Financial Pressure and Career Choices

Sergi Quintana Garcia*

October 14, 2024

Abstract

This paper studies the effect of higher education financing, with a special focus on student debt repayment plan, on the human capital and the field of study of the U.S. population. I develop a dynamic discrete continuous choice model that allows for different education, occupation and fields of studies heterogeneity and I estimate it with rich financial data also models student debt decisions of individuals. The results show that changing the student debt repayment plan towards an income driven one increases graduation on average by 3p.p. The effect is stronger for low income individuals,suggesting that future income uncertainty can generate debt aversion, generating a barrier for higher education. Furthermore, moving towards a safer, more comfortable, repayment plan also changes the distribution of fields of the economy. Individuals now move towards fields that are more vocational like education or humanities, and also towards health.

*I am very grateful to Joan Llull for his continuous support and mentorship during my doctoral studies I acknolwadge funding from XX and XX. All remaining errors are mine.

1 Introduction

In XX The US government introduced student loans as a way of financing higher eduation, allowing students to afford tuitions that would have otherwise been inaccesible. Since then it has gained popularity and becomed one of the main instruments in the financing of higher education. Unfortunately, although it has allowed many to access university, it has generated controversy on bla bla bla.

Individuals will enroll to college fields based on their preferences but also on the expected financial returns of a degree. As showed in XX, bla bla bla. This paper investigates the effect of exepcted earnings on field of study introducing the student loans repayment.

FOLLOWING EMILIO. In XXXX THe US governet introduced student debt as a way of bal bla ba. Since then blabla. It helps bla but the trend is to go to IDR. What is the effec in the composition of majors?

In front of that, some of the literature tries to address the question of whether individuals borrow too much, but other part of the literature focuses on whether individuals have access to all the credit they would like. Human capital investment theory has been focused on the role that financial constrains play at college access since decades. In one of the very early contributions, Becker, 1975 establishes that in the presence of non-borrowing constraints low-income individuals would borrow against their future income to optimally invest in education. The problem is that since human capital can not be used as collateral borrowing markets might not provide all the credit needed. For this reason, the federal student loan opportunities were created to allow individuals to borrow against future earnings up to some amount. Since then, lots of studies have focused on whether individuals are borrowing constrained and what are the effects on education enrollment.

What has not been paid much attention is what is the effect of borrowing constrains on the institution and major of choice by individuals. Borrowing constraints do not only affect college enrollment decisions, but very likely will also affect the amount of money to invest in education once the individual has already decided to attain college and potentially what is studied there. As described by Lochner and Monge-Naranjo, 2012 the relationship between ability and constrained human capital investment is rarely studied in the literature and therefore this work will try to fill this gap.

For this reason this works aim to shed some light on which individuals are financially constrained when choosing the amount of money to invest in education and how this affects the major they choose. This is the first step to in the future try to identify whether individuals attending college face borrowing constraints that affect their investment decisions.

The huge increase in student debt over the last decades follows some economic trends of the U.S economy. The main trend is the rising returns to schooling that makes college more attractive, responding to an increase of the demand of high skilled labor in the labor market. As a consequence, there is also an important increase in college tuition fees and prices following an increase in the demand for college education. Ovearll, the average student now faces more returns to college and a higher price, increasing financial constraints and rising the demand for credit. This is consequently translated into individuals using more student loans and increasing the aggregate amount of that credit. This raises questions of whether individuals are borrowing too much or who is benefiting more from using this debt opportunities.

The goal of this paper is to investigate the link between the way individuals finance higher education and the incentives it provides to enroll into school, and which field of study to choose, with particular emphasis on low-income individuals since they should be the more reactive to those policies. I compare a baseline scenario in which indivduals follow the standard 10year repayment plan with one in which the "Sav On Valuable Education" repayment plan is introduced, which allows for a more comfortable repayment. To have a better understanding of the role of loans in accessing higher education and affecting the choice of a major, I also evaluate different scenarios in which student loans are not possible. To analyze those policies, I specify a human capital investment model, in which individuals with heterogeneous backgrounds will make educational decisions based on their preferences, their financial resources and considering expected uncertain returns from their investment. Within the model, individuals will endogenously accumulate student loans to meet their financial needs and debt aversion and uncertainty about college returns will play a role in the field of study individuals decide to attend.

In the model, individuals will face a discrete choice at every period that might or not imply an enrollment decision. Individuals that decide to enroll into school will have then to chose how much loans to get.

I estimate the model using the National Longitudinal Survey of Youth of 1997 (NLSY97), something about the data. The NLSY97 allows to create a year by year panel of individual enrollment spells, with detailed information on the school of choice, the field of study and the financial information of those spells.

Estimation of dynamic discrete continuous choice models is not very common because of their computational complexity. I estimate the model by a combination of maximum likelihood and simulated method of moments using a variety of known results to build an estimation algorithm that can feasible estimate such a model. By using the empirical conditional choice probabilities and a sequential estimation method, I manage to estimate the model and generate counterfactuals.

I find that introducing a more comfortable, income-driven, repaymet plan has positive effects on enrollment and graduation, specially for low-income individuals. SOME RESULTS HERE BLA BLA BLA. To better understand the role of student debt, I also compare the results with an scenario in which student debt is not available. The results suggest that the availability of student loans increases enrollment and graduation by SHOW SOME RESULTS HERE.

2 Related Studies

The rapid increase in student debt over the past decades has gained a lot of attention among researchers. In Avery and Turner, 2012 the authors question whether students are borrowing too much. To their purpose the authors try to adress the question of wether individuals are able to correctly forecast future income at the time of making education investment decisons. Also, the authors show how the present discounted value of college net of tuition fee is higher than having not college degree, even after accounting for self-selection of indviduals into college. Not only this, but the difference is increasing suggesting that the positive impact of college has a rising trend. For this reason, the authors concluded that the total amount of student debt is not prejudicial, and that individuals are not borrowing too much. In a similar way Webber, 2016 examines the lifetime income of pursuing a college degree to see whether college is a profitable investment. The author concluded that it is for most of the individuals, although in some cases the positive returns from college are obtained very late in life or never. This is the case for individuals at the lower part of the ability distribution when they get very indebted. Overall, most of the studies provide evidence in favor of individuals not borrowing to much, and for this reason the other question is whether they can borrow enough.

The capacity of individuals to invest in human capital has been widely studied. In the seminal work of Becker, 1975 a theoretical framework was proposed to understand how is the role of parental income and ability in human capital investment decisions. According to Becker, 1975 individuals from low-income families would borrow to invest in the human capital of their children in the presence of none borrowing constraints. However, the author suggested that low-income families might have limited borrowing opportunities which difficult investment in human capital. Since human capital cannot be used as a collateral, it generated problems to access to credit. This inability to borrow against future earnings is what lead to the creation of the Guranteed Student Loan program, as explained in Kane, 1996. Avery and Turner, 2012 also explain that student loans can help reduce the problem of underinvestment in human capital by improving the efficiency of the capital markets.

There are a lot of studies that try to study this issue, some of them with an empirical approach and others with structural models. In an empirical work Kane, 1996 studies how borrowing constraint might be affecting delayed entrance to college for constrained individuals, since they fist need to work and save to afford the investment. As mentioned in this work, the author suggest that it is impossible to directly observe how students are constrained in the investment decision and for this reason most of the literature used structural models. However, some studies have managed to do it. In Stinebrickner and Stinebrickner, 2008 the authors directly ask to the individuals whether they would like to borrow more, and they found that 20% of the individuals are in fact credit constrained.

In a structural model created by Cameron and Heckman, 1998 the auhtors analyze with a life cycle perspective the role of family income on shooling decisions and found that family income does not affect college attendance decisions. In fact, the authors claim that the relationship between parental income and college enrollment is because parental income is a proxy for early achievement. Another structural model was estimated by Keane and Wolpin, 2001 who developed a dynamic discrete choice model incorporating the role of borrowing constraints by establishing an assets lower bound. As explained by Keane, 2002 in their model agents can finance college attendance from different sources. They can reach the assets lower bound, they can reduce their savings, they can receive parental transfers or they can work while at college. Interestingly, the authors found that financial constraints do not affect the schooling decisions and only affects the working decisions while at college. In a smiliar manner Cameron and Taber, 2004 also estimated a structural model to study if credit access has an impact on educational attainment. The authors concluded that borrowing constraints are not generating inefficiencies and that improving borrowing constrains will in fact have little impact on college attendance. Furthermore , they state that there is no under-investment in schooling resulting from credit access. Overall, the literature suggests that financial constraints at college do not play a role and that the important financial constraints are at early stages of life. In Keane, 2002 the author argue that financial constraints have little effect on the college attendance decision and that most of the inequality is generated in early stages of human capital investement. This is also shown by Carneiro and Heckman, 2002 who explain that the most important constraint is the inability of the child to buy the necessary parental environment.

Another part of the literature focuses on how individual response to changes in college costs. Literature has shown that the effect of tuition fees on college attendance is higher for individuals from low-income families. In a literature review Leslie and Brinkman, 1987 agrees that an increase of $ 1000 in tution cost will imply a 5% change on schooling attendance. This evidence is also predicted by the model of Keane and Wolpin, 2001. Overall, the literature seems to suggest that tuition fees subsides have a larger effect on college attendance than the availability of student loans. For this reason, both Keane, 2002 and Keane and Wolpin, 2001 suggest that an important policy to reduce future earnings inequality by college attendance is a tuition fee reduction.

