study just by affecting the way higher education is financed, which opens the door to future policies to target enrollment at specific fields.

References

Abbott, B., Gallipoli, G., Meghir, C., & Violante, G. L. (2019). Education policy and intergenerational transfers in equilibrium. *Journal of Political Economy*, *127*(6), 2569–2624.

Adda, J., & Cooper, R. W. (2003). *Dynamic economics: Quantitative methods and applications* (1st ed., Vol. 1). The MIT Press.

Aguirregabiria, V., & Mira, P. (2002). Swapping the nested fixed point algorithm: A class of estimators for discrete markov decision models. *Econometrica*, *70*(4), 1519–1543.

Ahn, T., Arcidiacono, P., Hopson, A., & Thomas, J. R. (2019, December). *Equilibrium Grade Inflation with Implications for Female Interest in STEM Majors* (NBER Working Papers No. 26556). National Bureau of Economic Research, Inc.

Altonji, J., Arcidiacono, P., & Maurel, A. (2016). Chapter 7 - the analysis of field choice in college and graduate school: Determinants and wage effects. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.). Elsevier.

Altonji, J. G. (1993). The demand for and return to education when education outcomes are uncertain. *Journal of Labor Economics*, *11*(1, Part 1), 48–83.

Altonji, J. G., Blom, E., & Meghir, C. (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annual Review of Economics*, *4*(Volume 4, 2012), 185–223.

Angrist, J., Autor, D., & Pallais, A. (2021). Marginal Effects of Merit Aid for Low-Income Students*. *The Quarterly Journal of Economics*, *137*(2), 1039–1090.

Arcidiacono, P. (2005). Affirmative action in higher education: How do admission and financial aid rules affect future earnings? *Econometrica*, *73*(5), 1477–1524.

Arcidiacono, P., Aucejo, E., Maurel, A., & Ransom, T. (2024). College attrition and the dynamics of information revelation. *Journal of Political Economy*, Forthcoming.

Arcidiacono, P., & Ellickson, P. B. (2011). Practical methods for estimation of dynamic discrete choice models. *Annual Review of Economics*, *3*(1), 363–394.

Arcidiacono, P., & Jones, J. B. (2003). Finite mixture distributions, sequential likelihood and the em algorithm. *Econometrica*, *71*(3), 933–946.

Arcidiacono, P., & Miller, R. A. (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica*, *79*(6), 1823–1867.

Avery, C., & Hoxby, C. M. (2004, July). Do and Should Financial Aid Packages Affect Students' College Choices? In *College Choices: The Economics of Where to Go, When to Go, and How to Pay For It* (pp. 239–302). National Bureau of Economic Research, Inc.

Beffy, M., Fougère, D., & Maurel, A. (2012). Choosing the Field of Study in Postsecondary Education: Do Expected Earnings Matter? *The Review of Economics and Statistics*, *94*(1), 334–347.

Belley, P., & Lochner, L. (2007). The Changing Role of Family Income and Ability in Determining Educational Achievement. *Journal of Human Capital*, *1*(1), 37–89.

Belzil, C., Hansen, J., & Pernaudet, J. (2023). Separating Structural and Composition Effects of Higher Education Financial Aid on Field of Study.

Beyer, H., Hastings, J., Neilson, C., & Zimmerman, S. (2015). Connecting student loans to labor market outcomes: Policy lessons from chile. *American Economic Review*, *105*(5), 508–13.

Biswas, S. (2020, November). *Federal student loans, college choice, and student welfare* [Unpublished manuscript].

Black, S. E., Denning, J. T., Dettling, L. J., Goodman, S., & Turner, L. J. (2023). Taking it to the limit: Effects of increased student loan availability on attainment, earnings, and financial well-being. *American Economic Review*, *113*(12), 3357–3400.

Bruneel-Zupanc, C. A. (2022). Discrete-Continuous Dynamic Choice Models: Identification and Conditional Choice Probability Estimation.

Burland, E., Dynarski, S., Michelmore, K., Owen, S., & Raghuraman, S. (2023). The power of certainty: Experimental evidence on the effective design of free tuition programs. *American Economic Review: Insights*, *5*(3), 293–310.

Cameron, S. V., & Heckman, J. J. (1998). Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of american males. *Journal of Political Economy*, *106*(2), 262–333.

Carneiro, P., Hansen, K. T., & Heckman, J. J. (2003). 2001 lawrence r. klein lecture estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review*, *44*(2), 361–422.

Carneiro, P., & Heckman, J. J. (2002). The evidence on credit constraints in post-secondary schooling*. *The Economic Journal*, *112*(482), 705–734.

Castleman, B. L., & Long, B. T. (2016). Looking beyond enrollment: The causal effect of need-based grants on college access, persistence, and graduation. *Journal of Labor Economics*, *34*(4), 1023–1073.

Catherine, S., Clanton, M. P., & Yannelis, C. (2024). The distributional effects of student loan forgiveness: An update on save and the covid-19 moratorium. *National Tax Journal*, *77*(3), 655–680.

Caucutt, E. M., & Lochner, L. (2020). Early and late human capital investments, borrowing constraints, and the family. *Journal of Political Economy*, *128*(3), 1065–1147.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 1–38.

Dynarski, S., Page, L., & Scott-Clayton, J. (2023). Chapter 4 - college costs, financial aid, and student decisions. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.). Elsevier.

Dynarski, S. M. (2003). Does aid matter? measuring the effect of student aid on college attendance and completion. *American Economic Review*, *93*(1), 279–288.

Field, E. (2009). Educational debt burden and career choice: Evidence from a financial aid experiment at nyu law school. *American Economic Journal: Applied Economics*, *1*(1), 1–21.

Bureau of Labor Statistics, U.S. Department of Labor. (1997-2021). National longitudinal survey of youth 1997 cohort, 1997-2021 (rounds 1-20) [Produced and distributed by the Center for Human Resource Research (CHRR), The Ohio State University. Columbus, OH: 2024].

Hai, R., & Heckman, J. J. (2017). Inequality in human capital and endogenous credit constraints [Special Issue on Human Capital and Inequality]. *Review of Economic Dynamics*, *25*, 4–36.

Heckman, J., & Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, *52*(2), 271–320.

Hendricks, L., & Leukhina, O. (2017). How risky is college investment? *Review of Economic Dynamics*, *26*, 140–163.

Hotz, V. J., & Miller, R. A. (1993). Conditional Choice Probabilities and the Estimation of Dynamic Models. *The Review of Economic Studies*, *60*(3), 497–529.

Iskhakov, F., Jørgensen, T. H., Rust, J., & Schjerning, B. (2017). The endogenous grid method for discrete-continuous dynamic choice models with (or without) taste shocks. *Quantitative Economics*, *8*(2), 317–365.

Johnson, M. T. (2013). Borrowing constraints, college enrollment, and delayed entry. *Journal of Labor Economics*, *31*(4), 669–725.

Kane, T. J. (1996). College cost, borrowing constraints and the timing of college entry. *Eastern Economic Journal*, *22*(2), 181–194.

Keane, M., & Wolpin, K. I. (2001). The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review*, *42*(4), 1051–1103.

Keane, M. P. (2002). Financial aid, borrowing constraints, and college attendance: Evidence from structural estimates. *American Economic Review*, *92*(2), 293–297.

Keane, M. P., & Wolpin, K. I. (1997). The career decisions of young men. *Journal of Political Economy*, *105*(3), 473–522.

Kennan, J., & Walker, J. R. (2011). The effect of expected income on individual migration decisions. *Econometrica*, *79*(1), 211–251.

Liu, Q. (2020). College enrollment, parental transfers, and student loans. In *Essays on student loans and returns to skill* (pp. 36–77).

Lochner, L. J., & Monge-Naranjo, A. (2015). Default and repayment among baccalaureate degree earners. In B. Hershbein & K. M. Hollenbeck (Eds.), *Student loans and the dynamics of debt* (pp. 235–286). W.E. Upjohn Institute for Employment Research.

Lochner & Monge-Naranjo, A. (2011). The nature of credit constraints and human capital. *American Economic Review*, *101*(6), 2487–2529.

Looney, A., & Yannelis, C. (2024). What went wrong with federal student loans? *Journal of Economic Perspectives*, *38*(3), 209–36.

Lovenheim, M., & Smith, J. (2023). Chapter 4 - returns to different postsecondary investments: Institution type, academic programs, and credentials. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.). Elsevier.

Lovenheim, M. F., & Reynolds, C. L. (2011). Changes in postsecondary choices by ability and income: Evidence from the national longitudinal surveys of youth. *Journal of Human Capital*, *5*(1), 70–109.

Luo, M., & Mongey, S. (2024, March). *Assets and job choice: Student debt, wages and amenities* [Unpublished manuscript].

Mueller, H., & Yannelis, C. (2022). Increasing enrollment in income-driven student loan repayment plans: Evidence from the navient field experiment. *The Journal of Finance*, *77*(1), 367–402.

Patnaik, A., Wiswall, M., & Zafar, B. (2021). College majors 1. *The Routledge handbook of the economics of education*, 415–457.

Rothstein, J., & Rouse, C. E. (2011). Constrained after college: Student loans and early-career occupational choices. *Journal of Public Economics*, *95*(1-2), 149–163.

Speer, J. D. (2017). The gender gap in college major: Revisiting the role of pre-college factors. *Labour Economics*, *44*, 69–88.

Appendices

A Additional Figures and Tables

A.1 Descriptive Statistics

Table A.1: Sample Summary Statistics: Parental Income

	Parental Income Quartiles							
	Q1		Q2		Q3		Q4	
	Share	*Avg*	*Share*	*Avg*	*Share*	*Avg*	*Share*	*Avg*
Two-Year School								
Grants	0.69	4,668	0.53	4,844	0.43	4,265	0.29	4,999
Parental Transfers	0.21	2,100	0.28	2,400	0.41	3,138	0.57	4,404
Loans Per Year	0.29	8,028	0.30	8,998	0.28	7,598	0.18	7,218
Work Full-Time	0.30	11.84	0.34	12.12	0.31	12.96	0.29	14.05
Four-Year School								
Grants	0.77	9,376	0.69	10,275	0.61	9,145	0.44	9,501
Parental Transfers	0.28	4,997	0.43	5,558	0.54	8,249	0.71	14,000
Loans Per Year	0.56	7,615	0.61	6,972	0.57	8,047	0.38	8,998
Work Full-Time	0.20	12.67	0.20	12.88	0.19	13.14	0.15	14.43
Graduate School								
Grants	0.33	7,716	0.33	14,858	0.38	13,417	0.42	21,027
Loans Per Year	0.59	12,787	0.56	15,166	0.47	20,280	0.43	26,187
Work Full-Time	0.69	20.49	0.54	18.23	0.49	17.73	0.45	18.98

Notes: This table reports the main sample statistics across parental income quartiles of individuals enrolled at the different institution types. There is no parental transfers information for graduate school, and for this reason it is not being reported. The average for "Work Full-Time" reports the average hourly wage among those individuals that are working full time while enrolled. Monetary amounts represent 2008 dollars. Data comes from the NLSY97 sample.

Table A.2: Sample Summary Statistics: Occupation

	Business	STEM	Social Sciences	Education	Humanities	Health	Sales & Office	Production
Demographics								
Share Female	0.51	0.21	0.69	0.79	0.52	0.81	0.61	0.28
Share Black	0.16	0.11	0.20	0.22	0.14	0.27	0.24	0.25
AFQT Score	58	67	60	57	63	45	45	37
Parental Income	101,206	101,565	93,620	86,894	103,316	74,794	75,408	67,652
Share ParInc Q1	0.17	0.10	0.23	0.19	0.13	0.25	0.28	0.30
Share ParInc Q4	0.38	0.39	0.30	0.29	0.37	0.21	0.21	0.16
Labor Market								
Wage Not Graduated	18.62	21.37	16.78	12.86	21.60	15.66	13.95	15.28
Wage Graduated	23.28	26.36	19.83	17.32	22.27	24.67	18.49	18.21
Working Full-Time	0.82	0.82	0.73	0.63	0.52	0.67	0.68	0.68

Notes: This table represents the main statistics of individuals that work at specific occupations. The row "Wage Not Graduated" represents the wage of an individual works at this occupation but does not have any higher education degree. Similarly, the "Wage Graduated" row is the wage of an individual with a degree. Data comes from the NLSY97 sample.

