I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16585

College Attrition and the Dynamics of Information Revelation

Peter Arcidiacono
Esteban Aucejo
Arnaud Maurel
Tyler Ransom

NOVEMBER 2023

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16585

College Attrition and the Dynamics of Information Revelation

Peter Arcidiacono
Duke University, IZA and NBER

Esteban Aucejo
Arizona State University, CEP, CEPR, CESifo and NBER

Arnaud Maurel
Duke University, IZA and NBER

Tyler Ransom
University of Oklahoma and IZA

NOVEMBER 2023

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

College Attrition and the Dynamics of Information Revelation*

We examine how informational frictions impact schooling and work outcomes. To do so, we estimate a dynamic structural model where individuals face uncertainty about their academic ability and productivity, which respectively determine their schooling utility and wages. Our framework accounts for heterogeneity in college types and majors, as well as occupational search frictions and work hours. Individuals learn from grades and wages in a correlated manner, and may change their choices as a result. Removing informational frictions would increase the college graduation rate by 4.4 percentage points, which would increase further by 2 percentage points in the absence of search frictions. Providing students with full information about their abilities would also result in large increases in the college and white-collar wage premia, while reducing the college graduation gap by family income.

JEL Classification: C35, D83, J24

Keywords: college dropout, dynamic discrete choice, learning, human capital

Corresponding author:
Tyler Ransom
University of Oklahoma
Department of Economics
308 Cate Center Dr
Room 158
Norman, OK 73072
USA

E-mail: ransom@ou.edu

* First version: June 2016. We thank the Editor, Jim Heckman, four anonymous referees, Stephane Bonhomme, Chris Flinn, Mitch Hoffman, Lance Lochner, Thierry Magnac, Bob Miller, Salvador Navarro, Derek Neal, Matt Wiswall and Ken Wolpin, and audiences at many seminars and conferences for helpful comments and suggestions at various stages of this research.

1 Introduction

Over 40% of the students entering four-year college in the United States do not earn a bachelor's degree within six years (National Center for Education Statistics, 2021b). To the extent that there is a large wage premium to receiving a four-year college degree (Heckman, Layne-Farrar, and Todd, 1996; Heckman, Lochner, and Todd, 2006; Goldin and Katz, 2008; Bound and Turner, 2011; Ashworth et al., 2021), this suggests that imperfect information and learning may be important to the decision to leave college. The aim of this study is to quantify the role of information frictions associated with one's own academic ability and labor market productivity in determining college outcomes and sorting in the workforce.

To this end, we characterize the impact of imperfect information on college enrollment, attrition, and re-entry by estimating a dynamic model of schooling and work decisions in the spirit of Keane and Wolpin (1997, 2000) with the crucial distinction that such decisions are allowed to depend on the arrival of new information about the abilities of individuals both in school and in the workplace. After graduating from high school, individuals decide in each period whether to attend college and/or work part-time, full-time or engage in home production. Should the individual attend college, he must also choose between attending a two-year college, a four-year college in a science major, or a four-year college in a non-science major. Moreover, he can decide to work in the blue-collar or, if he receives an offer, in the white-collar sector. Incorporating search frictions in the white-collar sector—frictions that depend in part on education—is important in order to match the age and education profiles of blue- and white-collar work. Upon college graduation, which is probabilistic from the individual's standpoint, the options available are reduced to working part-time or full-time in one of the two sectors or engaging in home production.

Importantly, individuals are allowed to have imperfect information about their abilities and enter each year with beliefs regarding their different kinds of schooling abilities as well as their skills in the workplace. At the end of each year, individuals update their beliefs given their grades for their particular schooling option (if they attended school) or their wages (if they worked). We account for the multidimensional nature of ability and allow the different kinds of schooling and workplace abilities to be arbitrarily correlated, implying that signals in one sector may be informative about abilities in another sector. As they accumulate ability signals, agents update their beliefs about their future potential wages in each sector, their flow utility payoffs from attending school, as well as their college graduation probabilities.

We estimate a richer model than previously possible by making use of innovations in the computation of dynamic models of correlated learning. Following James (2011), we (*i*) integrate out over actual abilities as opposed to the signals, and (*ii*) use the EM algorithm where at the maximization step ability is treated as known, resulting in a correlated learning model that is computationally feasible. Using results from Arcidiacono and Miller (2011, 2019), estimation continues to be computationally simple even in the presence of unobserved heterogeneity that is known to the individual. Leveraging this approach in our current context makes the estimation of our correlated learning model both feasible and fast. Importantly, it also allows us to easily take into account heterogeneity in schooling investments by distinguishing between two- and four-year colleges, as well as science and non-science majors for four-year colleges.[1]

We use the estimates of our model to quantify the importance of informational frictions in explaining college enrollment decisions, the observed transitions between college and work, and to evaluate the impact of imperfect information on ability sorting. We find that a sizable share of the dispersion in college grades and wages is accounted for by the ability components initially unknown to the individuals. Focusing on the ability components unknown to the individuals at the time of high school graduation, we find that schooling abilities are highly correlated across college types and majors (namely, 2-year college, 4-year college science major, 4-year college non-science major). We also show that the correlation in productivity in the blue-collar and white-collar sectors is positive and sizeable, highlighting the importance of allowing for correlated learning in this context. On the other hand, our estimation results point to positive but weaker correlations between the different types of schooling abilities and productivity in both sectors. This indicates that grades earned in college actually reveal relatively little new information about future labor market performance, once we account for background characteristics and college readiness. As a result, college graduates are somewhat well sorted on the basis of college abilities but less so in terms of work abilities.

We then simulate our model under a counterfactual scenario where all individuals have full information on their abilities by the end of high school. The goal of the simulation is to understand how information affects schooling choices, ability sorting patterns, and the earnings gap between college and non-college graduates in the different work sectors. We find that removing informational frictions would increase the four-year college graduation rate by 4.4 percentage points. This increase in graduation rate relative to the baseline is driven by efficiency gains that lead to a substantial decrease in the share of students dropping out of college despite a significant increase in the share who never attended college. Providing full information to the students also results in a smaller graduation-parental income gradient.

[1] See the surveys by Altonji, Arcidiacono, and Maurel (2016) and Altonji, Blom, and Meghir (2012), who discuss the importance of heterogeneity in human capital investments.

The mechanism for this convergence lies in differences in beliefs about the suitability of college between low- and high-income individuals. Many low-income individuals have priors that college is not a good match. Providing information reveals that some of them are indeed a good match, increasing their college graduation rates. The reverse holds true for high-income individuals; many high-income individuals have priors that college is a good match, with full information revealing that, for some, it is not. The slow revelation of information in the baseline leads to higher graduation rates for individuals from high-income families because of switching costs combined with the accumulation of some years of college experience. Results from an additional simulation exercise that removes occupational search frictions indicate that four-year college graduation rates would increase further by 2 percentage points if individuals always had the option to work in the white-collar sector.

Simulations also reveal that ability sorting would be much stronger in the perfect information scenario. While there is significant sorting on college abilities for college graduates, this sorting becomes even stronger in the counterfactual. But a new source of sorting also emerges. Namely, because the premium for being a college graduate—and, in particular, a science graduate—is especially high in the white-collar sector, those with high white-collar abilities are more likely to obtain college degrees, especially in the sciences. With individuals in the counterfactual now sorting into college based in part on white-collar abilities, the wage gap between college graduates and non-college graduates at age 28 increases by about 15 percentage points, and the wage gap between white-collar and blue-collar workers increases by as much as 43 percentage points.