However, literature has evolved and although there was agreement that parental income was not a key determinant in education atteinment some new evidence seems to change the situation. This is the argument of Belley and Lochner, 2007 who found using two different cohorts of the National

Longitudinal Survey that the effect of family income on educational attainment has increased, even after controlling for family background and ability. The authors suggest that borrowing constraints might have increased and that this might partially explain the increase in the college attainment gap by family income. Furthermore, the authors found a negative relationship between family income and working during college, consistent somehow with the results at Keane and Wolpin, 2001. This change of pattern on financial constraints has been further studied by Lochner and Monge-Naranjo, 2011 who developed a theoretical model to try to understand this increasing role of parental income explaining college attendance. After calibrating the model the authors found evidence in favor of borrowing constraints. The authors argue that this fact might be a consequence of increasing college costs and returns to college.

Studies that show IDRplans are not used Mueller and Yannelis, 2022 and some dynarsky!

This paper analyzes student debt effects on field Field, 2009.

Some papers study the effect of tuition on student debt Chakrabarti et al., 2022.

On the modeling side of the paper, this works contributes to the literature on the estimation of dynamic-discrete-continuous-choice models. There are many papers estimating dynamic discrete models or continuous ones, but the use of discrete-continuous choice models is less extended in the literature. In Delavande and Zafar, 2019 the authors study the effect of financial constraints on collge choice in Pakistan.

In Dynarski et al., 2021 they focus on the role of financial complexity in bla.

3 The U.S. Higher Education System

As mentioned in Lovenheim and Smith, 2022, the higher education system in the U.S. is a decentralized market. On the one hand, there are different colleges with heterogeneous characteristics and on the other side there are millions of students that demand different education. The aim of this section is to provide a short review of how the U.S. higher education system works to familiarize the reader.

On the college side, one of the best ways of categorize college is by whether they provide a degree or not. On that sense, there are degree-granting and non-degree-granting colleges. The former is the standard college we all have in mind, the later is an institution that provides knowledge but not degrees.

The next important categorization is between which types of degrees do colleges provide. The distinction is divided in associate degree (AA) or bachelor degree (BA). They are related to the next categorization, which according to the Department of Education colleges can be divided as four or more years, at least two-years but less than four, and less-than-two years. Four-year colleges have over two thirds of total enrollment and they tend to award more BA than AA. On the other hand, two-year mainly provide AA. Four-year and two-year colleges also differ in their tuition prices, the former has an average price of $20,977 while the later has an average price of $6,486.

Another way of categorizing institutions is by their control. In this category we have public or private institutions. Public institutions are operated by the state or local governments and they have a limited tuition and have a controlled admission policies among other features. On the other hand, private institutions can either be for-profit or non-for-profit.

Institutions also differ in their level of selectivity and how the application and admission process works. According to Lovenheim and Smith, 2022 half of the colleges have no admission criteria. However, it depends a lot on the institution type. While 92% of two-year colleges are open enrollment, only 25% of four-year colleges have no admission criteria. Among the college with admission criteria, they often require applicants to submit motivation letters, recommendation letters, and they grades in college entrance exams like SAT or ACT, together with the high school GPA.

Finally, another peculiarity of the U.S. higher education system is that it is very flexible for choosing majors. In particular, students might declare they intended major at the moment of applying into college but it is very easy to switch major during the first two years. Therefore, it has a very flexible system compared to other countries. In the next sections some descriptive evidence of the system will be shown.

4 Student Loan Acquisition and Repayment

The main elements of the student loan system are the acquisition and repayment mechanisms. The dynamics of loans accumulation depend on whether loans are subsidized by the U.S. federal government or acquired from private lenders. I will focus on subsidized student loans. The key features are the interest rate, which is set yearly by the federal government[1], and the loans limit. Individuals have a limited amount of student loans that can be obtained at subsidized rate, before having to rely on private lenders. The nature

The repayment of student loans is a crucial element and has been in the spotlight. As mentioned in Looney and Yannelis, 2024 the standard repayment plan is a 10-year repayment in which students start paying back their loans upon graduation, with a six-month grace period. The principal amount is divided in 120 equal monthly payments. If the individual fails to pay back their loans for 270 days, it is assumed to enter in default. Defaulting on the loans have differet consequences since you can't get rid of them! Until 2009 that was about it, but since then more Income-Driven repayment plans have been adopted, bla bla bla.

[1]See Appendix XX to see the evolution of interest rates and a detailed discussion.

5 Data and Descriptive Evidence

This study uses data from the National Longitudinal Survey of Youth 1997 (NLSY1997). 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 virtous of this data set is that it allows to create a year-by-year spell information on educational choices and labor market outcomes, with very rich information on financial aid, fields and colleges, wages and occupations, complemented with family background and other demographic characteristics. Finally, it also includes the outcome of the AFQT test score, which is very combinient since it allows to proxy for individual's cognitive ability. In Appendix XX I provide a description of the sample restriction 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 periods 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 also use data from the American Community Survey (ACS) pooling years 2009 to 2019.

5.1 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 XXXX. As Table **??** 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. The labor market outcomes are also very different across education levels, which account for returns to education but also perhaps a selection effect. XXX.XXXX.XXX

[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 fertility choices to be able to better accommodate the females in my sample.

Table ?? 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 recieve less grants. Conversely, in both two-year and four-year schools individuals with more parental income recieve parenal transfers more times, and when reciving transfers, the average amount is much higher. For Four-year schools an individual from the bottom quartil of the parental income distribution will recieve about 5,00dollars, compared with 14,000 dollars recieved from an individual from the top parenal income distribution.

5.2 Heterogeneous Returns to College

Others have documented heterogeneous returns across institution quality or etc. I will focus on the field of study as a determinant of heterogeneity that will affect the financial decisions of individuals.

6 A Human Capital Model with Financial Decisions

In this section, I develop a dynamic model of human capital investment decisions that incorporates student debt accumulation, where individuals are forward-looking. The model builds on the frameworks of Johnson, 2013 and Arcidiacono et al., 2023, 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.

6.1 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 fields.[3] The model starts when individuals complete high school and operates in discrete time fashion over 10 periods, 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 sectors 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 is considered unimportant.

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 major. The available fields of study include Business, STEM, Education, Social Sciences, Humanities and Arts, Health, or Other, capturing the heterogeneity

[3] Explain the notion of fields

[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 XX, I discuss the growth of the state space over time. Restricting the education phase to the first 10 periods is a common assumption, as in Arcidiacono et al. (2024).

[5] Appendix XX 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 individual preferences. Those that have an Associate Degree can only decide weather 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 a Graduate School. Finally, individuals that hold a graduate degree can't 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. 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, an important 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 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.

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.

6.2 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}. Depending on the intensive-margin, the hourly wage will become a yearly part-time salary or a full-time one. $l \in \{0, parttime, fulltime\}$

The wage will be determined by the individual state space $\mathcal{X}_{it}$ and an idiosyncratic shock $\varepsilon_{oit} \sim N(0, \sigma_o^2)$ assumed to be independent across individuals and time. The individual state space is composed by individual time invariant characteristics $\mathcal{X}_{it}^1$ and endogenously acquired human capital $\mathcal{X}_{it}^2$. The time invariant characteristics include parental income, ability, sex and race.[6] The vector of endogenously

[6]For parental income and ability the state space is break in quantiles of the distribution, so there are 4 possible states for each of them. For the race variable I am only modeling black vs non-black individuals. In total there are 4X4X2X2

acquired human capital includes labor market experience, education level and major of graduation, if any. [7] The wage equation will be affected by the elements in X_{oit}.

The log wage equation can then be expressed as:

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

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.

6.3 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 observable characteristics $\mathcal{X}_{it}$. 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 $e \in \{2y, 4y, grad\}$.[8] 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 $\mathcal{X}_{it}$ 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 with functions:

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

$$g_{eit} = X_{git}\lambda_{ge} + \varepsilon_{git} \tag{5}$$

where $\varepsilon^p_{it} \sim N(0, \sigma_p)$ and $\varepsilon^g_{it} \sim N(0, \sigma_g)$.

= 64 possible time-invariant states.

[7] In Appendix XXX I describe in detail the composition of the state space.

[8] 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.

Finally the graduation probabilities are assumed to follow a logistic distribution:

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

This 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 decresing in parental income and increasing in the individuals ability, and parental transfers will be determined by parental income. The coeffcient 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.

6.4 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. The education decision is denoted by $e \in \{0, 2, 4, G\}$ and is composed by non-education, 2y school, 4y school or graduate school. The labor supply decision can be $l \in \{0, part, full\}$ and denotes no labor supply decision, part time and full time work respectively. The occupation is denoted $o \in O$ and the field of study is denoted $f \in F$.[9] Importantly, an occupation can only be chosen if the individual is not studying. Otherwise occupation is considered irrelevant. In the same spirit, a field of study can only be chosen if the education decision is a 4y school. A choice is then $d_{it} = (e, l, o, f)$. Following Arcidiacono et al., 2023 the base alternative is home production, with $d_{it} = (0, 0, 0, 0)$.

I will now characterize the payoff associated with each possible discrete alternative. In particular, the main trade off individuals have to make each year 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 a function $g_j(X_{it})$ that will capture preferences or abilities towards specific fields of study or occupations. Therefore, the period specific payoff function is defined in Equation (7) as:

$$U_{elof}(\mathcal{X}_{it}, b_{it}, \varepsilon_{ieloft}) = g_{elof}(\mathcal{X}_{it}) + \frac{c_{ieloft}^{1-\sigma}}{1-\sigma} + \varepsilon_{ieloft} \tag{7}$$

where ε_{ieloft} is an unobserved i.i.d preference shock assumed distributed Type 1 Generalized Extreme Value (GEV). where each coefficient is alternative specific, so the same state space might affect in a different way each choice. For identification purposes I will normalize the function to $g_{home}(X_{it}) = 0$ in the event that the individual chooses home production. Therefore, the interpretation of all the coefficients should be made with respect to the home production base category.