A.2 Results

Table A.3: Logit Grants and Transfers

	Parental Transfers	Grants 2y	Grants 4y	Grants Grad
ParInc Q2	0.438	-0.563	-0.410	0.363
	(0.057)	(0.078)	(0.084)	(0.155)
ParInc Q3	0.856	-0.930	-0.710	0.526
	(0.055)	(0.079)	(0.080)	(0.151)
ParInc Q4	1.555	-1.492	-1.362	0.816
	(0.056)	(0.086)	(0.078)	(0.147)
AFQT Q2	-0.015	-0.076	0.076	0.033
	(0.060)	(0.077)	(0.094)	(0.226)
AFQT Q3	0.142	-0.075	0.149	1.580
	(0.059)	(0.081)	(0.090)	(0.182)
AFQT Q4	0.384	-0.019	0.501	2.568
	(0.061)	(0.090)	(0.091)	(0.178)
Female	0.008	0.515	0.304	0.449
	(0.035)	(0.056)	(0.044)	(0.079)
Black	-0.148	0.566	0.788	0.554
	(0.045)	(0.068)	(0.063)	(0.105)
Four-Year School	0.490			
	(0.037)			
Constant	-1.384	0.317	0.592	-6.743
	(0.065)	(0.082)	(0.098)	(0.196)

Notes: This table reports the estimates and standard errors in parenthesis of the coefficients of the financial functions, obtained during estimation. The values correspond to the coefficients governing all the logit financial functions of the model.

Table A.4: Financial Functions

	Parental Transfers	Grants 2y	Grants 4y	Grants Grad
ParInc Q2	0.130	-0.106	-0.053	-0.073
	(0.058)	(0.058)	(0.049)	(0.213)
ParInc Q3	0.454	-0.158	-0.207	0.037
	(0.056)	(0.063)	(0.048)	(0.205)
ParInc Q4	0.965	0.042	-0.212	0.353
	(0.054)	(0.075)	(0.049)	(0.197)
AFQT Q2	-0.008	0.164	0.124	-0.228
	(0.057)	(0.059)	(0.063)	(0.306)
AFQT Q3	-0.030	0.038	0.130	0.106
	(0.056)	(0.065)	(0.061)	(0.245)
AFQT Q4	0.165	-0.033	0.380	0.854
	(0.056)	(0.074)	(0.062)	(0.242)
Female	-0.043	-0.012	-0.020	-0.169
	(0.029)	(0.047)	(0.031)	(0.110)
Black	-0.230	0.346	0.110	0.115
	(0.041)	(0.051)	(0.038)	(0.147)
Four-Year School	1.006			
	(0.033)			
Constant	6.842	7.726	8.496	8.135
	(0.063)	(0.065)	(0.063)	(0.263)

Notes: This table reports the estimates and standard errors in parenthesis of the coefficients of the financial functions, obtained during estimation. The values correspond to the coefficients governing all financial transfers of the log functions.

Table A.5: Graduation Probablity

	Two-Year School	Four-Year School	Graduate School
ParInc Q2	0.219	-0.101	0.293
	(0.142)	(0.151)	(0.412)
ParInc Q3	0.219	0.111	-0.182
	(0.142)	(0.143)	(0.401)
ParInc Q4	0.113	0.117	-0.017
	(0.149)	(0.140)	(0.390)
AFQT Q2	0.124	0.066	-0.096
	(0.145)	(0.190)	(0.716)
AFQT Q3	0.280	0.130	0.442
	(0.146)	(0.182)	(0.648)
AFQT Q4	0.437	0.450	0.221
	(0.159)	(0.182)	(0.643)
Female	-0.079	0.291	-0.057
	(0.099)	(0.079)	(0.209)
Black	-0.370	-0.480	-0.548
	(0.131)	(0.109)	(0.298)

Notes: This table reports the estimates and standard errors in parenthesis of the coefficients of the graduation probability functions, obtained during estimation. The base alternative is "Not Graduation". The base category corresponds to "ParInc Q1", "AFQT Q1", "Male" and "Non-Black".

Table A.6: Estimates of Wage Parameters

	Enrolled	Business	STEM	Social Sciences	Education	Humanities	Health	Sales & Office	Production
ParInc Q2	0.038	0.094	0.041	0.107	0.184	0.144	0.064	0.035	0.090
	(0.013)	(0.033)	(0.056)	(0.064)	(0.050)	(0.125)	(0.025)	(0.013)	(0.012)
ParInc Q3	0.044	0.094	0.067	0.044	0.141	0.136	0.152	0.050	0.168
	(0.012)	(0.035)	(0.055)	(0.054)	(0.048)	(0.123)	(0.027)	(0.014)	(0.013)
ParInc Q4	0.123	0.190	0.100	0.335	0.227	0.212	0.193	0.162	0.217
	(0.012)	(0.033)	(0.053)	(0.061)	(0.049)	(0.118)	(0.029)	(0.014)	(0.014)
AFQT Q2	0.027	0.106	0.125	-0.036	0.138	-0.011	0.072	0.046	0.072
	(0.014)	(0.034)	(0.057)	(0.064)	(0.052)	(0.128)	(0.024)	(0.012)	(0.011)
AFQT Q3	0.031	0.102	0.210	0.118	0.151	-0.033	0.155	0.103	0.065
	(0.013)	(0.034)	(0.054)	(0.065)	(0.055)	(0.125)	(0.027)	(0.014)	(0.013)
AFQT Q4	0.027	0.142	0.244	0.025	0.195	0.041	0.249	0.102	0.002
	(0.014)	(0.036)	(0.054)	(0.067)	(0.056)	(0.127)	(0.033)	(0.016)	(0.016)
Female	-0.071	-0.101	-0.073	0.002	-0.107	-0.059	-0.021	-0.115	-0.172
	(0.008)	(0.021)	(0.038)	(0.043)	(0.042)	(0.078)	(0.023)	(0.010)	(0.010)
Black	-0.048	-0.057	0.011	-0.126	0.008	0.262	-0.034	-0.034	-0.091
	(0.010)	(0.029)	(0.047)	(0.049)	(0.041)	(0.098)	(0.022)	(0.011)	(0.011)
Experience	0.081	0.060	0.081	0.035	0.030	0.069	0.043	0.055	0.063
	(0.003)	(0.005)	(0.007)	(0.010)	(0.009)	(0.018)	(0.004)	(0.002)	(0.002)
Constant	2.196	2.361	2.358	2.369	2.153	2.459	2.285	2.267	2.247
	(0.015)	(0.038)	(0.062)	(0.070)	(0.063)	(0.134)	(0.034)	(0.015)	(0.012)

Notes: This table reports the estimates and standard errors in parenthesis of the coefficients of the wage functions, obtained during estimation. Coefficients for the education returns categories "Associate Degree", "Field of Study" or "Graduate Degree" are reported in Table A.7. The base category corresponds to "ParInc Q1", "AFQT Q1", "Male" and "Non-Black".

Table A.7: Education-Occupation Wage Complementarities

	Business	STEM	Social Sciences	Education	Humanities	Health	Sales & Office	Production
Business	0.352						0.486	0.371
	(0.033)						(0.026)	(0.055)
STEM	0.455	0.461				0.381	0.331	0.410
	(0.048)	(0.039)				(0.054)	(0.045)	(0.063)
Education	0.161			0.277			0.113	0.097
	(0.099)			(0.042)			(0.082)	(0.115)
Social Sciences	0.351		0.310	0.147			0.300	0.385
	(0.047)		(0.052)	(0.062)			(0.036)	(0.053)
Humanities	0.290		-0.010	0.167	0.103		0.222	0.299
	(0.044)		(0.074)	(0.054)	(0.107)		(0.027)	(0.044)
Health	0.168			0.267		0.599	0.450	0.409
	(0.136)			(0.101)		(0.039)	(0.082)	(0.099)
Other	0.205	0.269	0.157				0.136	0.282
	(0.077)	(0.094)	(0.082)				(0.061)	(0.074)
Associate Degree	0.162	0.160			-0.049	0.392	0.142	0.265
	(0.047)	(0.044)			(0.122)	(0.027)	(0.022)	(0.023)
Graduate Degree	0.189	-0.025	0.140	0.378	0.340	0.454	0.049	0.211
	(0.058)	(0.064)	(0.074)	(0.067)	(0.223)	(0.113)	(0.097)	(0.092)

Notes: This table reports the estimates and standard errors in parenthesis of the coefficients of the wage functions, obtained during estimation. The output only focuses on the returns to different fields or degrees. Since depending on which field of graduation only some occupations are feasible, some effects are not reported since they are unfeasible according to the model.

Table A.8: Flow-Payoff Parameters of Education Alternatives

	Associate	Business	STEM	Undeclared	Education	Social Sciences	Humanities	Health	Other	Grad School
ParInc Q2	0.21	0.38	0.33	0.33	0.16	0.05	0.24	0.40	0.11	0.21
	(0.37)	(0.72)	(0.69)	(0.89)	(0.84)	(0.80)	(0.78)	(0.82)	(0.84)	(0.93)
ParInc Q3	0.44	0.54	0.49	0.26	0.60	0.25	0.39	0.64	0.36	0.28
	(0.40)	(0.63)	(0.62)	(0.88)	(0.84)	(0.75)	(0.63)	(0.79)	(0.87)	(0.94)
ParInc Q4	0.47	0.66	0.56	0.67	0.55	0.43	0.66	0.85	0.53	0.20
	(0.43)	(0.62)	(0.60)	(0.80)	(0.88)	(0.59)	(0.55)	(0.78)	(0.86)	(0.91)
AFQT Q2	0.44	−0.36	−0.45	1.03	−0.41	−0.39	−0.34	−0.39	−0.57	−0.05
	(0.36)	(0.75)	(0.77)	(0.91)	(0.90)	(0.82)	(0.76)	(0.86)	(0.91)	(0.94)
AFQT Q3	0.85	0.04	0.03	2.08	−0.01	0.06	0.03	0.02	−0.20	0.09
	(0.40)	(0.66)	(0.70)	(0.80)	(0.85)	(0.71)	(0.69)	(0.87)	(0.87)	(0.91)
AFQT Q4	1.09	0.40	0.61	2.71	0.35	0.50	0.56	0.36	−0.09	0.01
	(0.48)	(0.68)	(0.68)	(0.75)	(0.80)	(0.69)	(0.66)	(0.82)	(0.82)	(0.87)
Female	0.28	0.03	−0.06	0.07	0.53	0.29	0.15	0.61	0.09	0.11
	(0.18)	(0.56)	(0.58)	(0.75)	(0.84)	(0.67)	(0.65)	(0.80)	(0.81)	(0.87)
Black	0.05	0.32	0.32	−0.07	−0.01	0.22	0.12	0.35	0.02	0.19
	(0.25)	(0.82)	(0.75)	(0.95)	(0.85)	(0.71)	(0.81)	(0.82)	(0.94)	(0.92)
Part-Time Work	−0.24	−0.34	−0.54	−0.71	−0.50	−0.38	−0.51	−0.51	−0.06	0.33
	(0.22)	(0.55)	(0.45)	(0.62)	(0.77)	(0.51)	(0.53)	(0.86)	(0.87)	(0.74)
Full-Time Work	−1.13	−1.54	−1.94	−2.28	−1.89	−2.09	−2.03	−1.96	−1.45	−0.36
	(0.28)	(0.59)	(0.59)	(0.84)	(0.83)	(0.67)	(0.67)	(0.88)	(0.89)	(0.64)
Switching Cost	−0.48	−2.76	−2.90	−1.58	−3.33	−2.79	−2.66	−2.96	−1.51	
	(0.31)	(0.81)	(0.78)	(0.93)	(0.93)	(0.83)	(0.85)	(0.83)	(0.77)	
Constant	−4.78	−6.79	−6.02	−7.55	−6.23	−5.94	−5.93	−7.32	−7.19	−3.54
	(0.59)	(0.74)	(0.72)	(0.72)	(0.78)	(0.73)	(0.73)	(0.79)	(0.80)	(0.78)

Notes: This table reports the flow-payoff estimates for each education category (occupational categories are reported in Table A.9). The reference alternative is "Home Production". The reference category includes "Parental Income Q1", "AFQT Q1", "Male", "Non-Black" and "Non-Work". There is no switching cost associated with the graduate school discrete alternative, for this reason there is no associated coefficient. I omit the following coefficients: "AFQT $\times$ Education Experience" that capture the learning about your own preferences and abilities effects. I also omit the "Work $\times$ Period" coefficients that capture different tastes for work across time.