Our analysis relates to early research by Manski and Wise (1983) and Manski (1989), which argues that college entry can be seen as an experiment that may not lead to a college degree. According to these papers, an important determinant of college attrition lies in the fact that, after entering college, students get new information and thus learn about their abilities. More recently, several other papers in the literature on college completion stress the importance of learning about schooling ability to account for college attrition (see, e.g., Altonji, 1993; Arcidiacono, 2004; Heckman and Urzúa, 2009; Hendricks and Leukhina, 2017; Larroucau and Rios, 2022). Of particular relevance to us are the articles by Stinebrickner and Stinebrickner (2012, 2014), who provide direct evidence, using subjective expectations data from Berea College (Kentucky), that learning about schooling ability is a major determinant of the college dropout decision.[2]

[2]See also Hastings et al. (2016) who provide evidence using a large-scale survey conducted in Chile that individual beliefs about earnings and costs of higher education at the time of college entry are associated with dropout outcomes.

Much of the learning literature assumes that the labor market is an absorbing state, implying that the decision to leave college is irreversible (Stange, 2012; Stinebrickner and Stinebrickner, 2012, 2014; Trachter, 2015).[3] In this paper, we relax this assumption, which is important to predict the substantial college re-entry rates of 40% among those who left college for at least a short period of time in our data. By quantifying the importance of learning on schooling abilities as well as labor market productivities, and evaluating the joint informational value of schooling and labor market outcomes, our paper brings together the literature on schooling choices and occupational choices under imperfect information (see, e.g., Miller, 1984; James, 2011; Antonovics and Golan, 2012; Papageorgiou, 2014; Sanders, 2014).[4] Beyond the context of educational choice and labor market sorting, our paper also fits within the rich empirical literature on dynamic learning models, which have often been estimated in marketing in particular since the seminal work of Erdem and Keane (1996) (see Ching, Erdem, and Keane, 2013 for a survey). A key additional challenge that arises in our context is that we do not observe individuals making the same choice multiple times, making it more difficult to identify the role of learning.

The remainder of the paper is organized as follows. Section 2 presents the data and provides descriptive evidence suggestive of learning. Section 3 describes a dynamic model of schooling and work decisions, where individuals have imperfect information about their schooling ability and labor market productivity, and update their beliefs through the observation of grades and wages. Section 4 discusses the identification of the model, with Section 5 detailing the estimation procedure. Section 6 presents our estimation results. Section 7 studies the role of informational frictions on educational and labor market outcomes. Finally, Section 8 concludes. All tables are collected at the end of the paper.

2 Data

We use data from the National Longitudinal Survey of Youth 1997 (NLSY97). The NLSY97 is a longitudinal, nationally representative survey of 8,984 American youth who were born between January 1, 1980 and December 31, 1984. Respondents were first interviewed in 1997 and have continued to be interviewed regularly on topics such as family background

[3]In a different context, Pugatch (2018) provides evidence that the option to re-enroll in high school in South Africa is an important determinant of the decision to leave school and enter the labor market.

[4]See also recent work by Proctor (2022), who builds on the framework developed in this paper to examine the extent to which family-driven prior ability beliefs contribute to intergenerational persistence in schooling and labor market outcomes.

characteristics (e.g., parental education, family income, race), labor market information (e.g., compensation, labor supply, occupation), as well as education (e.g., educational experiences, high school and college GPAs, SAT scores, field of study). Our estimation sample relies on the first 17 rounds of the survey, where we restrict the analysis to men who have graduated high school.[5] After imposing some additional data restrictions, the final sample includes 22,398 person-year observations for 2,300 men.[6]

2.1 Descriptive Statistics

Table 1 presents background characteristics conditional on the first college option chosen, where we distinguish four-year college options by entering as a science major, entering as non-science major, or not reporting a major. Individuals who attend college at some point and start at a four-year institution have, on average, higher SAT test scores, with science majors having higher scores than non-science majors, even for the verbal section. The same pattern holds for high school grades. Those who begin in a two-year college have worse academic credentials than those who start in a four-year college, but significantly stronger academic backgrounds than those who do not attend college at all. Those who begin in a four-year institution have higher parental income and parental education, with those who begin in a two-year college being stronger on these measures than those who never choose a college option.

Descriptive statistics by employment sector (blue or white collar) and by college graduation status are presented in Table 2.[7] The sample corresponds to individual-year observations. Using data from the March Current Population Survey (CPS), we classify each three-digit 2010 Census occupation code as white collar if 50% or more of male workers aged 18–64 in that occupation hold at least a bachelor's degree. The other occupations are classified as blue collar.[8] For both college graduates and non-college graduates older ages are associated with a higher frequency of white collar work. There are a number of explanations for this phenomenon. First is differential learning; for example, learning about abilities in both sectors may be faster in the blue collar sector. Second is that blue collar experience may have a higher return in the white collar sector than in the blue collar sector. Finally

[5]We restrict the analysis to males to avoid having to model fertility decisions that would considerably add to the complexity and computational burden of our model. Interviews for rounds 1–15 were conducted annually with biennial interviews in rounds 16 and 17. The last calendar year in our sample is 2015.

[6]Appendix A provides details on how we construct our key variables and final estimation sample.

[7]Appendix Table A2 reports the most common occupations by sector.

[8]See, among others, Clark, Joubert, and Maurel (2017) for a similar approach.

(and consistent with what we find), search frictions restrict access to white collar jobs early in the life cycle. The other patterns are as expected; conditional on graduation outcomes, individuals in the white-collar sector have higher SAT scores, high school GPA, and family income (when a teenager) than those in the blue-collar sector.

We next examine rates of college completion by initial college type. Table 3 provides frequencies related to three measures of interest: (*i*) continuous enrollment (in either two- or four-year college) until graduation from a four-year college; (*ii*) stopping out (i.e., leaving college before graduating from a four-year college and returning to school at some point); and (*iii*) dropping out (i.e., permanently leaving college before four-year graduation). These summary statistics show that stopping out is quite prevalent in our sample: approximately 25% left and returned to college at some later point. Similarly, dropout rates are large with almost 36% of students never earning a bachelor's degree.[9] Another empirical regularity that emerges from this table is that student behavior varies substantially depending on initial college and major. For example, dropping out and stopping out are more common in two-year colleges than in four-year colleges, with four-year science majors having the lowest proportions of dropping out and stopping out. Overall, the frequencies in Table 3 provide two main takeaways regarding modeling considerations. First, dropping out is not an absorbing state: more than 40% of the students who left college at some point returned in a later period.[10] Therefore, college re-entry might matter for understanding how information frictions affect completion. Second, there is important heterogeneity in completion outcomes depending on the type of college (i.e., two-year vs. four-year) and major.

2.2 Descriptive Investigation of Learning

The large dropout and stopout rates suggest that students are learning about their abilities while in college. To assess the relevance of this possible mechanism, we provide descriptive evidence consistent with the idea that students are discovering something new about themselves, as opposed to merely reflecting something they already knew at the time of enrollment.

[9]These figures match the national ones, but official statistics do not track stopout because it requires longitudinal data on enrollment gaps. Official statistics track degree completion within a set time (six years), which in our sample equals 45% (62% for those starting in four-year college; not reported in the table). This 62% figure exceeds the graduation rate for men starting four-year college in the early 2000s (57%) reported by National Center for Education Statistics (2021a). This difference stems from Table 3 having a longer time horizon than six years.

[10]This number is computed from the final column of Table 3. The stopout rate is divided by the sum of the stopout and dropout rates, i.e., $24.84\%/(24.84\% + 35.98\%) = 40.84\%$.

The first two panels of Table 4 show how student decisions to stay in college vary by their college grades. To the extent that college GPA works as a signal for students about their abilities, then we should expect that those receiving lower grades (i.e., negative signals) would be more likely to leave college. Panels A and B of Table 4 show that students who stay in college (either four-year or two-year college) at period $t+1$ do, in fact, have significantly higher grades in period t than those who were in school in t but not in $t+1$. While these differences may not necessarily reflect learning (lower grades may result from worse family backgrounds and/or lower ability), they are consistent with the idea that some of the students who leave college do so as a result of new information about their ability.