[9]Possible occupations are Business, STEM,... Possible fields are bla bla

In the next section I define the consumption and individual budget.

6.5 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 consumption levels. Following Arcidiacono et al., 2023 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 consumption equation is defined in Equation (8) as following:

$$c_{ieloft} = \begin{cases} W_{ilot} + G_{iet} + P_{iet} - \tau(j) + b_{elofit+1} - (1+r)b_{it} + \xi_{ieloft}, & \text{if } e \in \{2y, 4y, grad\} \\ W_{ilot} + R_j(b_{it}, W_{jit}) - (1+r)b_{it}, & \text{otherwise} \end{cases} \tag{8}$$

Where c_{ieloft} is the level of consumption of individual i in period t and alternative j. If an individual is not enrolled to school, consumption will depend in the yearly wage W_{ielot} [10] and the repayment scheme followed $R(b_{it}, W_{jit})$. If the individual decides to enroll into school it will face two more sources of income which are grants G_{jit} and parental transfers P_{jit}.[11] It will also need to pay the school tuition $\tau(j)$ 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 $\xi_{ijt} \sim N(0, \sigma_\xi^2)$. [12] Individuals will be then facing two sources of uncertainty regarding their budget constraint at each choice j. 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 the amount of parental transfers or grants that will be received. Notice that if the individual decides to engage in home production there will be no uncertainty in the budget constraint. [13] Because the different layers of uncertainty increase the computational complexity of the model, some simplifications will be done, as explained in Appendix XX.

An important element of the budget constraint is the student debt level $b_{j,i,t+1}$. This determines how much debt the agent decides to move for the next period. Notice that this will be either a decision

[10] In this case $W_{ilot} = \exp(w_{iot} + \varepsilon_{ijt}) \times 42 \times 40$ if the individual is working full time, $W_{jit} = \exp(w_{jit} + \varepsilon_{ijt}) \times 42 \times 20$ if working part-time and zero otherwise.

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

[12] In Appendix X a discussion regarding the consumption is made.

[13] Important, and following Arcidiacono et al., 2023 consumption will be bounded at a minimum value. This value will be set to be 3,842.25$ since it is the 2008 equivalent of the value found by HAI AND HECKMAN CITE THEM!

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.

6.6 Elements of $g(\mathcal{X}_{it})$

A key aspect of the flow-payoff is the elements that are included inside of the $g()$ function, which can be thought of controls for the non-monetary drivers of choices. Here I will include different families of parameters.

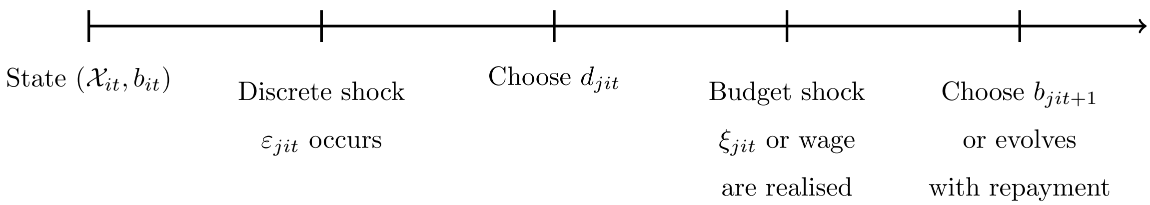
- Indicator variables for the different possible time-invariant state space characteristics and unobserved type effects. Those will vary at the e, o level.
- Switching costs to capture persistence in choices.
- Indicator variable to capture the fact that it might be the first time an individual is enrolled into a school. This will allow to capture persitance.
- 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., 2023 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.

6.7 Student Debt and Repayment Plan

A crucial element of the model is student debt b_{it}, the interest rate r, and the repayment plan $R(b_{it}, W_{it}, X_{it})$. Individuals will either endogenously accumulate student debt when choosing an educational alternative, or repay it under a 10 year repayment plan. To capture the dynamics of the Standard 10 Years Repayment Plan I will model the $R()$ function in such a way. For this reason, individuals will pay back the debt under the assumption that they are following a 10 year repayment plan. See Looney and Yannelis, 2024

6.8 Timming of the model

The timming of the model is such that the agents make the discrete choice observing the discrete-choice-specific shock ε_{jt}, but without observing the continuous-choice-specific shocks included in the wage and budget shock ξ_{jt}.



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 Blevins, 2014 and 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.

6.9 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:

$$\mathbb{E}\left[\sum_{t=1}^{T}\beta^{t-1}\sum_{j}^{J}d_{ijt}\left(g_j(\mathcal{X}_{it})+\frac{c_{ijt}^{1-\sigma}}{1-\sigma}+\varepsilon_{ijt}\right)\right] \tag{9}$$

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.

Applying the Bellman's optimality principle we can write the conditional value function as:

$$v_{jit}(X_{it}, b_{it}) = \begin{cases} g_j(\mathcal{X}_{it}) + \mathbb{E}\left[\max_{b_{ijt+1}}\left\{\frac{c_{ijt}^{1-\sigma}}{1-\sigma} + \beta V_{t+1}(X_{it+1}, b_{ijt+1})\right\}\right], & \text{if } e \in \{2y, 4y, grad\} \\ \\ g_j(\mathcal{X}_{it}) + \mathbb{E}\left[\frac{c_{ijt}^{1-\sigma}}{1-\sigma} + \beta V_{t+1}(X_{it+1}, R(X_{it}, b_{ijt}))\right], & \text{otherwise} \end{cases} \tag{10}$$

where the expectation denotes the fact that the utility of a specific alternative j is uncertain at the moment of making the decision given that the individual does not know the financial resources he will get. Furthermore, $V_t(X_{it}, b_{it})$ is the ex-ante value function which captures the expected discounted sum of future payoffs just before the idiosyncratic shock is revealed. Notice that for choices that do not imply an enrollment decision the conditional value function simplifies since the next period student debt level willl be determined by the repayment scheme. In that case, it looks like:

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

$$V_t(\mathcal{X}_{it}, b_{it}) = \ln\left(\sum_j \exp\{v_{jt}(\mathcal{X}_{it}, b_{it})\}\right) + \gamma \tag{11}$$

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.

6.10 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. To do so, it requires to assume what is the continuation value of the last period of the model, at a point at which agent decisions are not possible anymore. In this case, the terminal continuation value will be defined as the present discount value of the utility value of all future wages minus potential debt repayment scenarios. It is then defined as:

$$V_{T+1}(\mathcal{X}_{iT+1}, b_{iT+1}) = \sum_{l=35}^{60} \beta^{l-35} \frac{\left(\frac{w_l(\mathcal{X}_{iT+1}) - R(b_{iT+l})}{(1+r)^l}\right)^{(1-\sigma)}}{1-\sigma} \tag{12}$$

where the different wages $w_l(X_{i,T+1})$ will be estimated using ACS data.

7 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., 2023. 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 I can then estimate the parameters of the utility function. Sequential estimation will be crucial to address the computational complexity of the model.

First, consider the contribution to the likelihood function for an individual with unobserved schooling type k, given the observed sequence of choices d_{it}, wages w_{oit}, grants g_{eit}, parental transfers p_{eit}, graduation outcomes G_{eit}, and student debt decisions b_{it}. For simplicity, I will omit the dependence on observable characteristics, denoted as $\mathcal{X}_{it}$. The individual likelihood function, l_i, is given by:

$$\begin{aligned} l_i(k) &= L(d_{i1},...,d_{iT_i}, w_{oi1},...,w_{oiT_i}, g_{ei1},...,g_{eiT_i}, p_{ei1},...,p_{eiT_i}, G_{ei1},...,G_{eiT_i}, b_{i1},...,b_{iT_i}|k) \\ &= L_{dit}(k)L_{woit}(k)L_{geit}(k)L_{peit}(k)L_{Geit}L_{bit}(k) \end{aligned} \tag{13}$$

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_{Geit}, and student debt decisions, L_{bit}. [14] However, the individual's schooling type k is unobserved by the econometrician, which necessitates its integration 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:

$$lnl_i = \ln\left(\sum_{k=1}^{K}\pi_{ik}L_{dit}(k)L_{woit}(k)L_{geit}(k)L_{peit}(k)L_{Geit}L_{bit}(k)\right) \tag{14}$$

There are several important considerations to address in estimating the model's parameters. First, to account for the unobserved schooling taste type, I will integrate it out using the Expectation-Maximization (EM) algorithm. Additionally, the likelihood function is not additively 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 XXX.

[14] This follows from 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.

7.1 Summary Estimation Algorithm

The need to implement the EM-algorithm complicates the estimation by increasing the computational complexity. For this reason, I will build on the algorithm proposed by Arcidiacono et al., 2023 in which an auxiliary model with measures is used in estimation to identify the posterior distribution of unobserved types and take it as given in the estimation of the full model, allowing to just focus on the M-step where additive separaibility is reintroduced as noted in Arcidiacono and Jones, 2003.The estimation procedure is as follows:

1. Estimate an auxiliary model with measures. This will allow to obtain the posterior distributions for the unobserved type and the estimates for the CCPs. The use of measures will help in the identification and interpretation of the unobseved types. CITE THE PAPER THAT SAYS THAT!

2. Using the posterior probabilities as weights on the M-step, additive separability is reintroduced as mentioned in CITE ARCIDIACONO AND SOMEBODY, which implies

7.2 Estimation of the production functions

Thanks to the aditive separability properties re-introduced in the M-step, I can now bla bla bla. **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.