Table A.9: Flow-Payoff Parameters of Occupation Alternatives

	Business	STEM	Social Sciences	Education	Humanities	Health	Sales & Office	Production
ParInc Q2	0.17	0.28	−0.36	0.10	0.10	0.16	0.12	0.12
	(0.60)	(0.93)	(0.95)	(0.91)	(0.91)	(0.65)	(0.27)	(0.24)
ParInc Q3	0.16	0.35	0.17	0.20	0.43	0.17	0.12	0.11
	(0.58)	(0.93)	(0.86)	(0.90)	(0.84)	(0.72)	(0.30)	(0.26)
ParInc Q4	0.41	0.53	0.02	0.23	0.47	0.15	0.13	0.07
	(0.53)	(0.91)	(0.88)	(0.87)	(0.87)	(0.80)	(0.34)	(0.33)
AFQT Q2	0.09	0.06	0.01	0.10	0.55	−0.07	0.07	−0.01
	(0.62)	(0.93)	(0.90)	(0.92)	(0.89)	(0.70)	(0.36)	(0.27)
AFQT Q3	0.12	0.11	−0.09	−0.11	0.62	−0.18	0.02	−0.11
	(0.55)	(0.91)	(0.93)	(0.90)	(0.88)	(0.72)	(0.42)	(0.32)
AFQT Q4	−0.00	−0.04	−0.05	−0.10	0.48	−0.47	−0.11	−0.23
	(0.60)	(0.85)	(0.89)	(0.88)	(0.87)	(0.81)	(0.49)	(0.40)
Female	0.06	−0.87	0.40	0.71	−0.03	0.70	0.17	−0.27
	(0.41)	(0.95)	(0.82)	(0.79)	(0.85)	(0.70)	(0.16)	(0.14)
Black	−0.26	−0.50	−0.37	−0.14	−0.31	−0.26	−0.22	−0.24
	(0.86)	(0.97)	(0.94)	(0.86)	(0.94)	(0.74)	(0.25)	(0.19)
Full-Time Work	1.11	1.20	0.57	0.25	−0.37	0.11	0.18	0.16
	(0.70)	(0.79)	(0.81)	(0.72)	(0.83)	(0.38)	(0.24)	(0.23)
Switching Cost	−3.02	−3.75	−3.76	−3.53	−3.74	−3.63	−2.60	−2.44
	(0.53)	(0.95)	(0.93)	(0.92)	(0.91)	(0.87)	(0.12)	(0.09)
Constant	−5.40	−6.20	−6.87	−5.68	−5.80	−4.60	−2.83	−2.09
	(0.75)	(0.80)	(0.95)	(0.80)	(0.77)	(0.75)	(0.51)	(0.35)

Notes: This table reports the flow-payoff estimates for each occupation category (education categories are reported in Table A.8). The reference alternative is "Home Production". The reference category includes "Parental Income Q1", "AFQT Q1", "Male", "Non-Black" and "Part-Time". I omit the following coefficients "Full-Time × Period" that capture different tastes for work across time. I also omit the coefficients for the different effects of fields of study and graduate degrees on preferences, that are reported in Table A.10.

Table A.10: Education-Occupation Flow Payoff Complementarities

	Business	STEM	Social Sciences	Education	Humanities	Health	Sales & Office	Production
Associate Degree	0.15	0.75			0.24	0.28	0.04	0.01
	(0.77)	(0.84)			(0.84)	(0.61)	(0.37)	(0.31)
Business	1.59						1.15	0.52
	(0.66)						(0.62)	(0.78)
STEM	0.99	1.89				0.96	0.45	0.12
	(0.82)	(0.80)				(0.86)	(0.87)	(0.88)
Education	0.37			1.80			0.17	−0.06
	(0.94)			(0.86)			(0.93)	(0.92)
Social Sciences	0.74		1.67	1.05			0.37	0.05
	(0.71)		(0.81)	(0.88)			(0.80)	(0.76)
Humanities	0.93		1.22	1.34	1.36		0.77	0.42
	(0.72)		(0.90)	(0.81)	(0.87)		(0.66)	(0.73)
Health	0.59			1.65		1.85	0.58	0.57
	(0.90)			(0.92)		(0.88)	(0.95)	(0.95)
Other	1.55	1.84	2.10				1.05	0.97
	(0.96)	(0.92)	(0.94)				(0.97)	(0.97)
Graduate Degree	0.21	0.81	0.42	0.49	0.34	0.34	−0.47	−0.02
	(0.95)	(0.94)	(0.92)	(0.95)	(0.93)	(0.97)	(0.88)	(0.97)

Notes: This table reports the flow-payoff estimates for the educational effects on the different occupation categories. The coefficients represent the non-pecuniary effect of a degree or field of study return on the utility of a specific occupation. Because some occupations are not feasible for given fields, the coefficients are non-existent and therefore not reported in the table. See Appendix D for a description of which choices are feasible for each field of graduation.

Table A.11: Choice Frequencies by Data and Model

	All Periods		Period 1		Period 9	
	Data	Model	Data	Model	Data	Model
	(1)	(2)	(3)	(4)	(5)	(6)
Not Educ, Not Work	0.14	0.13	0.13	0.13	0.15	0.14
Not Educ, Part-Time	0.15	0.14	0.13	0.14	0.15	0.15
Not Educ, Full-Time	0.40	0.37	0.19	0.19	0.54	0.52
Two-Year, No-Work	0.03	0.03	0.06	0.07	0.02	0.02
Two-Year, Part-Time	0.04	0.05	0.11	0.10	0.02	0.02
Two-Year, Full-Time	0.04	0.04	0.05	0.05	0.03	0.03
Four-Year, No Work	0.06	0.08	0.16	0.15	0.02	0.03
Four-Year, Part-Time	0.08	0.09	0.14	0.14	0.01	0.02
Four-Year, Full-Time	0.05	0.06	0.03	0.04	0.03	0.03
Grad Sch, No Work	0.00	0.00	0.00	0.00	0.01	0.01
Grad Sch, Part-Time	0.00	0.00	0.00	0.00	0.01	0.01
Grad Sch, Full-Time	0.01	0.01	0.00	0.00	0.02	0.02

Notes: This table reports the frequency of each choice educational and working choice in the model. Choices are aggregated across multiple occupations and fields of study to obtain the reported statistics. Columns (1) and (2) nest all periods together. Columns (3) and (4) only use observations at period 1 and Columns (5) and (6) only use observations at the last period. Model choices are obtained by simulating 30 times each individual in the sample. See Section 5 for a detailed description of the simulation.

A.3 Counterfactual

Table A.12: Grant Equivalence Four-Year School at T=1

	Low Schooling Type			High Schooling Type		
	Baseline	Counterfactual	Grant	Baseline	Counterfactual	Grant
ParInc Q1	0.04	0.06 **(+44.82%)**	30,486$	0.35	0.41 **(+18.38%)**	17,506$
ParInc Q2	0.10	0.13 **(+35.11%)**	25,441$	0.48	0.54 **(+12.44%)**	11,083$
ParInc Q3	0.17	0.22 **(+27.69%)**	21,727$	0.61	0.66 **(+7.95%)**	7,294$
ParInc Q4	0.32	0.36 **(+11.54%)**	15,418$	0.74	0.77 **(+2.86%)**	2,640$
AFQT Q1	0.01	0.02 **(+48.85%)**	32,505$	0.24	0.31 **(+30.27%)**	21,905$
AFQT Q2	0.04	0.06 **(+39.95%)**	22,778$	0.47	0.53 **(+12.92%)**	7,987$
AFQT Q3	0.13	0.18 **(+37.13%)**	21,353$	0.71	0.75 **(+5.44%)**	4,768$
AFQT Q4	0.47	0.54 **(+15.00%)**	15,864$	0.81	0.83 **(+2.59%)**	3,028$
Male	0.13	0.17 **(+24.71%)**	27,620$	0.51	0.56 **(+8.86%)**	11,232$
Female	0.17	0.21 **(+19.93%)**	19,230$	0.57	0.63 **(+9.32%)**	8,441$
Non-Black	0.18	0.22 **(+21.46%)**	23,081$	0.57	0.61 **(+7.94%)**	9,147$
Black	0.08	0.10 **(+25.86%)**	24,696$	0.46	0.52 **(+13.22%)**	11,917$

Notes: This table shows the probability of enrollment in a four-year school for both latent unobserved types "Low" or "High" schooling type. The baseline probability refers to the probability obtained by the baseline model and the counterfactual probability is the one predicted by the SAVE repayment plan scenario. Enrollment probabilities represent the sum of individuals enrolling with different labor supply decisions and across all possible fields of study. Grants refers to an hypothetical grant equivalence at period 1 to obtain the same enrollment probabilities under the standard 10 year repayment plan. To find the grant amount I simulate 1,000 grant values and minimize the distance between the Counterfactual enrollment probability and the grants. Aggregation is weighted to respect the proportionality of each group.

Table A.13: Decomposition of New Graduates

			Counterfactual Decomposition		
	Baseline	Counterfactual	Never	Drop Out	Two Year
	(1)	(2)	(3)	(4)	(5)
ParInc Q1	0.10	0.12 **(+18.51%)**	0.11 **(+1.99%)**	0.75 **(+13.96%)**	0.14 **(+2.55%)**
ParInc Q2	0.16	0.18 **(+13.14%)**	0.07 **(+0.86%)**	0.73 **(+9.59%)**	0.20 **(+2.69%)**
ParInc Q3	0.28	0.31 **(+10.64%)**	0.07 **(+0.75%)**	0.71 **(+7.52%)**	0.22 **(+2.37%)**
ParInc Q4	0.42	0.44 **(+5.23%)**	0.06 **(+0.30%)**	0.79 **(+4.11%)**	0.16 **(+0.82%)**
AFQT Q1	0.06	0.07 **(+31.56%)**	0.15 **(+4.66%)**	0.73 **(+23.06%)**	0.12 **(+3.84%)**
AFQT Q2	0.15	0.17 **(+15.17%)**	0.04 **(+0.65%)**	0.72 **(+10.86%)**	0.24 **(+3.66%)**
AFQT Q3	0.27	0.29 **(+7.05%)**	0.04 **(+0.30%)**	0.73 **(+5.12%)**	0.23 **(+1.63%)**
AFQT Q4	0.50	0.53 **(+6.37%)**	0.08 **(+0.49%)**	0.77 **(+4.93%)**	0.15 **(+0.95%)**
Male	0.20	0.21 **(+9.23%)**	0.08 **(+0.72%)**	0.78 **(+7.18%)**	0.14 **(+1.33%)**
Female	0.27	0.30 **(+9.94%)**	0.07 **(+0.70%)**	0.72 **(+7.11%)**	0.21 **(+2.13%)**
Non Black	0.26	0.28 **(+8.55%)**	0.07 **(+0.62%)**	0.74 **(+6.31%)**	0.19 **(+1.62%)**
Black	0.16	0.18 **(+14.76%)**	0.08 **(+1.11%)**	0.75 **(+11.10%)**	0.17 **(+2.55%)**

Notes: This table reports the graduation effects of the policy, together with a decomposition of where are the new graduates coming from. Column (1) reports the graduation shares on the baseline scenario and Column (2) reports the graduation shares under the SAVE repayment plan, together with the percentage increase that the change represents. Columns (3), (4) and (5) focus on the new graduated individuals, and are decomposed in previously never enrolled, previously drop-out and previously two-year respectively. The first entry of each of those columns represents the share that they represent within the new graduates, and the second represents the percentage change that each group accounts.

Table A.14: Enrollment and Paths In Four-Year Schools After SAVE

	New Enrollment		Drop Out		Graduate	
	Share	High Type	Share	High Type	Share	High Type
	(1)	(2)	(3)	(4)	(5)	(6)
ParInc Q1	0.03	0.47	0.81	0.39	0.19	0.83
ParInc Q2	0.04	0.26	0.86	0.21	0.14	0.56
ParInc Q3	0.06	0.16	0.74	0.11	0.26	0.29
ParInc Q4	0.06	0.09	0.78	0.05	0.22	0.24
AFQT Q1	0.03	0.70	0.75	0.61	0.25	0.98
AFQT Q2	0.03	0.28	0.88	0.20	0.12	0.90
AFQT Q3	0.08	0.02	0.90	0.02	0.10	0.04
AFQT Q4	0.20	0.01	0.66	0.00	0.34	0.02
Male	0.04	0.29	0.85	0.26	0.15	0.47
Female	0.05	0.24	0.75	0.15	0.25	0.51
Non Black	0.05	0.22	0.80	0.17	0.20	0.39
Black	0.04	0.41	0.82	0.32	0.18	0.78

Notes: Columns (1) and (2) refer to the share of new enrolled individuals after SAVE is introduced. Column (1) reports the share of individuals that are not enrolled in the baseline but enroll after SAVE is introduced and Column (2) focuses on the unobserved schooling type composition, reporting the share of individuals with a high schooling type. The next categories ("Drop Out" and "Graduate") focus on the individuals that react to the policy, so its conditioning on being a new enrolled. Columns (3) and (4) report the share of individuals that drop out after being a new enrolled and the distribution of unobserved type. Columns (5) and (6) report the same statistic but focusing on those that achieve graduation.