In order to make further progress on the importance of learning, we next run a linear regression of college grades on a set of academic and family background characteristics (including race dummies, SAT scores, high school grades, parental education, age dummies, birth year, and whether the individual was working part- or full-time), and compute the residuals. Given that our goal is to further isolate learning from student background characteristics, we compare the mean residual at t for different educational choices at $t+1$ in panels C and D of Table 4. Despite the large set of controls included in the regression, we still find that those with higher grade residuals are more likely to stay in school. Again, while these patterns could still be partly driven by attriters having lower ability, of which they are aware all along but are not measured in the data, they are also consistent with learning about one's ability once in college.

Finally, to illustrate learning in the labor market as a reason for stopouts to return to college, we perform a similar analysis as in Table 4 but now using information on wages. The first panel of Table 5 presents mean log wages for those who have left college, broken out by next-period re-enrollment decision, while the second panel uses the mean difference between actual and expected log wages (i.e., residual) instead of the mean log wage. Overall, this table shows that those who decide to return to school earn lower wages than those who decide to stay in the labor market. For example, those who leave college for the labor force and then choose to return to school show a mean log wage of 2.19, while those who leave but do not return later show a mean log wage of 2.36. This pattern persists even after controlling for a rich set of individual, family background, schooling, and labor market experience variables (see Panel B). While these empirical regularities can be consistent with multiple explanations, the role of learning about labor market productivity in contributing to the decision to return to college is potentially an important one.

In summary, our descriptive analysis is consistent with individuals learning about their own abilities when participating in different sectors. In order to isolate and quantify the empirical relevance of learning, we next present a dynamic structural model of college and labor market decisions where individuals face uncertainty and learn about their academic ability and labor market productivity.

3 Model

3.1 Overview

Motivated by the descriptive patterns in the data, we now turn to our model of education and work decisions. After graduating from high school, individuals in each period from 1 to T^* make a joint schooling and work decision; from periods T^* to T individuals only make work decisions.[11] For those who have not graduated from a four-year college, their schooling options include whether to attend a two-year institution, a four-year institution as a science major, or a four-year institution as a non-science major.[12] Regardless of the schooling choice, individuals can also choose whether to work either full-time or part-time in the blue-collar or white-collar sector.[13] Individuals may always choose blue-collar work but face frictions in obtaining white-collar work. Finally, individuals may choose home production (i.e., neither work nor attending college), implying twenty possible choices for non-graduates. After graduation, schooling options are no longer part of the choice set. Thereafter, agents can only decide between home production and work in the blue-collar or white-collar sectors with different levels of intensity (i.e., full-time or part-time).

Individuals only have imperfect information about their abilities which are characterized by a multidimensional vector of five components. Namely, agents have different abilities for each of the three schooling options (two-year, four-year science, and four-year non-science), and two additional ones corresponding to the different types of labor markets (i.e., blue-collar and white-collar sectors). Throughout the paper, we denote A_i as the five-dimensional ability vector, $A_i \equiv (A_{i2}, A_{i4S}, A_{i4N}, A_{iW}, A_{iB})'$ (simply referred to as *ability* in the following), the elements of which correspond to the ability in two-year college, four-year college science major, four-year college non-science major, white-collar sector and blue-collar sector, respectively.

[11] For estimation, we do not need to take a stand on the values of T^* and T, provided that both are two periods beyond our sample period. See Subsection 3.7 for a detailed explanation of why this is the case.

[12] Recall that a substantial number of individuals entered 4-year colleges having not indicated a major. We treat this as the individual having made a choice of major, but this choice was not revealed in the data.

[13] See, e.g., Keane and Wolpin (2001) and Joensen (2009) who estimate dynamic structural models of schooling and work decisions and also allow for work while in college.

Individuals update their beliefs by receiving signals that depend on their choices: enrolling in school provides signals through grades, and working provides signals through wages. These signals then reveal different information regarding their abilities. Signals in one schooling option may be informative about the individual's abilities in the other schooling options, as well as their productivity in the labor market. Similarly, wages in the blue-collar sector may be informative not only with regard to productivity in this sector but also with regard to the individual's schooling abilities and productivity in the white-collar sector.

Agents are assumed to be forward-looking and choose the sequence of actions yielding the highest value of expected lifetime utility. Hence, when making their schooling and labor market decisions, individuals consider the option value associated with the new information acquired on different choice paths. Individuals who choose to work while in college will get two signals on their abilities and productivities: one through their grades, and one through their wages.

We now detail the main elements of the model. We first discuss the grade and wage equations as grades and wages provide the ability signals. We then describe how individuals update their beliefs, both about their abilities but also about their graduation status, white collar arrival rate, and the state of the labor market. Finally, we model the flow payoffs and the optimization problem the individuals face. Discussions of model identification and estimation are deferred to Sections 4 and 5, respectively.

3.2 Grades

In the following, we denote by $j \in \{2, 4S, 4N\}$ the type of college and major, where 2 denotes a two-year college, $4S$ a four-year college science major, and $4N$ a four-year college non-science major. Individuals are indexed by i.

We assume that grades in the college sector j depend on schooling ability A_{ij}, which is not directly observed by the agents. However, they form some initial beliefs about A_{ij} that are given by the prior distribution $\mathcal{N}(0, \sigma^2_{Aj})$. Grades also depend on a set of covariates for college sector j and period of college enrollment τ,[14] $X_{ij\tau}$, that is known to the individual and includes skill measures such as high school grades, indicators denoting participation in the labor market (i.e., working part-time or full-time), and background characteristics (i.e., age, race, and parental education).[15] In the following and throughout the paper, we assume

[14] τ is defined as the period of college enrollment irrespective of the type of college and major. For instance, someone who completes two years of a community college and then transfers to a four-year college will have $\tau = 3$.

[15] The specification for grades in two-year college also includes an indicator for whether an individual has

that unobserved ability A_i is independent of period-1 characteristics, X_{ij1}.

Grades in two-year colleges and in the first two years of four-year colleges are given by:

$$G_{ij\tau} = \gamma_{0j} + X_{ij\tau}\gamma_{1j} + A_{ij} + \varepsilon_{ij\tau} \tag{1}$$

The idiosyncratic shocks, $\varepsilon_{ij\tau}$, are mutually independent and distributed $\mathcal{N}(0, \sigma^2_{j\tau})$, and are also independent from the other state variables. Define the type-j (college, major) academic index of i in period τ, $AI_{ij\tau}$, as:

$$AI_{ij\tau} \equiv \gamma_{0j} + X_{ij\tau}\gamma_{1j} + A_{ij} \tag{2}$$

The academic index $AI_{ij\tau}$ is given by the expected grades conditional on knowing A_{ij} but not the idiosyncratic shock $\varepsilon_{ij\tau}$ (see Arcidiacono, 2004, for a similar ability index specification).

Finally, for four-year colleges and periods $\tau > 2$, we express grades relative to $AI_{ij\tau}$ as follows:

$$G_{ij\tau} = \lambda_{0j} + \lambda_{1j}AI_{ij\tau} + \varepsilon_{ij\tau} \tag{3}$$

Hence, the return to the academic index is allowed to vary over the periods of college enrollment and across majors. As such, while remaining parsimonious, this specification allows for different effects of ability on grades for lower- and upper-classmen.[16] Grade dynamics may also be different for science and non-science majors.

3.3 A Two-Sector Labor Market

Individuals who choose one of the work options (either full-time or part-time) receive an hourly wage that depends on their graduation status.[17] We assume that there are two sectors in the labor market, which are indexed by l and referred to as *white collar* ($l = W$) and *blue collar* ($l = B$). Workers face search frictions in obtaining employment in the white-collar sector. Each period, with some probability $\tilde{\lambda}_{it}^{(d_{t-1})}$ (which depends on individual characteristics and previous decision), the individual's choice set contains both white-collar

spent more than one year in this type of college.