7.3 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. A detailed description of this process is provided in Appendix XX. To summarize, 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 preceision I will perform Aguirregabira and Mira.

7.4 Estimation of the Flow-payoff parameters

As mentioned before, the estimation of the utility parameters will be done taking as given the estimates from the previous stage. Given the Type 1 GEV assumption of the idiosyncratic preference shock, the likelihood of the choices becomes a multionomial logit:

$$\sum_{i=1}^{N}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{j=1}^{J(it)} \hat{q}_{ik} d_{ijt} \ln\left(\frac{\exp(v_{jt}(\mathcal{X}_{it}, b_{it}) - v_{ht}(\mathcal{X}_{it}, b_{it}))}{1+\exp(v_{jt}(\mathcal{X}_{it}, b_{it}) - v_{ht}(\mathcal{X}_{it}, b_{it}))}\right) \tag{15}$$

And since identification is only obtained up to a reference category, I will set home production as the reference alternative normalizing $g_h(\mathcal{X}_{it}, b_{it}) = 0$ as has already been explained in Section XX. The main problem in the estimation of this likelihood is that the conditional value function depends in part on the continuation value, which is a computationally costly object to construct.[15]

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 (CITE STUDIES USING CCPS), the benefits of the usage of CCPs in this study are lower than expected because of the use of the Nested Pseudolikelihood Algorithm described in Aguirregabiria and Mira that will be explained later. In a nuthsell, to improve the precision of my estimates, I will re-update the CCPs using model estimates until convergece, which means that I still need to solve the model. For this reason, the usage of CCPs in this model comes from the fact that they will allow to obtain a very particual expression of the difference of continuation values.

$$v_{jt}(\mathcal{X}_{it}, b_t) - v_{ht}(\mathcal{X}_{it}, b_t) = \underbrace{g_j(\mathcal{X}_{it})}_{\substack{\text{Flow-payoff function} \\ \text{to be estimated}}} + \underbrace{\delta(\mathcal{X}_{it}, b_t, d_{it})}_{\substack{\text{Function controlling} \\ \text{differences in continuation} \\ \text{values}}} \tag{16}$$

where $\delta(\mathcal{X}_{it}, b_t, d_{it})$ is a function that captures differences in the sequence of conditional choice probabilities and terminal continuation values obtained by alternative j and the base category h.[16].

[15] The construction of the continuation value requires solving the conditonal value functions for each possible state space backwards since the last period of the model. In Appendix XX a discussion can be found about the computational complexity of the model and its state space.

[16] Visit appendix XX to see the derivation of this term

The key element of Equation 16 is that the parameters to be estimated enter linearlly in the equation only through $g_j(\mathcal{X}_{it})$, and once $\delta(\mathcal{X}_{it}, b_t, d_{it})$ has been computed, it can be included in the estimation as a nuisance term that controls for the difference in future payofs across the two alternatives. The fact that $\delta(\mathcal{X}_{it}, b_t, d_{it})$ has been constructed using CCPs means that it is not any function controlling for the difference in continuation values, but rather the one that controlls optimally (or with the optimal parameters). Effectively, this means that the parameters after the maximization of this likelihood are already consistent. 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.[17]

8 Results

In this section I present the estimated parameters and discuss the model fit.

8.1 Estimated Parameters

8.1.1 Wage Parameters

Table XX reports the estimates of the wage equations for the time invariant individual charactaristics and the level of labor market experience. Overall, there is a positive gradiant 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 ocupations 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 XX 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

[17] See appendix XX for a detail explanation of the functioning of the NPL algorithm and the modifications done in this work to accomodate the current estimaiton

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. bla bla bla. 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 return from a field is not only measured on the monetary effect on occupations, but also on which occupations allows to access.[18] In Table XX we can see that having specific majors open the door to particual 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.

8.1.2 Financial Parameters

The results from the Financial Functions are reported in Tables X, X and X 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 recieve 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 recieve grants at two and four-year schools, capturing the dynamics of affirmative action. However, when reciving grants, the effect on females is almost negligeble and black individuals are positive discriminated in total amounts.

[18] 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.

8.1.3 Graduation Probabilities

The graduation probabilities reveal dynamics. Invididuals with more parental income are more likely to graduate. The highest effect is the one by the highest quartile of the ability ditribution, 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.

8.1.4 Flow Payoff Parameters

The coefficients of the flow payoff utility are reported in Table XXX and reveal the preference patterns of different individuals.

8.2 Model Fit

To address the fit of the model I simulate choices by individuals of my sample. 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. To obtain future terms I first solve the model backwards starting at the terminal period. In table XX I show how the model matches different choices of the model. As mentioned in Arcidiacono et al., 2023 this is not surprising as choices are used during the estimation. The important is to asses the goodness of fit on dynamics.

Figure XX shows how the model performs in things that are not directly targeted by the estimation. As we can see, xxx.

The other important aspect is how is the distribution of majors of the economy.

Finally, student debt fit is :

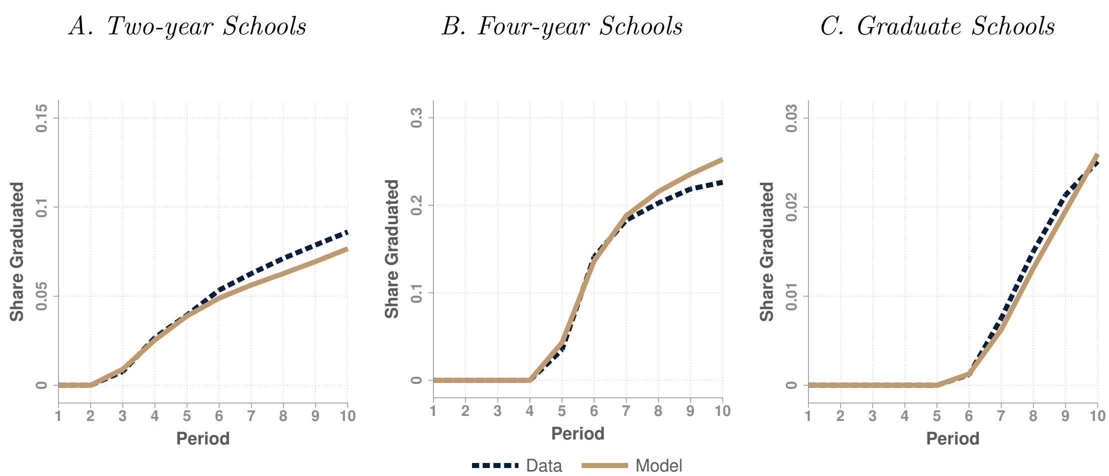
9 The Effect of Repayment on the Career

To evaluate the effects on the choice of field of study I will first evaluate two counterfactuals in which repayment plan is moved from a 10y repayment to an IDR. I will also compare those scenarios with an scenario in which student loans are not available and see what are the effects:

9.1 The Enrollment and Graduation Effect

9.2 The Field of Study Effect

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



Notes: This figure was created by simulating 30 individuals from each individual in the original sample. Those are untargeted moments that show the model fit in graduation dynamics. To see the same figure across parental income and ability distributions visit appendix XX.

Figure 2: **Graduation distribution by Field**

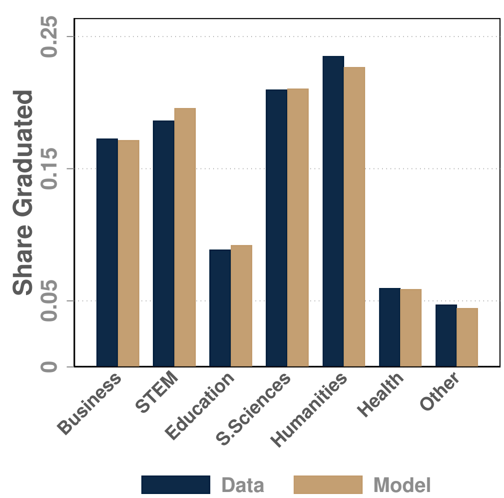


Table 1: Student Debt Fit by Ability and Parental Income

		Ability		Parinc	
		Data	Model	Data	Model
Q1	*Share*	0.70	0.62	0.78	0.72
	Avg	26757	18962	22931	22780
Q2	*Share*	0.74	0.63	0.78	0.73
	Avg	28404	21962	26005	23509
Q3	*Share*	0.70	0.70	0.74	0.72
	Avg	24801	23469	26448	22507
Q4	*Share*	0.58	0.64	0.49	0.57
	Avg	25762	20291	26596	19119

Notes: This table is created by simulating 30 times each individual of the sample, and simulating the discrete and educational choices.

References

Arcidiacono, P., Aucejo, E., Maurel, A., & Ransom, T. (2023). College attrition and the dynamics of information revelation. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4630803

Arcidiacono, P., & Jones, J. B. (2003). Finite mixture distributions, sequential likelihood and the em algorithm. *Econometrica*, *71*(3), 933–946. https://doi.org/https://doi.org/10.1111/1468-0262.00431

Avery, C., & Turner, S. (2012). Student loans: Do college students borrow too much–or not enough? *Journal of Economic Perspectives*, *26*(1), 165–92.

Becker, G. (1975). *Human capital: A theoretical and empirical analysis, with special reference to education, second edition*. National Bureau of Economic Research, Inc.

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.

Blevins, J. R. (2014). Nonparametric identification of dynamic decision processes with discrete and continuous choices. *Quantitative Economics*, *5*(3), 531–554. https://doi.org/https://doi.org/10.3982/QE117

Bruneel-Zupanc, C. A. (2022). Discrete-Continuous Dynamic Choice Models: Identification and Conditional Choice Probability Estimation. http://dx.doi.org/10.2139/ssrn.4072421

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.