Table A.15: Drop Out Decomposition

	Data	Baseline	SAVE	Previous Enrolled	New Enrolled
	(1)	(2)	(3)	(4)	(5)
ParInc Q1	0.68	0.60	0.58	0.54	0.81
ParInc Q2	0.59	0.56	0.55	0.52	0.86
ParInc Q3	0.46	0.43	0.41	0.38	0.74
ParInc Q4	0.38	0.36	0.35	0.33	0.78
AFQT Q1	0.75	0.63	0.60	0.56	0.75
AFQT Q2	0.63	0.54	0.52	0.48	0.88
AFQT Q3	0.53	0.48	0.48	0.45	0.90
AFQT Q4	0.34	0.36	0.35	0.33	0.66
Male	0.53	0.51	0.50	0.47	0.85
Female	0.46	0.41	0.40	0.37	0.75
Non Black	0.45	0.43	0.43	0.40	0.80
Black	0.65	0.55	0.53	0.50	0.82

Notes: This table shows the drop out rates in a four-year school. Column (1) reports the shares in the data. Column (2) reports the shares in the baseline simulation. Column (3) reports the share in the SAVE counterfactual. Column (4) reports the share of individuals that drop out under SAVE, conditional on attending school in the baseline. Column (5) reports the share of individuals that drop out, conditional on not having enrolled in the baseline. Drop out is considered if the individual has attended a four-year school and by the last period of the model has not achieved graduation.

Table A.16: Labor Supply Distribution While Enrolled

	All Individuals		Not Indebted		Indebted	
	Base	Counter	Base	Counter	Base	Counter
	(1)	(2)	(3)	(4)	(5)	(6)
Two-Year, No Work	0.26	0.30 **(+13.93%)**	0.27	0.33 **(+24.44%)**	0.25	0.24 **(-5.82%)**
Two-Year, Part-Time	0.39	0.37 **(-3.39%)**	0.41	0.38 **(-7.62%)**	0.34	0.36 **(+5.35%)**
Two-Year, Full-Time	0.35	0.33 **(-6.71%)**	0.32	0.29 **(-10.63%)**	0.40	0.40 **(-0.90%)**
Four-Year, No Work	0.34	0.37 **(8.19%)**	0.37	0.43 **(16.05%)**	0.32	0.32 **(0.16%)**
Four-Year, Part-Time	0.41	0.40 **(-3.66%)**	0.44	0.40 **(-9.50%)**	0.40	0.40 **(1.33%)**
Four-Year, Full-Time	0.24	0.23 **(-5.33%)**	0.19	0.17 **(-9.29%)**	0.28	0.28 **(-2.05%)**
Grad School, No Work	0.25	0.20 **(-17.28%)**	0.23	0.23 **(-0.41%)**	0.25	0.20 **(-19.61%)**
Grad School, Part-Time	0.24	0.27 **(+9.63%)**	0.28	0.27 **(-3.57%)**	0.24	0.27 **(+11.52%)**
Grad School, Full-Time	0.51	0.53 **(+3.71%)**	0.49	0.50 **(+2.24%)**	0.51	0.53 **(+4.12%)**

Notes: This table reports the change in labor supply decision of individuals under the new repayment plan, with a focus in the difference between individuals that were indebted or not indebted in the baseline scenario. Columns (1) and (2) report the baseline and counterfactual shares of choices, together with the percentage change. Columns (3) and (4) focuses on individuals that are not indebted in the baseline scenario and Columns (5) and (6) focuses on individuals that are indebted in the baseline scenario. To compute shares of choices I aggregate across all fields. Shares sum to one within each educational category ("Two-Year", "Four-Year" and "Grad-School").

Table A.17: Change in Student Debt Composition

	Share Indebted		Average Debt	
	Baseline	Counterfactual	Baseline	Counterfactual
Overall Effect	0.58	0.86 **(+48.28%)**	18, 358	22, 925 **(+24.87 %)**
ParInc Q1	0.70	0.82 **(+16.94%)**	22, 163	24, 070 **(+8.60%)**
ParInc Q2	0.71	0.84 **(+18.49%)**	23, 060	23, 135 **(+0.33%)**
ParInc Q3	0.71	0.86 **(+20.23%)**	20, 741	22, 753 **(+9.70%)**
ParInc Q4	0.43	0.89 **(+105.32%)**	12, 182	22, 735 **(+86.62%)**
AFQT Q1	0.65	0.79 **(+21.61%)**	20, 190	25, 164 **(+24.63%)**
AFQT Q2	0.66	0.83 **(+24.68%)**	20, 317	24, 268 **(+19.45%)**
AFQT Q3	0.66	0.85 **(+29.91%)**	22, 625	23, 174 **(+2.42%)**
AFQT Q4	0.53	0.88 **(+66.35%)**	15, 499	22, 418 **(+44.64%)**
Male	0.54	0.86 **(+58.79%)**	18, 439	23, 265 **(+26.17%)**
Female	0.61	0.87 **(+42.24%)**	18, 304	22, 679 **(+23.90%)**
Non-Black	0.57	0.87 **(+52.22%)**	18, 257	22, 88 **(+25.35%)**
Black	0.64	0.83 **(+30.99%)**	18, 919	23, 145 **(+22.33%)**

Notes: This table reports the share of indebted individuals and the average debt amounts among indebted individuals comparing the baseline and the counterfactual scenario. Shares and debt levels are computed at the year of graduation of individuals. Monetary amounts are reported in 2008 dollars.

Table A.18: Differences in Switching Field Behavior

	Not Indebted	Indebted
ParInc Q1	0.14	0.17 **(+24.01%)**
ParInc Q2	0.14	0.18 **(+32.67%)**
ParInc Q3	0.17	0.17 **(-0.26%)**
ParInc Q4	0.13	0.12 **(-5.43%)**
AFQT Q1	0.13	0.15 **(+13.64%)**
AFQT Q2	0.13	0.18 **(+34.29%)**
AFQT Q3	0.16	0.17 **(+6.52%)**
AFQT Q4	0.14	0.13 **(-5.26%)**
Male	0.13	0.14 **(+7.86%)**
Female	0.16	0.16 **(+4.21%)**
Non Black	0.15	0.15 **(+0.57%)**
Black	0.13	0.17 **(+30.17%)**

Notes: This table reports the share of individuals that switch fields under the counterfactual scenario. A switcher is defined as an individual that chooses enrollment in a four-year school in both the baseline and the counterfactual, but decides to switch field in the counterfactual scenario. I only identify switchers the first period that they switch. The different columns represent whether individuals where indebted or not in the baseline scenario.

Table A.19: Labor Supply Changes When Changing Field

	Not Indebted			Indebted		
	Same	Less	More	Same	Less	More
Overall Effect	0.22	0.62	0.16	0.28	0.38	0.34
ParInc Q1	0.23	0.59	0.19	0.29	0.31	0.40
ParInc Q2	0.21	0.64	0.15	0.30	0.30	0.41
ParInc Q3	0.24	0.62	0.14	0.29	0.35	0.36
ParInc Q4	0.22	0.62	0.17	0.27	0.49	0.24
AFQT Q1	0.25	0.61	0.14	0.26	0.36	0.39
AFQT Q2	0.22	0.62	0.16	0.27	0.32	0.41
AFQT Q3	0.22	0.62	0.16	0.29	0.32	0.39
AFQT Q4	0.22	0.62	0.16	0.29	0.45	0.25
Male	0.23	0.61	0.16	0.28	0.39	0.33
Female	0.22	0.62	0.16	0.29	0.37	0.34
Non Black	0.22	0.62	0.16	0.29	0.38	0.33
Black	0.23	0.60	0.17	0.28	0.35	0.37

Notes: This table reports the change in labor supply decision of individuals that decide to switch field. A switcher is defined as an individual that chooses enrollment in a four-year school in both the baseline and the counterfactual, but decides to switch field in the counterfactual scenario. I only identify switchers the first period that they switch. "Same" means the individual switched field but is working the same amount of hours, "Less" means the individual switched field and is working less hours and "More" means the individual is switching fields and is working more hours. I report the statistics both for individuals that in the baseline are indebted or not.

Table A.20: Grant Equivalence Two-Year School at T=1

	Low Schooling Type				High Schooling Type			
	Baseline	Counterfactual		Grant	Baseline	Counterfactual		Grant
ParInc Q1	0.16	0.22	**(+38.44%)**	54,126	0.32	0.33	**(+5.52%)**	5,938
ParInc Q2	0.20	0.26	**(+29.81%)**	41,249	0.28	0.27	**(-1.54%)**	0
ParInc Q3	0.25	0.30	**(+20.21%)**	29,342	0.24	0.22	**(-7.70%)**	0
ParInc Q4	0.22	0.26	**(+17.96%)**	20,305	0.16	0.15	**(-4.87%)**	0
AFQT Q1	0.13	0.18	**(+44.95%)**	63,253	0.34	0.38	**(+10.38%)**	8,491
AFQT Q2	0.20	0.28	**(+38.32%)**	53,857	0.32	0.30	**(-4.19%)**	0
AFQT Q3	0.29	0.35	**(+21.95%)**	21,541	0.19	0.16	**(-12.34%)**	0
AFQT Q4	0.21	0.22	**(+4.39%)**	2,437	0.13	0.12	**(-11.64%)**	0
Male	0.18	0.22	**(+26.13%)**	36,662	0.24	0.25	**(+4.37%)**	4,841
Female	0.23	0.29	**(+25.54%)**	36,945	0.26	0.25	**(-6.17%)**	0
Non-Black	0.21	0.26	**(+23.52%)**	33,848	0.24	0.23	**(-1.34%)**	0
Black	0.19	0.26	**(+32.98%)**	45,329	0.29	0.29	**(-0.49%)**	0

Notes: This table shows the probability of enrollment in a two-year school for both latent unobserved types "Low" or "High" schooling type. The baseline probability refers to the probability obtained by the baseline model and the counterfactual probability is the one predicted by the SAVE repayment plan scenario. Enrollment probabilities represent the sum of individuals enrolling with different labor supply decisions. Grants refers to an hypothetical grant equivalence at period 1 to obtain the same enrollment probabilities under the standard 10year repayment plan. Aggregation is weighted to respect the proportionality of each group.

Table A.21: Graduation Decomposition Graduate School

	Baseline	Counterfactual		Decomposition New Graduates			
				Previous Drop Out		Never Enrolled	
ParInc Q1	0.007	0.008	**(+22.69%)**	0.072	**(+1.64%)**	0.928	**(+21.05%)**
ParInc Q2	0.015	0.018	**(+18.27%)**	0.121	**(+2.21%)**	0.879	**(+16.06%)**
ParInc Q3	0.022	0.028	**(+26.83%)**	0.093	**(+2.49%)**	0.907	**(+24.34%)**
ParInc Q4	0.033	0.038	**(+14.03%)**	0.124	**(+1.73%)**	0.876	**(+12.29%)**
AFQT Q1	0.003	0.004	**(+38.67%)**	0.036	**(+1.38%)**	0.964	**(+37.29%)**
AFQT Q2	0.006	0.009	**(+35.51%)**	0.157	**(+5.59%)**	0.843	**(+29.92%)**
AFQT Q3	0.023	0.027	**(+18.87%)**	0.089	**(+1.68%)**	0.911	**(+17.18%)**
AFQT Q4	0.048	0.056	**(+15.94%)**	0.115	**(+1.82%)**	0.885	**(+14.11%)**
Male	0.015	0.017	**(+15.29%)**	0.085	**(+1.30%)**	0.915	**(+13.98%)**
Female	0.023	0.028	**(+22.04%)**	0.118	**(+2.60%)**	0.882	**(+19.43%)**
Non Black	0.022	0.026	**(+18.45%)**	0.114	**(+2.10%)**	0.886	**(+16.35%)**
Black	0.011	0.014	**(+24.40%)**	0.075	**(+1.82%)**	0.925	**(+22.58%)**

Notes: This table reports the change in the share of graduates at Graduate Schools across demographic characteristics. Baseline represents the share of individuals with a Graduate Degree on the Baseline. Counterfactual represents the new share and the percentage change once SAVE is introduced. The "Decomposition of New Graduates" analyzes the origin of the individuals that are graduating under SAVE and not in the Baseline. "Previous Drop Out" are individuals that were previously dropping out from Graduate School. "Never Enrolled" are individuals that were not enrolled to Graduate School in the baseline.

Table A.22: Education Level After SAVE

Baseline \SAVE	Never	Drop Out	Two-Year	Four-Year	Grad
Never	0.844	0.140	0.007	0.009	0.001
Drop Out	0.000	0.918	0.028	0.053	0.003
Two-Year Grad	0.000	0.040	0.881	0.122	0.010
Four-Year Grad	0.000	0.026	0.024	0.963	0.088
Grad	0.000	0.006	0.011	0.986	0.903

Notes: This table reports the share of individuals that are at each state after SAVE, conditional on the education level at the baseline. Notice that graduation states are not mutually exclusive, and individual can have a two-year degree, a four-year degree and a graduate degree at the same time.