[16] Arcidiacono, Aucejo, and Spenner (2012) shows, using data from Duke students, that grade distributions are more compressed for upper-year classes. A similar specification is also used in Arcidiacono (2004).

[17] Note that one can think of wages as a measure of performance on the job. As such, we do not need to assume that employers have perfect information about workers' abilities. Instead, we make a spot market assumption implying that workers are paid according to their realized productivity.

and blue-collar work options. With probability $1 - \tilde{\lambda}_{it}^{(d_{t-1})}$, the white collar options are not in the individual's choice set.

Log wages in sector l are allowed to differ based on whether the individual is currently attending college or not. For those who are not enrolled in school while working, log wages in sector l and calendar year t are assumed to depend linearly on sector-specific productivity A_{il}, a set of observed characteristics X_{ilt},[18] labor market conditions δ_t, and idiosyncratic shocks ε_{ilt}:

$$w_{ilt} = \delta_t + \gamma_{0l} + X_{ilt}\gamma_{1l} + A_{il} + \varepsilon_{ilt} \tag{4}$$

The returns to the various components in X_{ilt} are sector-specific. Note that this specification allows for human capital accumulation through schooling as well as on the job. The idiosyncratic shocks, ε_{ilt}, are assumed to be distributed $\mathcal{N}(0, \sigma_l^2)$ and are independent over time as well as across individuals and both sectors, and independent of the other state variables.

Like the model for grades, we define the sector-l productivity index of i in period t as

$$PI_{ilt} = \gamma_{0l} + X_{ilt}\gamma_{1l} + A_{il} \tag{5}$$

Then, for individuals working while in school, we express log wages relative to PI_{ilt} as

$$w_{ilt}^s = \delta_t + \lambda_{0l} + \lambda_{1l} PI_{ilt} + \varepsilon_{ilt}^s \tag{6}$$

where the idiosyncratic shocks, ε_{ilt}^s, are assumed to be distributed $\mathcal{N}(0, \sigma_{sl}^2)$. Hence, the return to the productivity index is allowed to vary over in-school work status and occupational sector. Importantly, such a specification allows for different signal-to-noise ratios for the wages received in and out of college.

Finally, we account for aggregate changes in wages over time through calendar year indicators, δ_t. If we did not control for these nonstationarities, we might falsely conclude that learning about ability is important when, in reality, workers are simply responding to aggregate shocks. The time dummies at t are observed in period t, but individuals must form expectations over this variable for periods $t + 1$ and beyond. We formalize this feature of the model in detail in Section 3.4.2.

[18]These include years of education, graduation, college major if graduated, an indicator for working part-time, labor market experiences, age, demographics, high school grades, and parental education. We assume that individuals who work part-time in a given year and sector accumulate half a year only of work experience in that sector.

3.4 Beliefs

Individuals are uncertain about (i) their future preference shocks, (ii) their schooling ability and labor market productivity, (iii) the evolution of the market shocks (the δ_t's), (iv) (four-year) college graduation, and (v) whether they can participate in the white-collar sector. The first component, future preference shocks, will be discussed in Section 3.5 when we describe preferences. We discuss the other components below.

3.4.1 Beliefs over schooling ability and labor market productivity

We assume that individuals are rational and update their beliefs in a Bayesian fashion. Their initial ability beliefs are given by the population distribution of A, which is supposed to be multivariate normal with mean zero and covariance matrix Δ. Importantly, we do not restrict Δ to be diagonal, thus allowing for correlated learning across the five different ability components.

At each period τ of college attendance, individuals use their realizations of grades and wages (if they work while in college) to update their beliefs about their schooling abilities in all college options $(A_{i2}, A_{i4S}, A_{i4N})$, as well as their labor market productivity in both sectors (A_{iW}, A_{iB}). Grade realizations provide noisy signals regarding abilities, with $S_{ij\tau}$ denoting the signal for individual i from a type-j college option at enrollment period τ. Specifically, for two-year colleges and the first two years of four-year colleges, the signal is given by:

$$S_{ij\tau} = G_{ij\tau} - \gamma_{0j} - X_{ij\tau}\gamma_{1j} \tag{7}$$

For four-year colleges in subsequent enrollment periods ($\tau > 2$), the index specification yields:

$$S_{ij\tau} = \frac{G_{ij\tau} - \lambda_{0j} - \lambda_{1j}(\gamma_{0j} + X_{ij\tau}\gamma_{1j})}{\lambda_{1j}} \tag{8}$$

Similarly, individuals who participate in the labor market update their ability beliefs after receiving their wages. The signal for those not in school and working in sector l and period t is given by:

$$S_{ilt} = w_{ilt} - \delta_t - \gamma_{0l} - X_{ilt}\gamma_{1l} \tag{9}$$

For those enrolled in school while working in sector l in period t, the signal is

$$S_{ilt} = \frac{w_{ilt} - \delta_t - \lambda_{0l} - \lambda_{1l}(\gamma_{0l} + X_{ilt}\gamma_{1l})}{\lambda_{1l}} \tag{10}$$

Finally, individuals may choose to work while in college, in which case they will receive two ability signals instead of one, $(S_{ij\tau}, S_{ilt})$.[19]

To describe the updating rules, we first introduce some notation. Let Ω_{it} be a 5×5 matrix with zeros everywhere except for the diagonal terms corresponding to the choices made by individual i in period t (namely two-year college, four-year college science major, four-year college non-science major, white- or blue-collar work). The diagonal elements corresponding to the choices made are given by the inverse of the variances of the idiosyncratic shocks.[20] The maximum number of positive diagonal elements is two, which corresponds to receiving two signals: one from grades in a particular schooling option and one from wages. Similarly, denote by $\widetilde{S}_{it}$ a 5×1 vector with zeros everywhere except for the elements corresponding to the choices in period t. Here, the non-zero elements are the ability signals received in this period.[21]

It follows from the normality assumptions on the initial prior ability distribution and on the idiosyncratic shocks that the posterior ability distributions are also normally distributed. Specifically, denoting by $E_t(A_i)$ and $\Lambda_t(A_i)$ the posterior ability mean and covariance at the end of period t, we have (see DeGroot, 1970):

$$E_t(A_i) = (\Lambda_{t-1}^{-1}(A_i) + \Omega_{it})^{-1}(\Lambda_{t-1}^{-1}(A_i)E_{t-1}(A_i) + \Omega_{it}\widetilde{S}_{it}) \tag{11}$$

$$\Lambda_t(A_i) = (\Lambda_{t-1}^{-1}(A_i) + \Omega_{it})^{-1} \tag{12}$$

As in canonical one-dimensional learning models, prior variances at the beginning of a given period decrease towards zero as individuals receive additional ability signals in the previous

[19] Once an individual graduates from college, learning in our model occurs through wage signals only.

[20] Note that, for the college options, the idiosyncratic variances will depend on the year of enrollment. For the work options, these variances will depend on school enrollment status.

[21] As an example, someone who chooses to work in the blue-collar sector while pursuing a 4-year non-science degree in $\tau \leq 2$ would have the following values for Ω_{it} and $\widetilde{S}_{it}$:

$$\Omega_{it} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma^2_{4Nt}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sigma^2_{sB}} \end{bmatrix}, \ \widetilde{S}_{it} = \begin{bmatrix} 0 & 0 & A_{i4N} + \varepsilon_{i4Nt} & 0 & A_{iB} + \varepsilon^s_{iBt} \end{bmatrix}'$$

periods, thus giving more weight to the prior ability and less to the current-period signal.