Cameron, S. V., & Taber, C. (2004). Estimation of educational borrowing constraints using returns to schooling. *Journal of Political Economy*, *112*(1), 132–182.

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

Chakrabarti, R., Fos, V., Liberman, A., & Yannelis, C. (2022). Tuition, Debt, and Human Capital. *The Review of Financial Studies*, *36*(4), 1667–1702. https://doi.org/10.1093/rfs/hhac065

Delavande, A., & Zafar, B. (2019). University choice: The role of expected earnings, nonpecuniary outcomes, and financial constraints. *Journal of Political Economy*, *127*(5), 2343–2393. https://doi.org/10.1086/701808

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. Retrieved September 19, 2024, from http://www.jstor.org/stable/2984875

Dynarski, S., Libassi, C., Michelmore, K., & Owen, S. (2021). Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students. *American Economic Review*, *111*(6), 1721–56. https://doi.org/10.1257/aer.20200451

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. https://doi.org/10.1257/app.1.1.1

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. Retrieved September 19, 2024, from http://www.jstor.org/stable/1911491

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. https://doi.org/10.1086/669964

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. https://doi.org/10.1257/000282802320189429

Leslie, L. L., & Brinkman, P. T. (1987). Student price response in higher education. *The Journal of Higher Education*, *58*(2), 181–204.

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

Lochner & Monge-Naranjo, A. (2012). Credit constraints in education. *Annual Review of Economics*, *4*(1), 225–256.

Looney, A., & Yannelis, C. (2024). What went wrong with federal student loans? *Journal of Economic Perspectives*, *38*(3), 209–36. https://doi.org/10.1257/jep.38.3.209

Lovenheim, M. F., & Smith, J. (2022). Returns to different postsecondary investments: Institution type, academic programs, and credentials. (29933). https://doi.org/10.3386/w29933

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. https://doi.org/https://doi.org/10.1111/jofi.13088

Stinebrickner, R., & Stinebrickner, T. (2008). The effect of credit constraints on the college drop-out decision: A direct approach using a new panel study. *American Economic Review*, *98*(5), 2163–84.

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

Appendices

A Additional Figures and Tables **29**

B Detailed Estimation Algorithm **60**

B.1 Summary of the Algorithm 60

B.2 The Expected-Maximization Algorithm 61

B.3 The Auxiliary Model 62

B.4 CCP Representation of the Conditonal Value Function 64

B.5 Estimation of the budget shock distribution 66

B.6 Aguirregabiria and Mira 2002 Expanded Algorithm 67

B.7 Gradiant of the Full-Model Likelihood 67

C Comments About the Budget **68**

D Major-Occupation **69**

E Data Cleaning Appendix **69**

Table 2: Change in Four-Year School Graduation

	SAVE	No Loans	Grants
Total	3p.p	8	8

A Additional Figures and Tables

Table 3: 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 Section 5. To obtain the age of the individual, I normalize age 18 to the year in which high-school is finalized.

Table 4: 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.

Table 5: 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.

Table 6: 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 bla bla bla

Table 7: Sample Summary Statistics: Labor Market for Fields

		Business	STEM	Education	Social Sciences	Humanities	Health	Other
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, as described in the data section. For the wage I report the average wage and the standard deviation in parenthesis. The "Overall" category includes individuals from ages 25 to 65.

Table 8: 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

This table bla bla

Table 9: 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

This table bla bla

Table 10: 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)
Ability 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)
Ability 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)
Ability 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 bla bla bla

Table 11: 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 bla bla bla

Table 12: 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 bla bla bla

Table 13: 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 bla bla bla

Table 14: 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 bla bla bla

Table 15: 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)
Ability 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)
Ability 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)
Ability 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. Estimates for the XXX are excluded and reported in Table X, estimates for the X are reported in Table X, and estimates for the period effects are excluded. The reference category includes "Parental Income Q1", "Ability 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.

Table 16: 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)
Ability 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)
Ability 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)
Ability 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 education category. Estimates for the XXX are excluded and reported in Table X, estimates for the X are reported in Table X, and estimates for the period effects are excluded. The reference category includes "Parental Income Q1", "Ability Q1", "Male", "Non-Black" and "Non-Work".

Table 17: 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 effects for each education cell on the choice probabilities of each occupation.

Table 18: Choice Frequencies by Data and Model

	All Periods		Period 1		Period 9	
	Data	Model	Data	Model	Data	Model
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 shows the frequency of each choice without considering the different occupations or fields. The first column shows all periods together, and period 1 represents the first period of the model and period 9 represents the last one.

Table 19: Grant Equivalence Two-Year School at T=1

	Low Schooling Type			High Schooling Type		
	Base	Counter	Grant	Base	Counter	Grant
ParInc Q1	0.159	0.220	54,126	0.317	0.335	5,938
ParInc Q2	0.202	0.262	41,249	0.278	0.273	3,009
ParInc Q3	0.246	0.296	29,342	0.238	0.220	8,33
ParInc Q4	0.217	0.256	20,305	0.159	0.151	3,97
AFQT Q1	0.127	0.184	63,253	0.345	0.381	8,491
AFQT Q2	0.202	0.279	53,857	0.316	0.303	1,038
AFQT Q3	0.288	0.352	21,541	0.186	0.163	25
AFQT Q4	0.212	0.221	2,437	0.132	0.116	230
Male	0.178	0.225	36,662	0.237	0.247	4,841
Female	0.232	0.292	36,945	0.264	0.248	356
Non-Black	0.209	0.258	33,848	0.237	0.234	3,044
Black	0.192	0.256	45,329	0.288	0.287	1,459

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 20: Grant Equivalence Four-Year School at T=1

	Low Schooling Type			High Schooling Type		
	Base	Counter	Grant	Base	Counter	Grant
ParInc Q1	0.042	0.061	30,486	0.349	0.413	17,506
ParInc Q2	0.096	0.130	25,441	0.484	0.545	11,083
ParInc Q3	0.169	0.216	21,727	0.612	0.661	7,294
ParInc Q4	0.322	0.359	15,418	0.744	0.766	2,640
AFQT Q1	0.013	0.020	32,505	0.238	0.311	21,905
AFQT Q2	0.040	0.056	22,778	0.468	0.528	7,987
AFQT Q3	0.131	0.180	21,353	0.710	0.748	4,768
AFQT Q4	0.469	0.539	15,864	0.806	0.826	3,028
Male	0.135	0.168	27,620	0.512	0.557	11,232
Female	0.172	0.207	19,230	0.572	0.625	8,441
Non-Black	0.178	0.216	23,081	0.569	0.614	9,147
Black	0.082	0.104	24,696	0.462	0.523	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 10year repayment plan. Aggregation is weighted to respect the proportionality of each group.

Table 21: Graduation at Two-Year Schools Under Different Scenarios

	Baseline	SAVE	No Debt	Grants
ParInc Q1	0.050	0.061	0.037	0.050
ParInc Q2	0.075	0.082	0.057	0.070
ParInc Q3	0.089	0.097	0.067	0.086
ParInc Q4	0.066	0.075	0.066	0.065
AFQT Q1	0.045	0.055	0.030	0.042
AFQT Q2	0.075	0.082	0.062	0.075
AFQT Q3	0.087	0.095	0.083	0.081
AFQT Q4	0.076	0.086	0.053	0.075
Male	0.056	0.066	0.043	0.055
Female	0.084	0.092	0.070	0.080
Non-Black	0.074	0.084	0.060	0.072
Black	0.057	0.063	0.045	0.055

Table 22: Graduation at Four-Year Schools Under Different Scenarios

	Baseline	SAVE	No Debt	Grants
ParInc Q1	0.101	0.120	0.073	0.094
ParInc Q2	0.156	0.181	0.125	0.154
ParInc Q3	0.282	0.316	0.259	0.284
ParInc Q4	0.414	0.438	0.412	0.419
AFQT Q1	0.056	0.074	0.027	0.052
AFQT Q2	0.147	0.171	0.119	0.143
AFQT Q3	0.269	0.290	0.227	0.268
AFQT Q4	0.502	0.543	0.519	0.509
Male	0.196	0.217	0.183	0.194
Female	0.273	0.303	0.243	0.273
Non-Black	0.260	0.285	0.240	0.260
Black	0.159	0.186	0.133	0.155

Table 23: Graduation at Graduate Schools Under Different Scenarios

	Baseline	SAVE	No Debt	Grants
ParInc Q1	0.007	0.008	0.003	0.006
ParInc Q2	0.015	0.018	0.009	0.014
ParInc Q3	0.023	0.027	0.013	0.021
ParInc Q4	0.033	0.039	0.022	0.029
AFQT Q1	0.003	0.004	0.001	0.002
AFQT Q2	0.006	0.007	0.003	0.005
AFQT Q3	0.023	0.028	0.013	0.021
AFQT Q4	0.048	0.057	0.033	0.044
Male	0.016	0.017	0.009	0.013
Female	0.023	0.029	0.015	0.021
Non-Black	0.022	0.026	0.014	0.020
Black	0.011	0.013	0.006	0.009

Table 24: Drop-Out Under Different Scenarios

	Data	Baseline	SAVE	No Debt	Grants
ParInc Q1	0.572	0.552	0.533	0.586	0.556
ParInc Q2	0.516	0.514	0.493	0.535	0.504
ParInc Q3	0.382	0.380	0.355	0.392	0.364
ParInc Q4	0.339	0.327	0.311	0.322	0.314
AFQT Q1	0.612	0.614	0.578	0.689	0.610
AFQT Q2	0.521	0.492	0.469	0.527	0.489
AFQT Q3	0.472	0.437	0.428	0.451	0.428
AFQT Q4	0.299	0.319	0.294	0.318	0.304
Male	0.461	0.472	0.455	0.470	0.458
Female	0.393	0.363	0.346	0.369	0.354
Non-Black	0.387	0.390	0.374	0.390	0.377
Black	0.555	0.507	0.484	0.527	0.501

Notes: This table reports the four-year school drop out rates across different characteristics of individuals. Drop-out rate is computed as the share of individuals that have ever attended a four-year school, but have not graduated and have chosen not to attend during the last period of the model.