Table A.23: Changes Across Fields

	Total	Switchers			New	Previous	Previous
	Change	In	Out	Net	Enrolled	Drop Out	Two-Year
Parental Income Q1							
Business	0.17	0.13	−0.21	−0.09	0.03	0.18	0.04
STEM	0.16	0.12	−0.18	−0.07	0.02	0.17	0.03
Education	0.36	0.20	−0.12	0.08	0.03	0.23	0.03
Social Sciences	0.12	0.06	−0.16	−0.10	0.02	0.17	0.03
Humanities	0.15	0.10	−0.17	−0.07	0.03	0.17	0.03
Health	0.45	0.38	−0.19	0.19	0.03	0.19	0.04
Other	0.29	0.16	−0.17	−0.01	0.03	0.22	0.04
Parental Income Q4							
Business	0.02	0.09	−0.15	−0.06	0.00	0.07	0.01
STEM	0.05	0.07	−0.10	−0.02	0.00	0.06	0.01
Education	0.09	0.14	−0.13	0.01	0.00	0.06	0.01
Social Sciences	0.02	0.08	−0.13	−0.05	0.01	0.05	0.01
Humanities	0.08	0.10	−0.11	−0.00	0.00	0.06	0.01
Health	0.10	0.15	−0.14	0.02	0.00	0.07	0.01
Other	0.04	0.09	−0.15	−0.05	0.00	0.08	0.02

Notes: This table reports the change across fields for individuals at the bottom and top of the parental income distribution. Total change represents the percentage change in individuals holding that degree. The "Switchers" represents individuals that are either leaving a degree, or coming from another one. "New Enrolled" are individuals that were not enrolled in the baseline. "Previous Drop Out" are individuals that were dropping out in the baseline scenario. "Previous Two-Year" are individuals that were graduating from a Two-Year school in the baseline.

Table A.24: Changes Across Fields

	Business	STEM	Education	Social Sciences	Humanities	Health	Other
Overall Effect	−4.90	1.68	8.08	−4.45	0.07	9.77	2.27
ParInc Q1	−1.22	−2.29	15.09	−5.20	−2.57	22.17	8.90
ParInc Q2	−4.78	2.96	5.74	−4.84	0.86	9.86	1.84
ParInc Q3	−8.59	5.58	8.79	−7.55	−1.98	17.06	4.46
ParInc Q4	−2.71	−0.27	3.66	−3.29	2.30	4.49	−0.88
AFQT Q1	−7.88	4.14	11.23	−0.83	−1.82	−12.03	3.81
AFQT Q2	−1.88	−7.77	9.44	−5.13	−2.56	20.78	3.76
AFQT Q3	−4.28	1.80	8.38	−5.59	−0.67	15.68	2.16
AFQT Q4	−6.38	4.30	3.86	−4.32	1.58	3.37	−3.48
Male	−3.83	0.98	6.19	−4.61	4.25	0.90	3.60
Female	−5.90	4.02	8.14	−4.43	−2.49	10.96	1.06
Non Black	−4.69	2.70	7.46	−5.60	0.21	9.09	2.37
Black	−5.63	−2.88	9.66	−1.88	−0.12	12.43	2.18
Low Schooling	0.53	7.15	4.64	−3.40	−4.56	−0.73	9.50
High Schooling	−5.25	0.77	9.03	−5.10	0.75	12.00	1.86

Notes: This table reports the change in the distribution of fields of the economy, and how the distribution of each subset of the population changes. Entries represent percentage amounts over the share that they represent on the population. The distribution of fields is obtained as the share of individuals that have a major at each field in the last period of the model.

Table A.25: Field To Field Transition

		Business	STEM	Education	Social Sciences	Humanities	Health	Other
All								
	Business	0.86	0.03	0.02	0.03	0.04	0.01	0.01
	STEM	0.02	0.91	0.01	0.02	0.03	0.01	0.00
	Education	0.01	0.02	0.89	0.02	0.03	0.02	0.01
	Social Sciences	0.02	0.03	0.02	0.87	0.04	0.02	0.01
	Humanities	0.02	0.03	0.02	0.02	0.89	0.01	0.01
	Health	0.02	0.02	0.02	0.02	0.03	0.88	0.00
	Other	0.02	0.03	0.01	0.03	0.04	0.01	0.85
Parental Income Q1								
	Business	0.84	0.03	0.04	0.03	0.04	0.01	0.01
	STEM	0.02	0.87	0.03	0.03	0.02	0.01	0.01
	Education	0.01	0.01	0.90	0.03	0.03	0.02	0.00
	Social Sciences	0.02	0.02	0.02	0.89	0.02	0.02	0.01
	Humanities	0.02	0.01	0.03	0.04	0.88	0.01	0.01
	Health	0.01	0.03	0.02	0.05	0.03	0.86	0.01
	Other	0.01	0.01	0.01	0.04	0.04	0.01	0.89

Notes: This table reports the transition patterns from individuals that change fields in the counterfactual simulation. I report both the overall change in the economy and the behavior of individuals from the bottom of the parental income distribution. I focus only on individuals that graduate in both scenarios, the baseline and the counterfactual.

Table A.26: Non-Consumption Change

	Field Leaving			Field Arriving		
	Working Hours			Working Hours		
	Same	Less	More	Same	Less	More
Business	0.73	0.89	0.58	0.59	0.67	0.28
STEM	0.76	0.91	0.50	0.54	0.59	0.37
Education	0.66	0.81	0.49	0.60	0.77	0.44
Social	0.65	0.88	0.41	0.67	0.75	0.42
Humanities	0.68	0.88	0.38	0.63	0.75	0.42
Health	0.65	0.75	0.41	0.63	0.79	0.40
Other	0.25	0.35	0.12	0.88	0.92	0.76

Notes: This table represents the percentage of individuals that experience an increase in non-consumption utility when changing their field of study decision with respect to the baseline. I document this change also considering whether the individual changed it's labor supply decision when changing field of study. The results are reported both for the field that were leaving, and for the field that are arriving.

Table A.27: Changes Across Fields by Demographics

	Business	STEM	Education	Social Sciences	Humanities	Health	Other
Male, Non-Black	−4.36	1.53	6.21	−5.67	4.83	−2.55	4.02
Male, Black	−1.18	−0.97	3.68	−0.89	0.26	20.97	0.76
Female, Non-Black	−4.94	7.12	7.33	−5.64	−2.76	10.76	0.75
Female, Black	−9.38	−5.90	11.31	−1.94	−0.38	12.32	2.64

Notes: This table reports the transition patterns from individuals that change fields in the counterfactual simulation, as in Table A.25. Here I focus in demographic characteristics that drive preferences for fields.

B Detailed Estimation Algorithm

B.1 Summary of the Algorithm

In this section I will describe in detail the estimation algorithm used in the paper to obtain all the parameters of the model. For simplicity, I will break the explanation into the first iteration of the algorithm, and then how I use the Nested Fixed Point Algorithm described in Aguirregabiria and Mira, (2002) to improve the precision of my estimates. It is important to mention that each of the iterations of the NFXP algorithm are consistent estimates of the structural parameters of the model, so the algorithm can be stoped at any iteration. In my case I repeat until convergence to make sure I have the best model fit.

I will now provide a description of the procedure, and in the following sections I will describe in detail each step. The algorithm is:

1. Estimate the posterior unobserved types probabilities using an auxiliary model expanded with measures, this will also provide the estimates for the CCPs.

2. Using the posterior probabilities and the CCPs, focus on the M-step of the full model estimation. Here the procedure is to estimate the model sequentially :

   (a) First obtain the parameters of the wage, grants, parental transfers and graduation functions.

   (b) Then estimate the budget shock distribution.

   (c) Finally, estimate the parameters of the flow-payoff using the CCPs as control functions.

   (d) At this point I will re-estimate the CCPs using the model estimates, and repeat the process from (b) until convergence in the parameters.

B.2 The Expected-Maximization Algorithm

As described in Section 4, the likelihood of the observed data follows a finite-mixture model and the estimation will be performed using the Expected-Maximization algorithm, which breaks the problem in an iterative two-step algorithm.[23] The algorithm is built under the fact that Dempster et al., (1977) realized that the first order conditions of the maximization problem of Equation B.1 can be also obtained from an equivalent problem expressed in Equation B.4, in which the population unobserved type probability are substituted by the posterior distribution of types.

$$(\hat{\pi}_k, \hat{\theta}) = \arg\max_{\pi,\theta} \ln \left(\sum_{k=1}^{K} \pi_{ik} L_{dit}(k) L_{woit}(k) L_{geit}(k) L_{peit}(k) L_{Geit} L_{bit}(k) \right) \tag{B.1}$$

[23]See Arcidiacono and Ellickson, (2011) for a detailed explanation of the usage of the EM-algorithm in this type of situations.

Under this set up, the EM-algorithm iterates between an expectation step (E-step) and a maximization step (M-step) to update both the type distributions and the model parameters, obtaining the desired estimates upon convergence.

Expectation Step

For any $m-th$ iteration of $\theta^{(m)}$ and $\pi_k^{(m)}$ this step is in charge of updating π_k^{m+1} as follows. First, using the current values the posterior distribution of unobserved types $q_i(k)^{(m+1)}$ is updated using :

$$q_i(k|\theta^{(m)}, \pi^{(m)}) = \frac{\pi_k l_i(k)}{\sum_{k=1}^{K} \pi_k l_i(k)} \tag{B.2}$$

Now taking this as given, the new guess of the distribution of types becomes:

$$\pi_k^{(m+1)} = \frac{1}{N} \sum_{i=1}^{N} q_k(k)^{(m+1)} \tag{B.3}$$

Maximization Step

The maximization step is in charge of updating $\theta^{(k+1)}$ taking as given the current value of $\pi_r^{(k+1)}$. To do so, this step builds on the fact that Dempster et al., (1977) realized that there is an equivalent maximization problem that provides the same first order conditions as the finite mixture that we are trying to optimize.

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{t=1}^{T} q_i(k|\theta, \hat{\pi}) \ln \left[L_{dit}(k) L_{woit}(k) L_{geit}(k) L_{peit}(k) L_{Geit} L_{bit}(k) \right] \tag{B.4}$$

The new problem is expressed in terms of the posterior distribution of the unobserved type variable. The benefits of this alternative representation of the problem is that , as noted in **Arcidiacono and Jones**, additive separability is re-introduced in this step, which implies that the maximization can be now done sequentially or by parts.

Unfortunately, as mention in Section 4 the computational complexity of performing the EM-Algorithm in the full model is high, and for this reason and following Arcidiacono et al., (2024) I will estimate the posterior distribution of the unobserved types with an auxiliary model and use those estimates as weights in the maximization step of the full-model estimation.

B.3 The Auxiliary Model

In this section, I describe the auxiliary model used to estimate the posterior distribution of unobserved types. The key idea behind the model is to leverage the same variation in the data as the full model, but with significantly reduced computational complexity. This approach allows for the estimation of the posterior distribution of the unobserved types without the high computational cost associated with the full model. Additionally, this method offers the advantage of providing estimates for the conditional choice probabilities (CCPs), which are crucial for the main model estimation.

The auxiliary model consists of two primary components: (1) a simplified static discrete choice model and (2) measures of school preference. I will now detail each component before presenting the full likelihood function of the model.

B.3.1 The Static Likelihood

The static component of the auxiliary model can be thought of as a myopic, simplified version of the full structural model. The choice set remains the same as in the full model, but the payoffs for each alternative no longer have a structural interpretation. Instead, the goal is to estimate the choice probabilities as accurately as possible, since these estimates will later serve as the CCPs in the full model estimation. To simplify estimation and make it computationally feasible I will not include continuation values in the auxiliary model. However, I will still incorporate consumption flows. In that sense, the choice probabilities of each alternatives will be affected by the expected consumption at each of the alternatives, denoted $\hat{C}_{ieolft}$, where the expectation is with respect to wages, grants and parental transfers, without considering debt decisions. Equation B.5 characterizes expected consumption, where expected wage, grants and parental transfers are computed as explained in Appendix C. Finally, to account for different choice probabilities based on the current debt status, I will include the current student debt effect on the flow payoff.

$$\hat{C}_{ieloft} = \hat{W}_{iot} + \hat{G}_{iet} + \hat{P}_{iet} - \tau(e) \tag{B.5}$$

The payoff for each alternative is modeled as:

$$\tilde{v}_{eloft}(Z_{it}, b_{it}) = \tilde{u}_{elof}(Z_{it}) + \beta_{1elo}\hat{C}_{ieloft} + \beta_{2elof}b_{it} + \epsilon_{ieloft} \tag{B.6}$$

The likelihood of choices for individual i will then be:

$$\tilde{L}^{a}_{dit}(k) = \prod_{e}^{E}\prod_{l}^{L}\prod_{o}^{O(e)}\prod_{f}^{F(e)} \tilde{P}_{ieolft}(Z_{it}, b_{it}, k)^{d_{ieloft}} \tag{B.7}$$

And the total likelihood of the static model is:

$$\tilde{l}_i(k) = \tilde{L}_{dit}(k)\tilde{L}_{woit}(k)\tilde{L}_{geit}(k)\tilde{L}_{peit}(k) \tag{B.8}$$

where the likelihood of wages, grants and parental transfers are defined as in the full model.