3.4.2 Beliefs over market shocks

We now specify how individuals form their beliefs about the aggregate labor market. Individuals observe the current value of δ_t. We assume that δ_t is the same for both employment sectors. We also assume that the aggregate shock follows an AR(1) process:

$$\delta_t = \phi\delta_{t-1} + \zeta_t \tag{13}$$

where ζ_t is i.i.d. $\mathcal{N}(0, \sigma_\zeta^2)$. The assumption that the aggregate shock follows an AR(1) process, or a discretized version of it (i.e., a Markov process of order 1) is common in the literature (see, e.g., Adda et al., 2010; Robin, 2011). Given the realizations of the δ_{t-1}'s, individuals then integrate over possible realizations of the ζ_t's when forming their expectations over the future.

3.4.3 Beliefs over graduation

We treat graduation as probabilistic. Individuals are only at risk of graduating if they have completed at least two years of college and if they are currently attending a four-year institution. Individuals in this risk set face a probability of graduation at the end of their τ-th period of college enrollment that depends on a set of characteristics $X_{ig\tau}$. This set of characteristics includes time-invariant measures like high school grades and demographics. It also includes time-varying components like years in each type of school (two-year or four-year), current college major, current work decisions, and the individual's prior beliefs about his four-year college abilities in science and non-science majors.[22] We then assume that the probability of graduating conditional on $X_{ig\tau}$ takes a logit form:

$$\Pr(\text{grad}_{i\tau} = 1 | X_{ig\tau}) = \frac{\exp(X_{ig\tau}\psi)}{1 + \exp(X_{ig\tau}\psi)} \tag{14}$$

Individuals are assumed to know the parameters ψ and form expectations over their probabilities of graduating using (14).

[22] While we do allow the prior ability beliefs, and thus prior college grades, to enter the graduation probability, we do not allow unobserved ability A_i itself to affect graduation. If we did so, then individuals could learn about their abilities through graduation realizations. This would substantially complicate our model by requiring, in particular, us to allow for correlated Bayesian learning based on a mixed continuous-discrete distribution of signals.

3.4.4 Beliefs over white-collar job offer arrival

While individuals can always choose to work in the blue-collar sector, they face search frictions associated with participation in the white-collar sector. In particular, white-collar job offers arrive at time t with probability $\tilde{\lambda}_{it}^{d_{t-1}}$ which is specified as follows:

$$\tilde{\lambda}_{it}^{d_{t-1}} = \begin{cases} \frac{\exp(\widetilde{X}_{it}'\delta_\lambda)}{1+\exp(\widetilde{X}_{it}'\delta_\lambda)} & \text{if } i \text{ did not work in the white-collar sector at time } t-1 \\ 1 & \text{if } i \text{ worked in the white-collar sector at time } t-1 \end{cases} \tag{15}$$

where $\widetilde{X}_{it}$ denotes a vector of state variables in period t (i.e., age and an indicator for college graduation in either major), and the superscript d_{t-1} emphasizes the dependence on the prior decision as outlined in (15). Similar to graduation, individuals are assumed to know the parameters δ_λ.

3.5 Flow utilities

We now define the flow payoffs for each of the schooling and work combinations. We denote in what follows the various schooling options by j, where $j \in \{2, 4S, 4N, 0\}$. Turning to the work options, we denote by $k \in \{p, f, 0\}$ part-time and full-time work, respectively, and by $l \in \{B, W\}$ the *blue-collar* ($l = B$) and *white-collar* ($l = W$) sector. The baseline alternative corresponds to the home production option, i.e., no work ($k = 0$) and no school ($j = 0$), which we denote with a slight abuse of notation by $d_{it} = (0, 0, 0)$.

We denote by Z_{1it} the variables that affect the utility of school, Z_{2it} the variables that affect the utility of work, C_{ijklt} the level of consumption associated with the school and work alternative (j, k, l) in period t, and by Z_{0it} the individual characteristics that affect flow payoffs through consumption only. The flow payoff for choice $d_{it} = (j, k, l)$ is assumed to be given by, letting $Z_{it} = (Z_{0it}, Z_{1it}, Z_{2it})$:

$$U_{jkl}(Z_{it}, \varepsilon_{ijklt}) = \alpha_{jkl} + \alpha_C E(U(C_{ijklt})) + Z_{1it}\alpha_j + Z_{2it}(\alpha_k + \alpha_l) + \varepsilon_{ijklt} \tag{16}$$

$$= u_{jkl}(Z_{it}) + \varepsilon_{ijklt} \tag{17}$$

where the idiosyncratic preference shocks ε_{ijklt} are i.i.d. following a Type 1 extreme value distribution.

This specification allows for level shifts of utility based on combinations of activities through α_{jkl}. The associated restriction on preferences here is that the characteristics of the individual do not affect the cost of pursuing two activities simultaneously. For example, certain characteristics may make part-time work more attractive but will affect the latent utility of part-time work in the same way regardless of sector and regardless of whether the individual is also choosing one of the educational options. We discuss below the remaining components of the flow utility payoffs, starting with the utility of consumption.

3.5.1 Consumption

Individuals make decisions in part based on their expected utility of consumption, which is supported by some combination of labor income, parental transfers, educational grants or loans, and the social safety net.

We follow a modified version of Johnson (2013) by assuming that the budget constraint binds in each period (we do not have savings or assets in our model, so we omit these components from Johnson's equation; however, students can take out loans to pay for their education). How consumption at time t for individual i choosing alternative (j, k, l), C_{ijklt}, is determined depends on whether the time period is before or after T^*. Prior to T^*, consumption depends on labor income, W_{ijklt}, and, if the individual chooses one of the college options, parental transfers, PT_{ijt}, educational grants (both need- and merit-based), GR_{ijt}, and educational loans, L_{ijt}, and tuition and fees, TF_j.[23] How parental transfers, grants, and loans are determined is given in Appendix D, with Appendix G detailing loan accumulation.

After T^*, no more educational decisions are made, and the loan repayment process begins. Consumption then depends only on W_{ijklt} and loan repayments, SLR_{it}.[24] Latent consumption is then defined as:

$$C^*_{ijklt} = \begin{cases} W_{ijklt} & \text{if } j = 0, k \neq 0, t < T^* \\ W_{ijklt} + PT_{ijt} + GR_{ijt} + L_{ijt} - TF_j & \text{if } j \neq 0, k \neq 0, t < T^* \\ PT_{ijt} + GR_{ijt} + L_{ijt} - TF_j & \text{if } j \neq 0, k = 0, t < T^* \\ W_{ijklt} - SLR_{it} & \text{if } j = 0, k \neq 0, t \geq T^* \\ 0 & \text{if } j = 0, k = 0 \end{cases} \tag{18}$$

To reflect the social safety net, realized consumption is then given by the maximum of

[23] Because our model has no notion of college choice or geographic space, tuition TF_j is the same for every individual and differs only by college sector (2-year vs. 4-year).

[24] In practice, as long as the loan repayment period starts after the estimation window, we do not need to take a stand on how repayments impact estimation. The loan repayment details are only relevant for our counterfactuals and are discussed in Appendix J.

C^*_{ijklt} and the consumption floor $\underline{C}$:

$$C_{ijklt} = \max(C^*_{ijklt}, \underline{C}) \tag{19}$$

Consumption is evaluated in terms of yearly consumption flow in 1996 dollars. We calibrate a consumption floor for individuals not working and not in school. Namely, we follow Hai and Heckman (2017) and set this value to $\underline{C} = \$2,800$.

We assume that individuals have CRRA preferences over their consumption, with a risk aversion parameter which we set equal to $\theta = 0.4$. We discuss in detail in Appendix D.4 how we compute the expected utility of consumption taking into account the different sources of uncertainty, with Appendix D.5 discussing how we choose the value of θ.