Table 25: Labor Supply Decision for Each Education Choice

	Baseline	SAVE	No Debt	Grants
Two-Year,No-Work	0.26	0.30	0.05	0.22
Two-Year,Part-Time	0.39	0.38	0.50	0.41
Two-Year Full-Time	0.35	0.33	0.45	0.37
Four-Year, No Work	0.34	0.37	0.24	0.31
Four-Year, Part-Time	0.41	0.40	0.48	0.44
Four-Year,Full-Time	0.24	0.23	0.28	0.25
Grad School, No Work	0.25	0.20	0.00	0.17
Grad School, Part-Time	0.24	0.27	0.33	0.27
Grad School, Full-Time	0.51	0.52	0.67	0.56

Notes: This table reports the labor supply shares for each educational choice under different scenarios. Labor supply shares sum to one within each educational choice.

Table 26: 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.

	Never	Drop Out	Two-Year Grad	Four-Year Grad	Grad
Never	0.862	0.128	0.006	0.005	0.000
Drop Out	0.000	0.942	0.025	0.033	0.003
Two-Year Grad	0.000	0.043	0.919	0.041	0.005
Four-Year Grad	0.000	0.044	0.006	0.952	0.106
Grad	0.000	0.027	0.004	0.969	0.906

Table 27: Education Level After SAVE

	Never	Drop Out	Two-Year Grad	Four-Year Grad	Grad
Never	0.652	0.268	0.026	0.054	0.001
Drop Out	0.000	0.856	0.035	0.110	0.006
Two-Year Grad	0.000	0.055	0.862	0.147	0.005
Four-Year Grad	0.000	0.045	0.024	0.944	0.080
Grad	0.000	0.028	0.007	0.966	0.862

Table 28: Education Level After SAVE

Table 29: Origin of New Four-Year Graduates

	Never	Drop Out	Two Year
ParInc Q1	0.108	0.754	0.138
ParInc Q2	0.066	0.729	0.205
ParInc Q3	0.071	0.707	0.222
ParInc Q4	0.057	0.786	0.157
AFQT Q1	0.148	0.731	0.122
AFQT Q2	0.043	0.716	0.241
AFQT Q3	0.043	0.726	0.231
AFQT Q4	0.077	0.774	0.149
Male	0.078	0.778	0.145
Female	0.071	0.715	0.214
Non Black	0.073	0.738	0.190
Black	0.075	0.752	0.172

Table 30: Decomposition of New Graduates at Four-Year Schools

	Total	Decomposition		
	Change	Never	Drop Out	Two Year
ParInc Q1	19.40%	1.88%	15.34%	2.18%
ParInc Q2	13.65%	0.88%	10.71%	2.06%
ParInc Q3	11.08%	0.72%	8.46%	1.90%
ParInc Q4	5.48%	0.23%	4.54%	0.71%
AFQT Q1	34.62%	5.04%	26.23%	3.35%
AFQT Q2	15.70%	0.49%	12.37%	2.85%
AFQT Q3	6.75%	0.25%	5.31%	1.20%
AFQT Q4	6.81%	0.45%	5.53%	0.82%
Male	9.60%	0.69%	7.87%	1.03%
Female	10.40%	0.62%	8.01%	1.77%
Non Black	9.01%	0.58%	7.06%	1.37%
Black	14.99%	1.02%	12.18%	1.79%

Table 31: Where Is New Enrollment Coming From

	All	Drop Out	Two-Year Grad	Four-Year Grad
ParInc Q1	0.35	0.35	0.27	0.29
ParInc Q2	0.29	0.29	0.30	0.21
ParInc Q3	0.21	0.21	0.26	0.31
ParInc Q4	0.15	0.15	0.17	0.19
AFQT Q1	0.38	0.38	0.33	0.36
AFQT Q2	0.30	0.31	0.27	0.14
AFQT Q3	0.22	0.23	0.30	0.14
AFQT Q4	0.10	0.08	0.09	0.36
Male	0.53	0.54	0.52	0.44
Female	0.47	0.46	0.48	0.56
Non Black	0.70	0.70	0.80	0.74
Black	0.30	0.30	0.20	0.26
Low Schooling	0.81	0.84	0.73	0.51

Table 32: Unobserved Type Distribution of New Enrolled

	All	Drop Out	Two-Year Grad	Four-Year Grad
ParInc Q1	0.70	0.74	0.43	0.17
ParInc Q2	0.81	0.83	0.71	0.44
ParInc Q3	0.89	0.91	0.88	0.71
ParInc Q4	0.95	0.97	0.98	0.76
AFQT Q1	0.61	0.66	0.39	0.02
AFQT Q2	0.88	0.90	0.75	0.10
AFQT Q3	0.99	0.99	1.00	0.96
AFQT Q4	0.99	1.00	1.00	0.98
Male	0.76	0.78	0.66	0.53
Female	0.87	0.90	0.79	0.49
Non Black	0.82	0.84	0.71	0.61
Black	0.79	0.82	0.79	0.22
All	0.81	0.84	0.73	0.51

Table 33: Elasticities to SAVE

	Never Enrolled	Drop Out	Switch	Two-Year
ParInc Q1	0.14	0.04	0.17	0.06
ParInc Q2	0.15	0.05	0.18	0.08
ParInc Q3	0.18	0.07	0.17	0.11
ParInc Q4	0.19	0.07	0.12	0.08
AFQT Q1	0.12	0.04	0.15	0.06
AFQT Q2	0.15	0.05	0.17	0.09
AFQT Q3	0.22	0.05	0.17	0.09
AFQT Q4	0.33	0.10	0.13	0.10
Male	0.14	0.05	0.13	0.07
Female	0.18	0.06	0.16	0.10
Non Black	0.16	0.06	0.15	0.08
Black	0.16	0.05	0.17	0.10
Low Schooling	0.14	0.03	0.11	0.05
High Schooling	0.35	0.08	0.16	0.11

Table 34: Changes Across Fields

	Total Change	Switchers In	Switchers Out	Switchers Net	New Enrolled	Previous Drop Out	Previous 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

Table 35: Changes Across Fields

	Business	STEM	Education	Social Sciences	Humanities	Health	Other
Overall	0.95	1.02	1.08	0.96	1.00	1.10	1.02
ParInc Q1	0.99	0.98	1.15	0.95	0.97	1.22	1.09
ParInc Q2	0.95	1.03	1.06	0.95	1.01	1.10	1.02
ParInc Q3	0.91	1.06	1.09	0.92	0.98	1.17	1.04
ParInc Q4	0.97	1.00	1.04	0.97	1.02	1.04	0.99
AFQT Q1	0.92	1.04	1.11	0.99	0.98	0.88	1.04
AFQT Q2	0.98	0.92	1.09	0.95	0.97	1.21	1.04
AFQT Q3	0.96	1.02	1.08	0.94	0.99	1.16	1.02
AFQT Q4	0.94	1.04	1.04	0.96	1.02	1.03	0.97
Male	0.96	1.01	1.06	0.95	1.04	1.01	1.04
Female	0.94	1.04	1.08	0.96	0.98	1.11	1.01
Non Black	0.95	1.03	1.07	0.94	1.00	1.09	1.02
Black	0.94	0.97	1.10	0.98	1.00	1.12	1.02
Low Schooling	1.01	1.07	1.05	0.97	0.95	0.99	1.10
High Schooling	0.95	1.01	1.09	0.95	1.01	1.12	1.02

Table 36: Decomposition of New Graduates

	Change	Never	Drop Out	Two Year
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%

Table 37: Elasticities to Loans

	Never Enrolled	Drop Out	Switch	Two-Year
ParInc Q1	0.20	0.06	0.34	0.22
ParInc Q2	0.24	0.07	0.29	0.22
ParInc Q3	0.27	0.08	0.29	0.24
ParInc Q4	0.11	0.04	0.11	0.06
AFQT Q1	0.19	0.05	0.28	0.15
AFQT Q2	0.19	0.06	0.29	0.20
AFQT Q3	0.31	0.09	0.27	0.25
AFQT Q4	0.31	0.04	0.16	0.06
Male	0.19	0.05	0.18	0.11
Female	0.25	0.08	0.23	0.22
Non Black	0.22	0.06	0.20	0.17
Black	0.22	0.06	0.29	0.24
Low Schooling	0.18	0.02	0.14	0.03
High Schooling	0.50	0.12	0.22	0.24