B.3.2 Measures of Schooling Preferences

To improve identification of the model, I will expand the model with measures that capture information about the latent schooling preference type. For those measures I will use a simplified version of

Arcidiacono et al., (2024) and I will include:

- School tardiness: An indicator for whether the individual was late to school more than three times without a valid reason.
- Summer school attendance: An indicator for whether the individual has ever taken summer classes.

Importantly, these measures are informative about the individual's latent schooling preference type but provide no information about individuals labor market ability. The likelihood of these binary outcomes is modeled using logistic regressions, which depend on X_{im} that include time-invariant individual's characteristics (parental income, afqt, sex and race):

$$L_{it}^{m} = P(measure = 1|X_{im}, k) \tag{B.9}$$

The main difference with this likelihood and the static model is that this measure is only observed at one period of time, and for this reason it does not use the panel structure of the model.

B.3.3 Auxiliary Model Likelihood

The overall likelihood function for the auxiliary model combines the static discrete choice model and the schooling preference measures. For individual i, the likelihood is:

$$L_i^a = \sum_{k=1}^{K} \pi_k \tilde{L}_{dit}(k) \tilde{L}_{woit}(k) \tilde{L}_{geit}(k) \tilde{L}_{peit}(k) L_{m=1i}(k) L_{m=2i}(k) \tag{B.10}$$

The estimation of this likelihood will be done following Section EM-agorithm, and as a result we will obtain the posterior type distributions q_{ik}, the type distribution π_k and the estimates of the CCPs $\hat{P}_{ieloft}$. Table B.1 shows the results of the measurement system, and we can see that the effect of the unobserved schooling type is negative in both measures, suggesting that this type identifies individuals that are not late to school and did not take summer classes.

B.4 CCP Representation of the Conditonal Value Function

In this section, I will explain how to express the continuation value as a function of conditional choice probabilities (CCPs) in order to generate the necessary control function. As initially demonstrated by Hotz and Miller, (1993), and subsequently utilized in many studies, the continuation value can be mapped as a function of CCPs and inverted accordingly. Specifically, under the assumption of a Type-1 Generalized Extreme Value (GEV) distribution, the mapping between the CCPs and the continuation value is:

$$V_t(Z_{it}, b_{it}) = v_{kt}(Z_{it}, b_{it}) - \log\left(P_{kt}(Z_{it}, b_{it})\right) \tag{B.11}$$

Table B.1: Estimates of the measurement system

	Arrive Late	Summer Courses
ParInc Q2	−0.11	−0.12
	(0.96)	(0.07)
ParInc Q3	−0.17	−0.22
	(0.89)	(0.07)
ParInc Q4	0.08	−0.23
	(0.67)	(0.07)
Ability Q2	−0.11	−0.22
	(0.90)	(0.08)
Ability Q3	−0.35	−0.45
	(0.89)	(0.08)
Ability Q4	−0.63	−0.83
	(0.96)	(0.16)
Female	−0.06	−0.12
	(0.96)	(0.06)
Black	0.05	0.05
	(0.66)	(0.06)
Constant	−1.25	0.10
	(0.45)	(0.09)
High Schooling Taste	−0.48	−0.31
	(0.59)	(0.06)

Notes: This table reports the estimated coefficients of the measurement system obtained performing the EM-algorithm estimation. Note that "High Schooling Taste" refers to the coefficient of the unobserved type estimated to be high schooling taste. The other unobserved type effect is normalized to zero.

where $P_{kt}(\mathcal{X}_{it}, b_{it})$ is the conditional choice probability of choosing alternative k at period t given a state space (Z_{it}, b_{it}). The intuition of this result implies that the continuation value can be expressed as the value of any of the possible alternatives (in this case k), plus a correction term that accounts for the fact that the alternative might be suboptimal (in this case $\log(P_{kt}(Z_{it}, b_{it}))$). Now, consider setting k as home production, implying the individual makes the minimum consumption level. The continuation value then becomes:

$$V_t(Z_{it}, b_{it}) = \frac{\underline{C}^{1-\sigma}}{1-\sigma} - \log(P_{kt}(Z_{it}, b_{it})) + \beta V_{t+1}(Z_{it+1}, b_{it+1}) \tag{B.12}$$

The choice of home production is intentional, as it provides several desirable properties. First, because home production is assumed to be the base category in the model estimation, the non-consumption utility u_h is normalized to zero. Moreover, since home production implies that the individual consumes at the minimum consumption level, this alternative eliminates budget uncertainty, simplifying the computational process.[24] Finally, the state space evolution under home production is straightforward, as individuals do not accumulate labor market or educational experience during this period. By substituting Equation B.11 into Equation B.12, and iterating until the final period of the model—when individuals receive the terminal continuation value—I obtain the following expression:

$$\begin{aligned} v_{eloft}(Z_{it}, b_t) &= u_{elof}(Z_{1it}) + \sum_{p=t+1}^{T} \beta^{p-t} \frac{\underline{C}^{1-\sigma}}{1-\sigma} + \\ &\mathbb{E}\left[\max_{b_{eloft+1}} \left\{ \frac{C_{eloft}^{1-\sigma}}{1-\sigma} + \sum_{p=t+1}^{T} -\beta^{p-t} \ln P_{home,p}(Z_{ip}, b_{ip}) + \beta^{T-t} V_{T+1}(Z_{T+1}, b_{T+1}) \right\}\right] \end{aligned} \tag{B.13}$$

This expression illustrates that the conditional value function for alternative (e, l, o, f) can be decomposed into the non-consumption effect of the flow payoff, captured by $u_{elof}(Z_{1it})$, and a term that captures the expected utility of consumption under this alternative. This term accounts for the optimal debt level after the budget realization, along with a discounted sum of minimum utility payoffs, adjusted by the probabilities that home production is chosen in each periods until the last, and finally, the terminal continuation value implied by this sequence of choices[25].

Under this expression, it is then clear to see that the difference of a conditional value function with respect to the base category (home production) is:

$$\begin{aligned} &v_{eloft}(Z_{it}, b_t) - v_{ht}(Z_{it}, b_t) = u_{elof}(Z_{1it}) + \\ &\mathbb{E}\left[\max_{b_{eloft+1}} \left\{ \frac{C_{eloft}^{1-\sigma}}{1-\sigma} + \sum_{p=t+1}^{T} -\beta^{p-t} \ln P_{home,p}(x_p, b_p) + \beta^{T-t} V_{T+1}(Z_{iT+1}, b_{iT+1}) \right\}\right] \\ &- \left[\frac{\underline{C}^{1-\sigma}}{1-\sigma} + \sum_{p=t+1}^{T} -\beta^{p-t} \ln P_{home,p}(x_p, b_p) + \beta^{T-t} V_{T+1}(Z_{iT+1}, b_{iT+1})\right] \end{aligned} \tag{B.14}$$

Next, I define $\delta(Z_{it}, b_t, d_{it})$ as the component of the conditional value function capturing the difference in future sequences of conditional choice probabilities and terminal values:

$$\begin{aligned} \delta(Z_{it}, b_t, d_{it}) = &\mathbb{E}\left[\max_{b_{ieloft+1}} \left\{ \frac{C_{ieloft}^{1-\sigma}}{1-\sigma} + \sum_{p=t+1}^{T} -\beta^{p-t} \ln P_{home,p}(x_p, b_p) + \beta^{T-t} V_{T+1}(Z_{iT+1}, b_{iT+1}) \right\}\right] \\ &- \left[\frac{\underline{C}^{1-\sigma}}{1-\sigma} + \sum_{p=t+1}^{T} -\beta^{p-t} \ln P_{home,p}(x_p, b_p) + \beta^{T-t} V_{T+1}(Z_{iT+1}, b_{iT+1})\right] \end{aligned} \tag{B.15}$$

To finally get:

$$v_{eloft}(Z_{it}, b_{it}) - v_{ht}(Z_{it}, b_{it}) = u_{elof}(Z_{1it}) + \delta(Z_{it}, b_{it}, d_{it}) \tag{B.16}$$

[24] Specifically, this allows for direct insertion of the minimum consumption level, instead of solving for the optimal consumption at different points of a Gaussian quadrature.

[25] Notice that here we could be done, since expression includes the parameters that we want to estimate, $u_{elof}(Z_{1it})$, and a term that controls for all the future sequence of utilities. However, since identification is only achieved with respect to a base category, I will show the expression in difference.

As mentioned in Section 4, the importance of this expression is the fact that the difference in continuation values can be expressed as a linear term of the parameters that want to be estimated plus a correction term controlling for the difference in future continuation values. This will allow for a simple multinomial logit estimation with a nuisance term.

B.5 Estimation of the budget shock distribution

In this section, I will outline the estimation of the budget shock distribution, conditional on the parameters obtained in the earlier stages of the estimation process. As previously discussed in Section 4, identification relies on variations in student debt decisions. However, two key challenges must be addressed: (1) the student debt decision lacks a closed-form expression in relation to the budget shock, and (2) the empirical distribution of student debt is censored at zero, as individuals with sufficiently large budgets do not take on debt due to the model's assumption that savings are not permitted. According to the model, student debt decisions will be made based on the following expression (for simplicity I remove the dependence on the alternative (e, l, o, f)):

$$b^*_{it+1} = \arg\max_{b_{it+1}} \left(\frac{(C(Z_{it}, b_{it+1}, \xi_{it}))^{1-\sigma}}{1-\sigma} + \beta V_{t+1}(Z_{it+1}, b_{it+1}) \right)$$

Where $C(Z_{it}, b_{it+1}, \xi_{it})$ just emphasizes that consumption depends on the desired debt level. Using the CCPs and in the same spirit as the previous section, can be expressed as:

$$\begin{aligned} b^*_{it+1} = \arg\max_{b_{it+1}} & \frac{(C(Z_{it}, b_{it+1}, \xi_{it}))^{1-\sigma}}{1-\sigma} + \\ & \sum_{p=t+1}^{T} -\beta^{p-t} \ln P_{home,p}(x_p, b_p) + \beta^{T-t} V_{T+1}(Z_{iT+1}, b_{iT+1}) \end{aligned}$$

The use of CCPs guarantees that the estimation of the budget shock provides consistent estimates for the distribution even withouht the flow-payofs parameters, since the parameters of the flow-payoff do not enter the expression.

Given the absence of a closed-form expression for student debt as a function of the budget shock, likelihood-based models, including Simulated Likelihood methods, are not possible. Therefore, estimation will be conducted using the Simulated Method of Moments (SMM).[26] Let $m(b^*_{it})$ represent the vector of empirical moment conditions, which in this case are the average student debt among indebted individuals and the share of non-indebted individuals. I will simulate budget shocks $\xi_{it}(\mu, \sigma^2)$ for the individuals in the sample and obtain their simulated student debt decisions $b^*_{it+1}(\xi_{it})$. I will repeat this process S times to obtain S different samples of simulated student loans decisions, fixing the distribution parameters μ, σ^2. The idea is then to minimize the weighted distance in moments to find the parameters that better explain the data. However, to standardize the units of the different moments, I will use as objective function the percentage distance in moments:

[26] See Chapter 4 of Adda and Cooper, (2003) for a detailed description of this method.

$$(\hat{\mu}, \hat{\sigma}^2_{\xi}) = \arg\min_{\mu, \sigma^2} \left[\frac{m(b^*_{it}) - \frac{1}{S}\sum_{s=1}^{S} m(b^*_{it+1}(\xi_{it}(\mu, \sigma^2)))}{m(b^*_{it})} \right]' W^{-1}$$
$$\left[\frac{m(b^*_{it}) - \frac{1}{S}\sum_{s=1}^{S} m(b^*_{it+1}(\xi_{it}(\mu, \sigma^2)))}{m(b^*_{it})} \right]$$

In the previous expression W^{-1} is the identity matrix. Finally, in order to capture all the budget heterogeneity in my model, I will estimate a different distribution for each parental income and ability cell.

B.6 Nested Pseudo-Likelihood Expanded Algorithm

In this section I will describe how I finally build the estimation algorithm putting together the different pieces obtained from the different sections. In particular, I will use a Nested Pseudo-Likelihood algorithm from Aguirregabiria and Mira, (2002) in which the CCPs are re-updated with model predictions until convergence is found. The procedure is:

1. Estimate the posterior distribution of the unobserved schooling preference variable and the CCPs.
2. Estimate the parameters of the production functions (wages,grants, parental transfers and graduation probabilities).
3. Estimate the flow-payoff and budget shock distribution parameters using the empirically estimated CCPs.
4. Using the obtained parameters, re-estimate the CCPs with the model and go back to step (3) until convergence.