3.5.2 Controls

We now turn to the variables Z_{1it} and Z_{2it}:

- Z_{1it} and Z_{2it} include demographics, family background characteristics, year of birth, measures of academic performance, and controls for the previous choice (to allow for switching costs similar in spirit to Keane and Wolpin, 1997). Our approach is consistent with students accumulating human capital while in college, while also allowing for non-pecuniary payoffs associated with it. In particular, heterogeneity in the consumption value of attending two- and four-year colleges, is captured with the observed individual characteristics.[25]

- Specific to Z_{1it} is the expected ability in schooling option j (which is computed with respect to individual i's prior ability distribution at the beginning of period t), expected graduation probability in the next period, years of college completed, and work intensity (whose effects are allowed to vary by labor market sector).[26] Expected ability is included in Z_{1it}, which can be interpreted either as individuals valuing grades for non-pecuniary reasons or as an inverse proxy for the cost of effort.[27] In our model, students learn passively about their college abilities. We show in Appendix E that it

[25]See Heckman, Humphries, and Veramendi (2018) for an empirical analysis of the non-market returns to education.

[26]Given that the individual's baseline choice is home and its normalized value remains the same before and after college graduation, one needs to include expected graduation probability at $t+1$ in the flow utility to account for any differences in the value of home before and after graduation.

[27]This reduced-form specification has been used in various prior studies in the empirical schooling choice literature. See, for instance, Arcidiacono (2004) and Stinebrickner and Stinebrickner (2012).

is possible to frame our model instead around learning about per-unit study cost.[28]

- Finally, specific to Z_{2it} are indicators for four-year college graduation and for participation in the white-collar sector (when working part-time). Note that although we do not directly measure non-pecuniary returns to schooling, we allow for such returns by including four-year college graduation as a preference shifter for the work alternatives.

The home production sector is chosen as a reference alternative, and we normalize accordingly the corresponding flow utility to zero. The flow utility parameters, therefore, need to be interpreted relative to this alternative.

3.6 The optimization problem

Individuals are forward-looking and choose the sequence of college enrollment and labor market participation decisions yielding the highest present value of expected lifetime utility. The individual chooses $(d_{it})_{t=1\ldots T}$, a combination of schooling and work decisions, to sequentially maximize the discounted sum of payoffs:

$$E\left[\sum_{t=1}^{T}\beta^{t-1}\sum_{j}\sum_{k}\sum_{l}(u_{jkl}(Z_{it})+\varepsilon_{ijklt})1\{d_{it}=(j,k,l)\}1\{\text{Offer}_{ilt}=1\}\right] \tag{20}$$

where $\beta \in (0,1)$ is the discount factor. The expectation is taken with respect to the distribution of the future idiosyncratic preference shocks, the signals associated with the different choice paths, the beliefs over the aggregate shock to wages, the probability of receiving an offer in sector l ($1\{\text{Offer}_{ilt}=1\}$ above), and the probability of graduating.

Let $V_t(Z_{it})$ denote the ex-ante value function at the beginning of period t, that is, the expected discounted sum of current and future payoffs just before the current period idiosyncratic shock is revealed. The conditional value function $v_{jkl}(Z_{it})$ is given by:

$$v_{jkl}(Z_{it}) = u_{jkl}(Z_{it}) + \beta E_t\left[V_{t+1}(Z_{it+1})|Z_{it}, d_{it}=(j,k,l)\right] \tag{21}$$

where the term $E_t[\cdot|\cdot]$ is indexed by t to highlight the fact that this expectation is conditional on the information set of the individual at the beginning of period t, which includes, in particular, the sequence of ability signals received from periods 1 up until $t-1$. Denoting by $\tilde{\lambda}_{it}^{(d_{t-1})}$ the probability of receiving an offer in the white-collar sector in period t given

[28] Absent information on effort itself, we conjecture that a model having students learn about both their schooling abilities and their study costs would not be identified.

previous decision d_{t-1}, and assuming that the ε's are i.i.d. Type 1 extreme value yields the following weighted log-sum formula:

$$v_{jkl}(Z_{it}) = u_{jkl}(Z_{it}) + \beta\tilde{\lambda}^{(jkl)}_{i,t+1}E_t\left[\ln\left(\sum_j\sum_k\sum_l \exp(v_{jkl}(Z_{it+1}))\right)\middle| Z_{it}, d_{it}=(j,k,l), \text{Offer}_{ilt+1}=1\right]$$
$$+\beta\left(1-\tilde{\lambda}^{(jkl)}_{i,t+1}\right)E_t\left[\ln\left(\sum_j\sum_k \exp(v_{jkB}(Z_{it+1}))\right)\middle| Z_{it}, d_{it}=(j,k,l), \text{Offer}_{ilt+1}=0\right]+\beta\gamma \quad (22)$$

where γ denotes Euler's constant.

3.7 Finite dependence

Following Arcidiacono and Miller (2011, 2019), we re-express the future payoffs so we can avoid solving the full backward recursion problem. Note that the expected value function at time $t+1$ can be expressed relative to the conditional value function for one of the choices, plus a (known) function of the choice probabilities. This feature, coupled with the fact that it is the differences in the conditional value functions that are relevant in estimation, makes it possible in some cases to express the estimation problem such that the differenced future utility term only depends on a-few-period-ahead choice probabilities and flow utilities. The key is to choose choice sequences—given two different initial choices—such that the same distribution of states results from both paths.

We now discuss how this concept can be applied in our setup, and defer the formal derivation of the finite dependence property to Appendix F. Denote first by h the triplet of home production choices $(0,0,0)$. Our goal is to find an expression for the difference in the conditional value functions, $v_{jkl} - v_h$, where only a few period ahead flow utilities and conditional choice probabilities are needed to characterize the differenced future value term.[29] This is simple to do for alternatives that do not include working in the white collar sector. To see this, consider the conditional value function at time t for $\{4S, p, B\}$, so that the individual is attending school in the sciences while working part-time in the blue collar sector. Regardless of the transitions of the other state variables, we express the future utility terms at $t+1$ and $t+2$ relative to the conditional value function for the home option, h. Now consider conditional value function at time t associated with h. Expressing the future utility

[29] As a reminder, j denotes the schooling option, k refers to the work options, l determines the sector, and t the period. $v_{000t} \equiv v_{ht}$ refers to home production, which is the baseline alternative.

term at $t+1$ relative to $\{4S, p, B\}$ and at $t+2$ relative to h yields the same distribution of states as the sequence that began with $\{4S, p, B\}$, and hence the same expected future utility at $t+3$:

$$\begin{aligned} E_t\left[V_{t+3}(Z_{t+3})|d_t = \{4S, p, B\}, d_{t+1} = \{0,0,0\}, d_{t+2} = \{0,0,0\}, Z_t\right] &= \\ E_t\left[V_{t+3}(Z_{t+3})|d_t = \{0,0,0\}, d_{t+1} = \{4S, p, B\}, d_{t+2} = \{0,0,0\}, Z_t\right] & \end{aligned} \tag{23}$$

In both cases, the lagged choice (d_{t+2}) is home, the individual will have an additional half-year of work experience and an additional year of science education, and with the choice ordering also not affecting the ability signal distributions, graduation probabilities, and market state.