Table 38: Changes Across Fields Introducing Loans

	Business	STEM	Education	Social Sciences	Humanities	Health	Other
Overall	1.02	0.90	0.98	1.11	1.04	0.86	0.99
ParInc Q1	0.94	0.93	0.93	1.10	1.05	0.70	0.97
ParInc Q2	0.98	0.90	1.01	1.06	1.09	0.89	1.13
ParInc Q3	1.05	0.90	0.98	1.09	1.06	0.87	0.93
ParInc Q4	1.03	0.92	0.95	1.10	1.02	0.90	1.01
AFQT Q1	0.90	0.97	0.93	1.15	1.12	0.62	0.98
AFQT Q2	0.98	1.01	0.97	1.07	1.09	0.81	0.85
AFQT Q3	0.96	0.98	0.98	1.15	1.02	0.82	0.90
AFQT Q4	1.05	0.91	0.93	1.10	1.03	0.92	1.03
Male	1.05	0.91	1.02	1.12	1.00	1.01	1.05
Female	1.00	0.95	0.95	1.10	1.05	0.82	0.95
Non Black	1.05	0.89	0.98	1.10	1.04	0.88	1.00
Black	0.88	1.03	0.95	1.13	1.04	0.80	0.95
Low Schooling	0.99	0.90	0.92	1.06	1.07	0.96	0.98
High Schooling	1.03	0.90	0.98	1.12	1.03	0.85	0.99

Table 39: Changes Across Work

	Same	Less	More
ParInc Q1	0.27	0.39	0.34
ParInc Q2	0.27	0.40	0.33
ParInc Q3	0.27	0.44	0.29
ParInc Q4	0.25	0.54	0.21
AFQT Q1	0.25	0.43	0.31
AFQT Q2	0.25	0.41	0.34
AFQT Q3	0.26	0.42	0.31
AFQT Q4	0.27	0.52	0.22
Male	0.26	0.46	0.27
Female	0.26	0.46	0.28
Non Black	0.26	0.47	0.27
Black	0.26	0.43	0.31

B Detailed Estimation Algorithm

B.1 Summary of the Algorithm

In this section I will describe in detail all 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 to improve the precision of my estimates. It is important to mention that each of the iterations of the NFXP algorithm are consisten estimates of the structural parameters of the model, so the algorithm can be stoped at any interation. In my case I repeat until convergence to make sure I have the best model fit.

As mentioned in section XX the likelihood of the model is a finite mixture model, which will be estimated using the EM-algorithm. In the next section I explain with details how I use the EM-algorithm, but the probelm in this model is that it would be computationally unfeasible to perform the EM-algorithm in the full structural model. For this reason, and following CITE ARCIDIACONO, I will obtain the posterior unobserved type probabilities of the individuals of my sample using an auxiliary model expanded with measures to improve the identification of the latent types. Once I have obtained the posterior unobserved types probabilites distributions, I can use them as weights in the M-step of the full model to estimate the structural parameters. The procedure will be:

1. Estimate the posterior unobserved types probabilities using an auxiliary model expanded with measures.

2. Using the posterior probabilities, focus on the M-step of the full model using the previous estimates as weights. Here the procedure is :
   (a) Using the fact that additive separability is reintroduced in the M-step, as noted in Arcidiacono and Jones, I will estimate the model sequentially. This means I will first obtain the parameters of the wage, grants, parental transfers and graduation functions.
   (b) With those parameters as given, I will estimate the CCPs and use them to estimate the parameters of the flow-payoff equations.
   (c) Finally, I will use the likelihood of the student debt decisions to update the distribution of the unobserved budget shock. This will be done with Simulated Method of Moments because of the inexistance of a closed form expression.

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

B.2 The Expected-Maximization Algorithm

As described in Section XX, 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. [19] The algorithm is built under the fact that Dempster et al., 1977 realized that the first order conditons of the maximization problem expressed in equation XX, which is teh same here for simplicity, can be also obtaned from an equivalent problem expressed in Equation 20 in which the population unobserved type probabiliti is substituted by the .

$$(\hat{\pi}_k, \hat{\theta}) = \arg\max_{\pi,\theta} \ln \left(\sum_{k=1}^{K} \pi_{ik} L_{dit}(k) L_{woit}(k) L_{geit}(k) L_{peit}(k) L_{Geit} L_{bit}(k) \right) \tag{17}$$

Under this set up, the EM-algorithm corresponds a method to do that hta tna dthatn. Explain here the $k - th$ iteration procedure and the fact taht it shouñd be repoeated until convergence. Because of the computationally complexity of performing the EM-algorithm in a model such as the one described in this paper, I will then use the procedure described in arcidiacono et all 2024. The EM-algoritm is an iterative algorithm that proceeds as follows:

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 :

$$q_i(k|\theta^{(m)}, \pi^{(m)}) = \frac{\pi_k l_i(k)}{\sum_{k=1}^{K} \pi_k l_i(k)} \tag{18}$$

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

$$\pi_k^{(m+1)} = \frac{1}{N} \sum_{i=1}^{N} q_k(k)^{(m+1)} \tag{19}$$

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

[19]See Arcidiacono Practical Methods Paper for a detailed explanation of the usage of the EM-algorithm in this type of situations.

maximization problem that provides the same first order conditions as the finite mixture that we are trying to optimize.

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{t=1}^{T} q_i(k|\theta, \hat{\pi}) \ln \left[L_{dit}(k) L_{woit}(k) L_{geit}(k) L_{peit}(k) L_{Geit} L_{bit}(k) \right] \tag{20}$$

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 XX the computational complexity of permorning the EM-Algorithm in the full model is high, and for this reason and following Arcidiacono et al., 2023 I will estimate the posterior distribution of the unobserved types with an auxiliar 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 estimating 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.

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 $\hat{c}_{jit}$, where the expectation is with respect to wages, grants and parental transfers, without considering debt decisions. Equaiton (XXX) characterizes expected consumption, where expected wage, grants and parental transfers are computed as explained in Appenidx XXX. 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.

$$\hat{c}_{ieolft} = \hat{W}_{ioletft} + \hat{G}_{oelitf} + \hat{P}_{oeloft} - \tau(j) \tag{21}$$

The payoff for each alternative is modeled as:

$$\tilde{v}_{jt}(X_{it}, b_{it}) = \tilde{g}_j(\mathcal{X}_{it}) + \beta_{1elo}\hat{c}_{jit} + \beta_{2elof}b_{it} + \epsilon_{jit} \tag{22}$$

The likelihood of choices for individual i will then be:

$$\tilde{L}^a_{dit}(k) = \prod_{j=1}^{J} \tilde{P}_{itj}(\mathcal{X}_{it}, b_{it}, k)^{d_{it}} \tag{23}$$

And the total likelihood of the static model is:

$$\tilde{l}_i(k) = \tilde{L}_{dit}(k)\tilde{L}_{woit}(k)\tilde{L}_{geit}(k)\tilde{L}_{peit}(k) \tag{24}$$

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

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., 2023 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.

These measures are informative about the individual's latent schooling preference type. The likelihood of these binary outcomes is modeled using logistic regressions:

$$L^m_{it} = P(measure = 1|X_{it}, k) \tag{25}$$

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.

Auxiliar 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:

$$L_i^a = \sum_{k=1}^{K} \pi_k^a \tilde{L}_{dit}(k)\tilde{L}_{woit}(k)\tilde{L}_{geit}(k)\tilde{L}_{peit}(k) \times L_{m=1i}(k) \times L_{m=2i}(k) \tag{26}$$

The estimation procedure here is the one described in the Section before using the EM-algorithm. This will generate the CCPs estimates and the posterior probabilities for the unobserved type distribution.

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), such that it remains independent of the parameters governing the function $g_j(\mathcal{X}_{it})$. As initially demonstrated by Hotz and Miller, and subsequently utilized in many studies, including Arcidiacono et al., 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 continuation value can be represented as:

$$VT(\mathcal{X}_{it}, b_{it}) = v_{kt}(\mathcal{X}_{it}, b_{it}) - \log\left(P_{kt}(\mathcal{X}_{it}, b_{it})\right) \tag{27}$$

where $P_{kt}(\mathcal{X}_{it}, b_{it})$ is the conditional choice probability of choosing alternative k at period t given a state space $(\mathcal{X}_{it}, b_{it})$. Now, consider setting k as home production, implying the individual makes the minimum consumption level. The continuation value then becomes:

$$VT(\mathcal{X}_{it}, b_{it}) = \frac{c_{min}^{1-\sigma}}{1-\sigma} + \beta VT(\mathcal{X}_{it+1}, b_{it+1}) - \log\left(P_{kt}(\mathcal{X}_{it}, b_{it})\right) \tag{28}$$

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 function $g_j(\mathcal{X}_{it})$ is normalized to zero. Moreover, since home production implies that the individual consumes at the minimum level, this alternative eliminates budget uncertainty, simplifying the computational process.[20] 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

[20] 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.