C Comments About the Expected Budget

In this section I describe in detail the creation of the consumption variable and how I compute the expected budget. The main complication when dealing with consumption is that it is unknown by the agent at the moment of making the discrete choice decision. As described during Section 3, this implies that the agent will have to form expectations about the possible consumption values for each alternative before making the discrete decision. This introduces a layer of computational complexity, since for each discrete alternative I will have to compute expectations with respect to consumption. Not only this, but because for each possible budget state there are many possible debt scenarios, I also need to evaluate this expectation for any possible choice of student debt to find which debt is the optimal. Given that consumption depend on wages, grants, parental transfers and the budget shock, only one expectation implies several integrals that can easily made the model unfeasible to be estimated. The computational complexity is much higher since for each discrete alternative there is an associated student debt choice that considers the future repayment. For this reason, the expectation of consumption today is not only with respect to the financial resources that would be available at the current period, but also for all the future sequence of financial resources that would be available at any possible future choice, since this will determine the student debt decision. It is then clear that this introduces a challenge.

To address this issue and simplify the computational complexity of the model I estimate all the expectations using Gaussian quadratures, and in particular Gauss-Hermite quadrature of degree five, since they are specially convinient for integrals of functions of normal distributions, which is the case here. However, even with the computational reduction experienced by approximating the integrals, the complexity is still very high because I need to evaluate consumption for each possible debt decision, not only today but also in the future.

To further reduce the complexity, I will break the problem in the choices that imply an education decision and those that don't. During non-education choices the complexity is much lower, since the expectation is only with respect to wages, and this is feasible to compute. During education spells the layers of uncertainty increase, since individuals will now face uncertainty at the wage, grants and parental transfers levels. This would imply in some cases I need to compute the expectation with respect to wages, grants probabilities, expected grants, parental transfers probabilities and expected parental transfers as well as the budget shock. Although this would technically be feasible, with my current resources it would imply that the model takes several months to be estimated. For this reason, and as a temporary correction, I will load all the financial uncertainty of education choices to the budget shock. Because the budget shock estimation will be estimated with student debt choices, it can be used to capture all the residual uncertainty departing from expected financial resources. I will now describe what this means.

In the estimation of the model, I will modify the budget constraint of educational choices, so that the agents get the expected value of each financial source and the only uncertainty that remains is at the budget shock level. This implies that the budget constraint for education choices now is:

$$C_{ieloft} = \begin{cases} \hat{W}_{ilot} + \hat{G}_{iet} + \hat{P}_{iet} - \tau(e) + b_{elofit+1} - (1+r)b_{it} + \xi_{ieloft}, & \text{if } e \in \{2y, 4y\}, l \neq 0 \\ \hat{G}_{iet} + \hat{P}_{iet} - \tau(e) + b_{elofit+1} - (1+r)b_{it} + \xi_{ieloft}, & \text{if } e \in \{2y, 4y\}, l = 0 \\ \hat{W}_{ilot} + \hat{G}_{iet} - \tau(e) + b_{elofit+1} - (1+r)b_{it} + \xi_{ieloft}, & \text{if } e = G, l \neq 0 \\ \hat{G}_{iet} - \tau(e) + b_{elofit+1} - (1+r)b_{it} + \xi_{ieloft}, & \text{if } e = G, l = 0 \end{cases} \tag{C.1}$$

with:

$$\hat{W}_{oieat} = \exp\left(X_{oit}\hat{\gamma} + \frac{\hat{\sigma}_o^2}{2}\right)$$

$$\hat{G}_{eit} = \frac{\exp(X_{git}\hat{\Gamma}_{ge})}{1 + \exp(X_{git}\hat{\Gamma}_{ge})} \times \exp\left(X_{git}\hat{\lambda}_{ge} + \frac{\hat{\sigma}_{ge}^2}{2}\right)$$

$$\hat{P}_{eit} = \frac{\exp(X_{pit}\hat{\Gamma}_{pe})}{1 + \exp(X_{pit}\hat{\Gamma}_{pe})} \times \exp\left(X_{pit}\hat{\lambda}_{pe} + \frac{\hat{\sigma}_{pe}^2}{2}\right)$$

There are several gains from doing this. First of all, and most importantly, the uncertainty about consumption will now only depend on the budget shock, which will capture deviations from the sum of parental transfers, grants and wages, which sometimes crowd-out each other. Since I will identify the budget shock distribution from residual variation after the expected transfers,the uncertainty distribution will be as close as possible to the real one, but with just one integral.

D State Space and Choice Set

In this section I describe in detail the state space of the model, together with the possible choices that can be done at different stages.

D.1 State Space

In Table D.1 I report the different variables that compose the state space of the model. In Table D.2 I report the total amount of possible state spaces in the model, without considering student debt discrete grid, which would multiply the total by $\times 100$. In total there are 398,656,000 state space points considering student debt. To simplify once graduation is obtained, years of education are not tracked anymore since it has no longer an effect once graduation is achieved.

Table D.1: Time-Invariant State Space Composition

Time Invariant Characteristics	
Parental Income	4 Quartiles
AFQT Score	4 Quartiles
Female	True/False
Black	True/False
Latent Schooling Type	Low/High
Total States	$4 \times 4 \times 2 \times 2 \times 2 = 128$
Endogenous Characteristics	
Labor Market Experience	Maximum 10 years
Two-Year School Experience	Maximum 10 years
Four-Year School Experience	Maximum 10 years
Grad-School Experience	Maximum 6 years
Associate Degree	True/False
Bachelor Degree	True/False
Graduate Degree	True/False
Last School Period	Maximum 9
Field of Graduation	Maximum 8
Last Sector Choice	Maximum 10
Student Debt	Discrete Grid

Table D.2: State Space per Period

	Endogenous States	Total States	Cummulative Total States
Period 1	1	128	128
Period 2	18	2,304	2,432
Period 3	73	9,344	11,776
Period 4	233	29,824	41,600
Period 5	590	75,520	117,120
Period 6	1294	165,632	282,752
Period 7	2,589	331,392	614,144
Period 8	4,785	612,480	1,226,624
Period 9	8,237	1,054,336	2,280,960
Period 10	13,325	1,705,600	3,986,560

D.2 Choice Set

At every period t individuals can choose among several education, occupation and labor market participation possibilities. In Table D.3 I report the total possible amount of choices of the model. However, individuals of the model will never face so many choices, since the actually choice set will depend on the state space.

Table D.3: Total Model Choices

	Number of Choices
Education Choices	
Two-Year Institutions: 3 Labor Supply Decisions	3
Four-Year Institutions: 3 Labor Supply Decisions × 8 Fields of Study	24
Graduate School: 3 Labor Supply Decisions	3
Non-Education Choices	
Work Choices: 2 Labor Supply Decisions × 8 Occupations	16
Non-Work	1
Total Possible Choices	47

For educational choices the restrictions are: (1) graduate school can only be chosen if a Bachelor degree has been obtained, (2) graduated individuals will not be able to choose education choices anymore, (3) if an individual has more than two years of four-year school, undeclared field is not an option anymore.

For occupational choices the restrictions are a bit more complex. I will assume that depending on the field of graduation some occupations will not be possible choices. The reason is that for some fields I observe very few individuals choosing specific occupations, so the identification will be complicated.

This means that if an individual is not graduated, all occupations are within the choice set, but the moment an individual graduates from either a two-year school or a four-year school, some restrictions will be implemented. To apply the restrictions I will use two reasons: (1) the empirical frequences of occupations chosen by fields of graduation, reported in Table D.4, (2) the distribution of fields within an occupation, reported in Table D.5. The idea is that if a field rarely chooses an occupation, then this occupation will not be within the choice set of individuals graduated from that field, unless the field represents a big chunk of the people within that occupation. The final distribution of possible choices for each field of graduation is reported in Table D.6.

Table D.4: Distribution of Occupations Chosen by Human Capital Cells

Human Capital	Occupations							
	Business	STEM	Social Sciences	Education	Humanities	Health	Sales & Office	Production
No Degree	5.47	1.42	1.06	1.90	1.25	6.05	30.16	52.69
Business	40.38	4.28	2.55	3.21	3.95	0.82	36.27	8.55
STEM	17.63	41.02	2.66	5.54	3.22	8.09	12.42	9.42
Education	6.17	0.22	3.52	73.79	0.88	1.32	7.49	6.61
Social Sciences	21.43	3.81	24.57	9.52	3.81	3.05	20.95	12.86
Humanities	17.63	3.10	5.42	13.41	12.30	1.89	30.44	15.82
Health	7.94	1.27	4.13	10.16	0.00	57.46	9.84	9.21
Other Major	20.21	7.32	16.72	3.83	2.79	4.53	22.30	22.30
Associate Degree	8.34	7.17	1.54	3.93	3.51	16.73	26.02	32.77

Table D.5: Distribution of Fields by Occupations

	Human Capital								
Occupations	No Degree	Business	STEM	Education	Social Sciences	Humanities	Health	Other Major	Associate
Business	49.48	18.40	5.96	1.05	8.43	7.68	0.94	2.17	5.88
STEM	34.17	5.19	36.96	0.10	4.00	3.60	0.40	2.10	13.49
Social Sciences	34.60	4.21	3.26	2.17	35.01	8.55	1.76	6.51	3.93
Education	36.54	3.11	3.98	26.67	7.96	12.42	2.55	0.88	5.89
Humanities	47.10	7.51	4.54	0.63	6.26	22.38	0.00	1.25	10.33
Health	69.11	0.47	3.46	0.28	1.52	1.04	8.57	0.62	14.92
Sales & Office	80.64	4.89	1.24	0.38	2.44	3.92	0.34	0.71	5.43
Production	91.06	0.75	0.61	0.21	0.97	1.32	0.21	0.46	4.42

Table D.6: Occupations That Can be Chosen for Each Human Capital Cell

	Occupations							
Human Capital	Business	STEM	Social Sciences	Education	Humanities	Health	Sales & Office	Production
No Degree	✓	✓	✓	✓	✓	✓	✓	✓
Business	✓	✗	✗	✗	✗	✗	✓	✓
STEM	✓	✓	✗	✗	✗	✓	✓	✓
Education	✓	✗	✗	✓	✗	✗	✓	✓
Social Sciences	✓	✗	✓	✓	✗	✗	✓	✓
Humanities	✓	✗	✓	✓	✓	✗	✓	✓
Health	✓	✗	✗	✓	✗	✓	✓	✓
Other Fields	✓	✓	✓	✗	✗	✗	✓	✓
Associate Degree	✓	✓	✗	✗	✓	✓	✓	✓

E Occupations and Fields Classification

To classify occupations I have used the 2002 Census Occupation Classification and to classify fields of study I have used the Census detailed classification code use at the American Community Survey. Table E.1 reports the main occupations within each occupation category and Table E.2 report the main fields within each field of study possibility.

Table E.1: Main Occupations

Occupation	Frequency
Business	
Managers, All Other	0.20
Accountants and Auditors	0.10
Chief Executives	0.06
Financial Managers	0.06
Legislators	0.05
STEM	
Computer Software Engineers	0.17
Computer Support Specialists	0.09
Engineers, All Other	0.08
Engineering Technicians, Except Drafters	0.07
Computer Programmers	0.07
Social Sciences	
Lawyers	0.28
Social Workers	0.26
Counselors	0.19
Clergy	0.12
Psychologists	0.05
Education	
Elementary and Middle School Teachers	0.38
Postsecondary Teachers	0.15
Teacher Assistants	0.13
Other Teachers and Instructors	0.10
Secondary School Teachers	0.09
Humanities & Arts	
Designers	0.27
News Analysts, Reporters and Correspondents	0.11
Writers and Authors	0.08
Artists and Related Workers	0.08
Musicians, Singers, and Related Workers	0.08

Health	
Nursing, Psychiatric, and Home Health Aides	0.29
Physicians and Surgeons	0.12
Licensed Practical and Licensed Vocational Nurses	0.10
Diagnostic Related Technologists and Technicians	0.05
Clinical Laboratory Technologists and Technicians	0.04
Sales & Office	
First-Line Supervisors/Managers of Retail Sales Workers	0.11
Secretaries and Administrative Assistants	0.10
Cashiers	0.10
Retail Salespersons	0.10
Customer Service Representatives	0.07
Production	
Driver/Sales Workers and Truck Drivers	0.06
Janitors and Building Cleaners	0.05
Chefs and Head Cooks	0.05
Laborers and Freight, Stock, and Material Movers, Hand	0.04
Motion Picture Projectionists	0.04

Note: This table reports the main occupations at the five-digit level for each of the model possible occupation categories. This table was constructed using data from the ACS and computing the relative frequency of each five-digit occupation within each model occupation.