The problem becomes more complicated for choices involving the white collar sector as there is no guarantee that an individual choosing home will receive a white collar offer in the next period. Consider the same conditional value function as before, but replacing the blue collar sector with the white collar sector: $\{4S, p, W\}$. We again normalize the future utility at time $t+1$ and $t+2$ relative to h for the path that begins with $\{4S, p, W\}$. To ensure that the future value at time $t+3$ from this path will cancel with the corresponding term in the conditional value function for h, note that the future value term along the path that begins with h can be expressed relative to any linear combination of future decisions as long as the weights on those decisions sum to one (Arcidiacono and Miller, 2019). Specifically, given that white-collar job offers arrive with probability $\tilde{\lambda}_{t+1}^{(h)}$ after making decision h at time t, along the path where the individual receives a white collar offer we place a weight $1/\tilde{\lambda}_{t+1}^{(h)}$ on the alternative $\{4S, p, W\}$ and a weight $1-(1/\tilde{\lambda}_{t+1}^{(h)})$ on the home alternative. Along the path where the individual does not receive an offer at $t+1$, we express the future value term relative to the home alternative h. In Appendix F, we show that this weighting scheme results in the same distribution of states at $t+3$ that would result from making the sequence of choices $\{4S, p, W\}$ and then expressing the future value term relative to home at $t+2$ and $t+3$. It follows that two-period finite dependence can be achieved, even though whether the individual receives a white collar offer is uncertain.

4 Identification

We first discuss in Subsection 4.1 how our model accommodates permanent unobserved heterogeneity before turning in Subsection 4.2 to the identification of the model parameters.

4.1 Unobserved heterogeneity and measurement system

The presence of unobserved preferences and skills (to the econometrician, but known by the agents) that could be correlated over time constitute a threat to identifying the coefficients of interest. We address this issue by allowing for unobserved heterogeneity types in the spirit of Heckman and Singer (1984) and Keane and Wolpin (1997). Namely, we include type-specific components, capturing permanent characteristics of the individual, which enter the model in the form of location shifters. These type-specific shifters are allowed to affect the outcomes (college grades and log-wages), college graduation probability, white-collar job offer arrival rate and flow utilities associated with the different choice alternatives. We allow for three unobserved binary heterogeneity factors, namely one for schooling ability, one for schooling preferences, and one for work motivation. It follows that the agents in our model are supposed to belong to one of eight unobserved heterogeneity types, where each of the three latent factors can take on one of two possible values, low (L) or high (H).[30] We assume that (initially unknown) unobserved ability and unobserved heterogeneity types are mutually independent, and that they are also independent from the observed covariates at period $t = 1$. In practice we rely on an auxiliary measurement system, which is primarily used to identify the distribution of permanent unobserved heterogeneity. Beyond identification, another advantage of using these measurements is that they render greater interpretability to the heterogeneity types (Carneiro, Hansen, and Heckman, 2003). Appendix C describes the measurements we use and how each one of them relates to the unobserved heterogeneity types.

4.2 Model identification

We now discuss how the model parameters are identified. Identification proceeds sequentially. We identify in a first step the unobserved heterogeneity distribution, the conditional choice probabilities, the outcome equations and the distribution of the agents' beliefs. The structural utility parameters are then identified in a second step. We end this subsection by explaining how white-collar job offers are separately identified from preferences for white-collar work.

[30] The possible combinations for schooling ability, schooling preferences, and work motivation that lead to eight types are as follows: (H,H,H), (H,H,L), (H,L,H), (H,L,L), (L,H,H), (L,H,L), (L,L,H), and (L,L,L).

4.2.1 Unobserved heterogeneity distributions, conditional choice probabilities, continuous outcome equations and agents' beliefs

Distribution of unobserved permanent heterogeneity that is known to the agent Grades and log wages are linear functions of (*i*) observed covariates, (*ii*) type-specific unobserved heterogeneity (known to the agents), (*iii*) unobserved heterogeneity initially unknown to the agents, and (*iv*) idiosyncratic shocks. Although we assume that the random factors (*ii*) and (*iii*) are mutually independent, these factors are generally correlated once we condition on the outcome being observed by the analyst. For instance, grades in four-year college science major are only observed for students who enrolled in this specific type of college and major, a decision which likely depends on their (known) unobserved heterogeneity as well as their beliefs about their initially unknown unobserved heterogeneity. It follows that one cannot directly apply the identification arguments from, e.g., Carneiro, Hansen, and Heckman (2003) or Heckman and Navarro (2007), to identify the distribution of the factors from the realized grades and log wages.

Instead, we use the auxiliary measurement system discussed in Subsection 4.1 for identification. Namely, we rely on the fact that we have access to a set of selection-free measurements for the three unobserved discrete heterogeneity factors. This then allows us to use existing identification results for finite mixture models with multiple measurements. Specifically, one can apply Theorem 8 of Allman, Matias, and Rhodes (2009) to our context to identify the distribution of the unobserved heterogeneity types, along with the distributions of the measurements conditional on the heterogeneity types.

Type-specific conditional choice probabilities The choice probabilities conditional on the observed state variables and the measurements, which are directly identified from the data, can be expressed as a finite mixture over the heterogeneity types of the type-specific conditional choice probabilities. The mixture weights are identified from the previous step. Key to the identification of the type-specific conditional choice probabilities is then the assumption that measurements are independent from choices once we condition on heterogeneity types.[31] Conditional on the vector of observed state variables, the finite mixture model can then trivially be expressed as a linear system with N_M (number of distinct values taken by the vector of measurements) equations and N_T (number of points of support of the vector of types) unknowns. Provided that the measurements are relevant measurements of

[31] It follows that measurements only affect the choice probabilities through the mixture weights. See Henry, Kitamura, and Salanie (2014), who use similar exclusion restrictions to (partially) identify finite mixture models in a more general context with unknown mixture weights.

types (i.e., the type distribution conditional on measurements is a non-trivial function of at least a subset of the measurements), the type-specific choice probabilities will generally be over-identified in our context.

To fix ideas, consider the case of the type-specific conditional choice probabilities in the first period, assuming one unobserved heterogeneity factor with two points of support (i.e., two types), $R \in \{1, 2\}$. While simpler than our specification, this setup remains rich enough to convey the main identification arguments. Under the previous assumptions, the choice probabilities conditional on the measurements $M = m$, $\Pr(D|M = m)$, can be written as a finite mixture of the type-specific conditional choice probabilities, where the mixture weights, $\pi_1^{M=m}$ and $\pi_2^{M=m} = 1 - \pi_1^{M=m}$, are identified from the earlier step (we omit the conditioning on the observed covariates for simplicity here). Namely:

$$\Pr(D|M = m) = \Pr(D|M = m, R = 1)\pi_1^{M=m} + \Pr(D|M = m, R = 2)(1 - \pi_1^{M=m}) \tag{24}$$

Evaluating the choice probabilities conditional on measurements at two different points m_1 and m_2 in the support of the measurements M yields the following linear system:

$$\begin{pmatrix} \Pr(D|M = m_1) \\ \Pr(D|M = m_2) \end{pmatrix} = \begin{pmatrix} \pi_1^{M=m_1} & 1 - \pi_1^{M=m_1} \\ \pi_1^{M=m_2} & 1 - \pi_1^{M=m_2} \end{pmatrix} \begin{pmatrix} \Pr(D|R = 1) \\ \Pr(D|R = 2) \end{pmatrix} \tag{25}$$

This identifies the vector of type-specific conditional choice probabilities $\begin{pmatrix} \Pr(D|R = 1) \\ \Pr(D|R = 2) \end{pmatrix}$ provided that the rank condition $\pi_1^{M=m_1} \neq \pi_1^{M=m_2}$ holds.