27 into Equation 28, and iterating until the final period of the model—when individuals receive the terminal continuation value—we obtain the following expression:

$$
\begin{aligned}
&v_{jt}(\mathcal{X}_{it}, b_t) = g_j(\mathcal{X}_{it}) + \sum_{p=t+1}^{T} \beta^{p-t} \frac{c_{min}^{1-\sigma}}{1-\sigma} + \\
&\mathbb{E}\left[\max_{b_{jt+1}} \left\{ \frac{c_{jt}^{1-\sigma}}{1-\sigma} + \sum_{p=t+1}^{T} -\beta^{p-t} \ln P_{home,p}(x_p, b_p) + \beta^{T-t} V_{T+1}(x_{T+1}, b_{T+1}) \right\}\right]
\end{aligned}
\tag{29}
$$

This expression illustrates that the conditional value function for alternative j can be decomposed into the linear effect of the flow payoff, captured by $g_j(\mathcal{X}_{it})$, and a term that reflects 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 period, and finally, the terminal continuation value implied by this sequence of chioces

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:

$$
\begin{aligned}
&v_{jt}(\mathcal{X}_{it}, b_t) - v_{ht}(\mathcal{X}_{it}, b_t) = g_{jt}(\mathcal{X}_{it}) + \\
&\mathbb{E}\left[\max_{b_{jt+1}} \left\{ \frac{c_{jt}^{1-\sigma}}{1-\sigma} + \sum_{p=t+1}^{T} -\beta^{p-t} \ln P_{home,p}(x_p, b_p) + \beta^{T-t} V_{T+1}(x_{T+1}, b_{T+1}) \right\}\right] \\
&- \left[\frac{c_{min}^{1-\sigma}}{1-\sigma} + \sum_{p=t+1}^{T} -\beta^{p-t} \ln P_{home,p}(x_p, b_p) + \beta^{T-t} V_{T+1}(x_{T+1}, b_{T+1}) \right]
\end{aligned}
\tag{30}
$$

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

$$
\begin{aligned}
\delta(\mathcal{X}_{it}, b_t, d_{it}) = {} & \mathbb{E}\left[\max_{b_{jt+1}} \left\{ \frac{c_{jt}^{1-\sigma}}{1-\sigma} + \sum_{p=t+1}^{T} -\beta^{p-t} \ln P_{home,p}(x_p, b_p) + \beta^{T-t} V_{T+1}(x_{T+1}, b_{T+1}) \right\}\right] \\
& - \left[\frac{c_{min}^{1-\sigma}}{1-\sigma} + \sum_{p=t+1}^{T} -\beta^{p-t} \ln P_{home,p}(x_p, b_p) + \beta^{T-t} V_{T+1}(x_{T+1}, b_{T+1}) \right]
\end{aligned}
\tag{31}
$$

To finally get:

$$
v_{jt}(\mathcal{X}_{it}, b_t) - v_{ht}(\mathcal{X}_{it}, b_t) = g_{jt}(\mathcal{X}_{it}) + \delta(\mathcal{X}_{it}, b_t, d_{it}) \tag{32}
$$

As mentioned in section XX, 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 optimal continuation value. This allows for a simple multinomial logit estimation.

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 XX, 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:

$$b^*_{jt+1} = \arg\max_{b_{jt+1}} \left(\frac{(budget_{jit} + b_{jit} + \xi_{jit})^{1-\sigma}}{1-\sigma} + \beta VT(\mathcal{X}_{t+1}, b_{jt+1}) \right) \tag{33}$$

which can be expressed using CCPs as:

$$\begin{aligned} b^*_{jt+1} = & \arg\max_{b_{jt+1}} \frac{(budget_{jit} + b_{jit} + \xi_{jit})^{1-\sigma}}{1-\sigma} + \\ & \sum_{p=t+1}^{T} -\beta^{p-t} \ln P_{home,p}(x_p, b_p) + \beta^{T-t} V_{T+1}(x_{T+1}, b_{T+1}) \end{aligned} \tag{34}$$

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 constinuation values are constructed using the optimal variation.

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 infeasible. Therefore, estimation will be conducted using the Simulated Method of Moments (SMM).[21] 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:

[21] See Chapter 4 of Adda and Cooper for a detailed description of this method.

$$
\begin{aligned}
(\hat{\mu}, \hat{\sigma}^2_{\xi}) = \arg\min_{\mu,\sigma^2} & \left[\frac{m(b^*_{it}) - \frac{1}{S}\sum_{s=1}^{S} m(b^*_{it+1}(\xi_{it}(\mu,\sigma^2)))}{m(b^*_{it})} \right]' W^{-1} \\
& \left[\frac{m(b^*_{it}) - \frac{1}{S}\sum_{s=1}^{S} m(b^*_{it+1}(\xi_{it}(\mu,\sigma^2)))}{m(b^*_{it})} \right]
\end{aligned}
\tag{35}
$$

In the previous expression W^{-1} is the weighting matrix that can be estimated using a two-step SMM procedure to get the optimal weighting 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 Aguirregabiria and Mira 2002 Expanded Algorithm

In this section I will describe how I finally build the estimation algorithm puting together the different pieces obtained from the different sections. The Nested Pseudolikehood estimation algorithm will improve the precision of my estimates. The procedure is then:

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 and go back to step (3) until convergence.

B.7 Gradiant of the Full-Model Likelihood

The optimization of the likelihood function is done using XXX, a gradiant based method. Because of the high number of parameters of the model, using a numerical approximation of the derivative increases the computational time. To adress this concern and speed up the estimation, I will instead use the theoretical expression of the gradiant of the likelihood. The objective function for this step is the weighted log likelihood of the flow-payoff parameters:

$$
\ln L_{dit} = \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J(it)}\sum_{r=1}^{R} \hat{q}_{ir}
$$

C Comments About the Budget

In this section I describe in detail the creation of the consumption variable. 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 XXX, 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. To address this issue I will explain how consumption will be obtained.

In order to obtain consumption during non-education decisions, I will model it as the expected wage based on the wage distribution of each of the different occupations.

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. In total this sums to many possible scenarios that become prohibitive even with the usage of Gaussian quadrature to perform the integrals. For this reason, I will instead compute expected wages, grants and transfers values and plug that in the consumption equation. Therefore :

[equation arcidiacono]

However, to still have one layer capturing all the uncertainty, I have introduced the budget shock that is capturing all the remaining budget that deviates from the expected one.

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

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

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

Now, the expected utility is:

$$E[\frac{C}{1-\sigma}] \tag{36}$$

D Major-Occupation

E Data Cleaning Appendix

In this part of the paper I will discuss the data cleaning and sample selection performed for the final data.

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 comissions, bonuses, tips or others. To learn about how this variables where created visit here https://nlsinfo.org/content/cohorts/nlsy97/topical-guide/employment/wages

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 44 reports the number and share of missing observations for each relevant variable. I will now describe the process of data inputation, 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 bla bla bla.

Table 40: 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.

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 41: Distribution of Occupations Chosen by Human Capital Cells

	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 42: Caption

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

Table 43: Occupations That Can be Chosen for Each Human Capital Cell

Table 44: Missing Data Employment Variables

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

Table 45: Employment Status on the Data

	Number	Share
Total Person-Month Observations	2,138,844	100
Status		
No Information	12,120	0.57
Not Associated / Not in Labor Force	60,318	2.82
Not Working (Unkown Reason)	15,755	0.74
Associated XXXX	4,983	0.23
Unemployed	129,844	6.07
Not in Labor Force	495,660	23.17
Active Militar Service	38,474	1.80
Employed	1,381,690	64.60

Table 46: Missing Occupatoin

	Number	Share
Total Missing	12,815	0.93
Process		
Input Mode from same Academic Year	9,579	0.69
Input Previous or Next	292	0.58
Code as Missing Type	0	0.00

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. This can be found here.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. I HAVE MERGED THIS DATA WITH GRADUATION DATE, IDENTIFIED INCONSISTENCIES AND INPUT SOME VALUES TO MAKE THE PATHS MORE CONSISTENT.

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. [22] 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 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. PUT TABLE HERE ABOUT WHEN TERM INFO SHOULD BE AVAILABLE. Also, this only applies to undegraduates, 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 missings at a non-negligible rate. Table 46 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 inputed 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

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

Table 47: Missing Data of College Variables

	Number	Share
Total Person-Term Observations	49,988	100
Variables		
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

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. As described in Table 47 I have first inputed 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.

Table 48: Missing Data of College Variables

	Number	Share
Total Missing	21	100
Process		
Input Academic Year from Same Round	14	0.04
Input Year of Interview	0	0

Major. The major variable is obtained from "First Major" variable (YSCH-21300). In Appendix XX I describe the creation of the different groups. In total 634 observations are missing (1.27%). To understand the major inputation 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. As described in Table 49 this will only leave 15 observations without major (which correspond to 3 individuals). I will for now code them as group Other Field.

Table 49: Missing Major

	Number	Share
Total Missing	634	100
Process		
Input Major from same Academic Year	402	XXX
Input Previous or Next	15	0.03
Code as Other Field	0	0.00

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. This variable is obtained from two different sources. For undergraduate level it is obtained from the college experience term rooster. For the graduate level enrollment it is obtained from the yearly reported graduate loans variables. In total there are 2,339 missing observations which represent 4.68 % of the total loans observations. The imputation process will be different than for the other variables, since here it is important to recover every single observation, even if for a given academic year I already observe some loans. Notice that some individuals do not know the exact value of the loans they took, but they provide a range. In that case I have imputed the lower bound for that range. Then, the first imputation will be that for some individuals they don't remember the exact amount of loans they took, but they know the current value of the loans at the moment of the interview,

Table 50: Missing College Type

	Number	Share
Total Missing	696	1.27
Process		
Input Type from same Academic Year	601	1.20
Input Previous or Next	292	0.58
Code as Missing Type	0	0.00

which is not that far from the actual value given that interviews are conducted yearly.

Table 51: Missing Loans

	Number	Share
Total Missing	2.339	4.68
Process		
Imput Current Loans	1,630	3.26
Imput Average Loans	394	0.79
Code as Missing Type	0	0.00

Grants. This variable is obtained from two different sources. For undergraduate level it is obtained from the college experience term rooster. For the graduate level enrollment it is obtained from the yearly reported graduate loans variables. In total there are 3,296 missing observations

Employer Asisstance. This variable is obtained from schooling rooster.

Table 52: Missing Grants

	Number	Share
Total Missing	2.339	4.68
Process		
XXX	XXX	3.26
XXX	XXX	0.79
XXX	0	0.00

Table 53: Missing Employer Assistance

	Number	Share
Total Missing	2.339	4.68
Process		
Imput Current Loans	1,630	3.26
Imput Average Loans	394	0.79
Code as Missing Type	0	0.00