Table E.2: Main Fields

Occupation	Frequency
Business	
Business Management and Administration	0.31
General Business	0.20
Accounting	0.20
Marketing and Marketing Research	0.10
Finance	0.10
STEM	
Biology	0.14
Electrical Engineering	0.08
Computer Science	0.08
Mathematical Engineering	0.07
Mathematics	0.06
Education	
Elementary Education	0.30

General Education	0.26
Physical and Health Education Teaching	0.06
Art and Music Education	0.05
Secondary Teacher Education	0.05
Social Sciences	
Psychology	0.33
Political Science and Government	0.18
Economics	0.16
Sociology	0.12
Family and Consumer Sciences	0.06
Humanities & Arts	
English Language and Literature	0.17
History	0.12
Communications	0.12
Liberal Arts	0.07
Fine Arts	0.06
Health	
Nursing	0.26
Treatment Therapy Professions	0.14
Pharmacy, Pharmaceutical Sciences, and Administration	0.11
Communication Disorders Science and Services	0.09
Medical Technologies Technicians	0.07
Other	
Criminal Justice and Fire Protection	0.28
Physical Fitness, Parks, Recreation and Leisure	0.16
Animal Sciences	0.05
General Agriculture	0.05
Agriculture Production and Management	0.04

Note: This table reports the main fields at the detailed code level for each of the model possible fields of study categories. This table was constructed using data from the ACS and computing the relative frequency of each detailed code field within each model field of study, without considering undeclared field.

F Data Appendix

In this part of the paper I will discuss the data cleaning and sample selection performed for the final data. To facilitate the discussion, I will divide into how choices are defined and how the variables are created.

F.1 Choice Definition

In the model, choices are a combination of an occupation, field of study, education and labor supply decision.

In terms of employment, I will consider the labor supply decision as the most common labor supply status within the academic year, without considering December of the summer period. If the most common labor supply decision was less than 10 hours, the individual is considered not to work, if it is between 10 and 35 the individual is considered to work part-time, and if it is more than 35 the individual is considered to work full-time.

Once employment is obtained, occupation is coded as the most common occupation within that period in terms of hours worked at that occupation.

In terms of education, I will classify enrollment decisions following the choices of the different terms being reported. If an individual reports a enrollment spell during an academic year, the individual is assumed to be enrolled. Based on the college of choice I will report if it is a two-year or a four-year school. Some individuals report both during the same academic year, in that case I will consider the first one being reported.

Similarly, I will classify majors as the major reported during a particular academic year. If the individual reports more than one major, I will assume the choice is the first one being reported.

If an individual is coded as enrolled during an academic year, I will assume that if working the occupation is coded as enrolled, which means it is irrelevant.

Individuals classified as not working and not enrolled will be assumed to be engaged in home production.

I will now proceed to explain how I obtain the different information from the raw data.

F.2 Data Cleaning

I will divide the data into demographics, labor market and education variables.

Demographics

The demographic information used in the paper is sex, ethnicity, parental income and AFQT. For the creation of parental income, I collect self reported parental income by the individual as well as information reported by parents in early rounds. Once I have a notion of parental income, I classify individuals in the four different quartiles of the parental income distribution. The same is done to AFQT.

Wages, Occupation and Employment

To create this variables I have used data from the employment section of the NLSY data. This section has a rooster structure which means that at each survey round the respondent reports all the new jobs she has had since the date of the last interview. For each job the respondent has had there is information in the hourly pay rate and hourly monetary compensation, the average hours per week worked, and the occupation code of the job among other things. Furthermore, the respondents report the starting date and the stop date of each of the jobs, which allows me to track all employment spells. The objective is to obtain labor force participation and wage at the academic year level. To do so, I have created a monthly panel with all the labor market information of the individuals, that I will later aggregate at the academic year level.

To measure hourly pay rate and hourly monetary compensation I use NLSY created variables that are based on multiple questions that try to capture all the possible sources of income such as hourly compensation but also non-wage payments like commissions, bonuses, tips or others.[27]

To obtain the occupation code of each period, where a period is a month, I use the occupation code of the job where the respondent worked the majority of the time, since it might be that an individual works at multiple jobs at the same time.

To construct the data some decisions were made. First of all, I dropped all observations for which job starting date or stopping date was not available, since it was not possible to match this wage data into a specific time period. Also, whenever an individual changed job at a specific month, I am using the old job information to identify that time period. In periods in which individuals are working in more than one firm at the same time, I am summing the weekly hours of both to obtain the total amount of hours worked, and I am computing the weighted average of the hourly wage and hourly compensation using hours worked as weights.

Unfortunately, before I can aggregate data at the academic year level I need to deal with some missing information that I will input to maximize sample size. Table F.1 reports the number and share of missing observations for each relevant variable. As we can see, the level is not worrisome, and the variable with more missing information is hourly wage with 6,59% missing information. I will now describe the process of data imputation, using employment status as a reference variable.

Employment Status. This is a created variable by the NLSY archivist that reports at the weekly level the employment status of the individuals.

Occupation. Occupation is reported for each job. In total 0.93% of all spells have missing information, that I will input using the mode of the current academic year, the previous or next job spell

[27]To learn about how this variables where created visit here https://nlsinfo.org/content/cohorts/nlsy97/topical-guide/employment/wages

Table F.1: Missing Data Employment Variables

	Number	Share
Total Person-Month Observations	1,391,690	100
Missing Observations per Variables		
Hourly Wage	91,087	6.59
Hourly Pay	52,614	3.81
Occupation	12,815	0.93
Weekly Hours	54,771	3.96

Notes: This table reports the total amount of person-month observations only focusing on months with labor supply participation, before imposing the final sample restrictions.

information, or code them as missing.

Hourly Wage. For individuals with missing hourly wage I will proceed as follows, by order of priority: (1) input hourly pay, (2) input the wage from the same occupation and year, (3) input the average wage in the same academic year, (4) input the average wage in the same calendar year, (4) input the most recent or future wage, depending on which is closer in calendar time.

Enrollment, Major and Financial Information

To construct the different variables regarding enrollment, major and financial information I have relied on two different sources from the NLSY. To create the enrollment variables I am using data from the monthly event study variable section of the NLSY97. [28] This section is created by the archivist at NLSY and provides a simplified version of NLSY raw data. In particular, it contains month-by-month information of college and graduate school experience which allows to track at a monthly basis the enrollment status of the individuals in the sample.

Unfortunately, not all college information data is available in this cleaned version of the NLSY archive. In order to recover relevant financial variables and major information I have used college part of the schooling rooster. In this part of the questionnaire students are asked at each interview retrospectively about all schools attended since the date of last interview.[29] This will have two important implications. The first one is that if individuals missed some interview dates, we will still recover the full path of college enrollment. The second important implication is that the different college spells being reported at a specific point in time do not correspond with the college dates at which the individual was actually enrolled. The data of the college experience section of the schooling rooster is

[28] Visit here for more information.

[29] Visit the following sice for more info: https://www.nlsinfo.org/content/cohorts/nlsy97/topical-guide/education/education-training-achievement-scores-introduction

reported at the term level, which means that the NLSY inverviewer will first identify all the different college-terms that the individuals have attended since the date of last interview and will then loop over them to recover all relevant information. It is important to notice that the criteria for a term to be reported in the term rooster section changes over time, so I have homogenize it. Also, this only applies to undergraduates, graduate school is reported at the year level without the term structure.

Since the data is reported at the term level for undergraduates, and at the year level for graduate school, I have created a panel at the term level to later aggregate it at the academic year level. Notice that because some of the information is being reported retrospectively with respect to the past, some variables have missing at a non-negligible rate. Table F.2 reports the distribution of information across college-term variables. As we can see the variables with more missing information are the ones with respect to family financial help. I will now report the process of creation of each of the different variables and how I have imputed the values whenever it was possible to maximize sample size.

Academic Year. This variable is created using the enrollment dates for each term being reported by the individual (variable YSCH-20400). Whenever this variable was missing I used the enrollment dates reported in the NEWSCHOOL rooster (NEWSCHOOL-START). I have defined an academic year as the period between September of a year and August of the next year. For example, an individual reporting an spell that starts in March 2005 will be assumed to belong to the 2004 academic year. Unfortunately, there are 21 observations missing, which means those are spells that I can't match the actual academic year. I have first imputed as academic year the academic year if some other terms are being reported during the same round, in the absence of this information, I have reported as academic year the year at which the interview was conducted.

Major. The major variable is obtained from "First Major" variable (YSCH-21300). To understand the major imputation process it is important to know how I will aggregate academic years. For each academic year I observe different term spells, and it is possible that the individuals report different majors across terms. In the event that all majors are the same across terms within the same academic year, it is then obvious that this will correspond to the major of that academic year. However, when multiple majors are reported within the same academic year, I need to assume which is the major of that corresponding year. Since in my model I am modeling major preferences year-by-year, I want to understand what made an individual choose a specific major given the situation she was at the begining of the academic year, for this reason I have decided that I will take as major the first major the individual was reporting during this specific academic year. For this reason, if within an academic year I observe the fist major and not others, I will take that as the corresponding major for that academic year. In the same sense that since I will aggregate majors by academic years, as long as I observe one major within that year, I can assume that this was the major the individual was attending. Now, for individuals that I do not observe a major within an academic year, I will input the next observable major. If this is not available because it is the last spell of the individual, I will input the previous major. This will only leave 15 observations without major (which correspond to 3 individuals) that I will for now code them as group "Other".

Table F.2: Missing Data of College Variables

	Number	Share
Total Person-Term Observations	49,988	100
Observations Missing per Variable		
Academic Year	21	0.04
Major	634	1.27
College Type	696	1.39
College Level	1	0.00
Any Loans	373	0.75
Amount Loans	2,339	4.68
Any Grants	373	0.75
Amount Grants	3,293	6.59
Any Employer Assistance	386	0.77
Amount Employer Assistance	591	1.18
Any Work Study	386	0.77
Amount Work Study	680	1.36
Any Other Help	373	0.75
Amount Other Help	740	1.48
Family Help	5,777	11.56
Family Loans	5,725	11.45

Notes: This table reports the total amount of person-term observations before imposing all the final sample restrictions. The focus is just on college spells, so it does not consider working spells that have no college enrollment implications. The observations missing correspond to the total amount of missing observations for each of the variables, and the share that they represent over the total.

College Type. This variable refers to the ownership of the college, and it can be "public", "private non for-profit" , "private for-profit" or "missing type". The variable is created by the archivist of the NLSY and is obtained from the IPEDS data base. It refers to the variable (CV_COLLEGE_TYPE). The procedure here will be the same as with major information, I will aggregate by academic years prioritizing the first school reported. In the absence of an school during a given academic year I will input the next one, unless it is not available, in which case I will input the previous one. For individuals that have never reported a schooling type, I will input them as "missing type".

College Level. It is coded as "Four Year", Two year or Grad School. It only has one missing that disappears once I aggregate by academic year.

Student Loans, Grants and Other Financial Aid: For all the financial variables the imputation will be very similar. First, I will input the indicator variable for receiving a specific type of aid on a specific term. To do so, I will assume that if the individual received that specific type of aid during that same academic year it also received in the missing term. Notice this is innocuous since when aggregating at the academic year the variable would still be coded as one. To input the total amounts the procedure will be the following. If an individual has missing amounts for a financial aid that has never reported received, I will code as 0 the amount received. If individuals report missing information for a financial aid that has received, I will input in the following other of priority: (1) try to set the average financial aid from that source at that same specific institution (since I have college ID), (2) if that is not possible, set the individual average in that institution type ("Two-Year","Four-Year", "Grad School"). Finally, yearly student loans are top-coded to the 99% of the distribution.

Total Education Expenditure: Once I have all the sources of income used to finance a specific college term, I sum all of them to obtain an estimate of the tuition paid by the individual.

G No Loans Counterfactual

Here I report the results of introducing student loans relative to a scenario in which loans are not available. Results show that the introduction of loans is highly beneficial for low-income individuals, but has a smaller effect for high-income ones.

Table G.1: Decomposition of New Graduates

	Change (1)	Never (2)	Drop Out (3)	Two Year (4)
ParInc Q1	40.15%	5.52%	25.22%	9.41%
ParInc Q2	26.45%	2.81%	16.59%	7.04%
ParInc Q3	8.50%	0.65%	5.18%	2.67%
ParInc Q4	0.71%	0.03%	0.53%	0.14%
AFQT Q1	101.59%	23.52%	62.12%	15.94%
AFQT Q2	21.55%	2.02%	13.18%	6.34%
AFQT Q3	20.81%	0.82%	12.80%	7.19%
AFQT Q4	0.00%	0.00%	0.00%	0.00%
Male	6.04%	0.74%	4.20%	1.09%
Female	13.74%	1.18%	8.30%	4.27%
Non Black	8.32%	0.78%	5.36%	2.18%
Black	20.72%	2.31%	12.69%	5.72%

Notes: This table reports the change in graduation at four-year schools comparing a situation in which there are no loans, with one in which loans are introduced under the 10-year repayment plan. Column (1) reports the percentage increase in graduation. Column (2) reports the percentage increase coming from individuals that were not enrolled in the baseline. Column (3) reports the percentage increase coming from individuals that were drop outs in the baseline. Column (4) reports the percentage increase coming from individuals that were graduating from a two-year school in the baseline.