Similar arguments can be used to establish identification of the type-specific conditional choice probabilities in the more general case of a vector of unobserved discrete heterogeneity factors.[32]

Outcome equation parameters (grades and log wages) An important feature of our model is that grades and wages are all subject to sample selection. For instance, grades in four-year science are, by definition, only observed for those who chose to enroll in a four-year college, science major. The same holds true for wages in blue- and white-collar sectors. Importantly though, *conditional on the heterogeneity type*, grades and wages are subject to

[32] For periods $t > 1$, the mixture weights are given by the type probabilities conditional on the measurements and the state variables, which, unlike for the first period, include prior choices and ability beliefs. Identification of these weights exploits the normality assumption on the outcome equations (grades and log wages).

selection on observables only. It follows that identification of the outcome equation parameters can be established using a control function approach. Identification of the outcome equation parameters proceeds along these two steps:

Step 1: Identification of the distribution of the outcome Y_τ^j conditional on heterogeneity type (R), past schooling/work choices ($D_{\tau-1} = j, D_{\tau-2}...$), past ability signals ($S_{\tau-1}, S_{\tau-2}...$), and other state variables $X_{\tau-1}^j$

This follows from a similar reasoning as for the type-specific conditional choice probabilities, after replacing the conditional choice probabilities by the densities of the outcomes Y_τ^j conditional on past choices, past outcomes, and other state variables.

Step 2: Identification of the outcome equation parameters

These parameters are identified from the theoretical regression of $E(Y_\tau^j|R, D_{\tau-1} = j, D_{\tau-2}, ..., S_{\tau-1}, S_{\tau-2}, ..., X_{\tau-1}^j)$, which is identified from the previous step, on R, $X_{\tau-1}^j$, the interaction terms $D_{\tau-2}S_{\tau-1}$, $D_{\tau-2}D_{\tau-3}S_{\tau-2}$, ..., and an intercept. Identification of the ability signals S_τ follows.

Distribution of unobserved ability factors $A = (A_2, A_{4S}, A_{4N}, A_W, A_B)'$ that are initially unknown to the agent The joint distribution of the unobserved factors A is identified primarily from the covariances between the outcomes (namely, grades and log wages) and past ability signals.

In our model, conditional on observed covariates, both type-specific unobserved heterogeneity and unobserved factors A generate persistence in the outcomes. The assumption that the distribution of unobserved heterogeneity known by the agent is discrete helps identifying the distribution of A. In particular, denoting by Y the outcomes, S the past ability signals, X the observed covariates, R the heterogeneity type dummy which follows a Bernoulli distribution (e.g. schooling ability type for grades, where we assume here to simplify the exposition that there are only two types), and D the selection dummy, $\text{Cov}(Y, S|X, D = 1)$ as well as $\text{Cov}(Y^2, S|X, D = 1)$ are both functions of the same covariance term $\text{Cov}(R, S|X, D = 1) = \text{Cov}(R^2, S|X, D = 1)$. The share of the covariance $\text{Cov}(Y, S|X = x, D = 1)$ that is attributable to the unobserved ability A is then identified from the covariances $\text{Cov}(Y, S|X = x, D = 1)$ and $\text{Cov}(Y^2, S|X = x, D = 1)$, letting x vary on the support of X. This, in turn, identifies the signal-to-noise ratios associated with the different ability signals, along with the variance-covariance matrix of A.

Individual beliefs about unobserved abilities Individual prior beliefs about abilities A, at the beginning of any given period, are identified as a byproduct of the previous steps, combined with the maintained normality assumption for the distribution of the ability signals.

4.2.2 Offer arrival rates and utility function parameters

We now discuss the identification of the conditional value functions before turning to the flow payoffs. As these arguments take as given the conditional choice probabilities, we start by discussing the identification of the white-collar job arrival rates and probabilities of choosing to work in a white-collar job.

White-collar job arrival rates and choices First, note that individuals with arbitrarily large prior abilities in the white-collar sector are predicted to choose to work in the white-collar sector—conditional on receiving an offer in that sector—with a probability approaching one.[33] In other words, it follows from the specification of our choice model that:

$$\lim_{y\to\infty} \Pr(D_t = W|\text{Offer}_t = 1, Z_t, E_{t-1}(A_{-W}), E_{t-1}(A_W) = y, R) = 1 \tag{26}$$

where we denote by $E_{t-1}(A_{-W})$ the vector of prior abilities for all three college options and the blue-collar sector, and we suppress the individual subscripts to ease the notational burden.

It follows that, for this subset of individuals, the offer arrival rates are equal to the probabilities of working in the white-collar sector unconditional on receiving an offer in that sector, namely $\Pr(D_t = W|Z_t, E_{t-1}(A_{-W}), E_{t-1}(A_W), R)$. These probabilities are identified for this subset of individuals in a similar fashion to the type-specific conditional choice probabilities discussed in the previous subsection. Under the maintained assumption that the offer arrival rates do not depend on prior ability, this, in turn, identifies the arrival rates, which we denote here by $\widetilde{\lambda}_t^R$ to make the dependence on R explicit. Finally, it follows that the probabilities of working in the white-collar sector conditional on receiving an offer, which are given by $\Pr(D_t = W|\text{Offer}_t = 1, Z_t, E_{t-1}(A_{-W}), E_{t-1}(A_W), R) = \Pr(D_t = W|Z_t, E_{t-1}(A_{-W}), E_{t-1}(A_W), R)/\widetilde{\lambda}_t^R$, are identified for the whole population. In practice, separate identification of the white-collar job offer arrival rates and the probabilities of choosing to work in the white-collar sector conditional on receiving an offer in that sector

[33] Implicit here is the assumption that prior ability, and thus expected wages in the white-collar sector are positively associated with the flow utility for holding a white-collar job.

is further facilitated by our parametric and distributional assumptions, combined with additional exclusion restrictions between the offer arrival rates and the flow utility of working in the white-collar sector (see Table B2 in Appendix B).

Conditional value functions Having identified the type-specific conditional choice probabilities, one can then identify the conditional value functions associated with each alternative using standard identification arguments from the dynamic discrete choice literature.[34] In particular, it follows from the assumption that the idiosyncratic preference shocks are drawn from a Type 1 extreme value distribution that the conditional value functions—up to a reference alternative (home production)—are identified by inverting the type-specific conditional choice probabilities as in Hotz and Miller (1993). This identification result does require knowledge of the distribution of the preference shocks. However, similar inversion results can be obtained for a more general class of error distributions, including generalized extreme value distributions (see, e.g., Arcidiacono and Miller, 2011, Chiong, Galichon, and Shum, 2013).

Flow utilities Finally, having identified the type-specific conditional choice probabilities and conditional value functions from the earlier steps, the flow utilities associated with the different alternatives are identified by applying to our context the results of Arcidiacono and Miller (2015)[35] on utility functions that do not depend on time itself (see Subsection 4.2 in Arcidiacono and Miller, 2015).[36]

5 Estimation

For expositional reasons, we first present the estimation procedure for a specification without type-specific unobserved heterogeneity (Subsections 5.1-5.4). We then discuss how the

[34]Note that it follows from the final paragraph of Section 4.2.1 that the posterior ability vector can be treated at this stage as an observed continuous state variable.

[35]This is a working paper version of Arcidiacono and Miller (2020) available at `http://comlabgames.com/ramiller/working_papers/identification-manuscriptJune2015pa.pdf`.

[36]This identification result relies on a rank condition on the choice-specific transition matrices. An alternative approach to showing identification of the flow utilities is to note that, with our finite dependence representation, the expressions for the differenced conditional value functions have, at this stage of identification, only the non-home utilities for one choice as the unknowns and where the the equation is linear in the unknowns. Using the differences in log odds between the choice and home coupled with coverage of the log odds for every state yields a linear system.

procedure can be extended to allow for unobserved heterogeneity types (Subsection 5.5).[37]

5.1 Additive separability

Assuming that the idiosyncratic shocks are mutually and serially uncorrelated and in the absence of type-specific unobserved heterogeneity, the model can be estimated sequentially. Broadly speaking, the estimation, in this case, proceeds in two key stages. In the first stage, one can estimate the parameters from the grade and wage processes in addition to the choice probabilities associated with all schooling and work alternatives, while the second stage is devoted to estimating the flow utility parameters, taking as given the first-stage estimates.[38] The validity of this sequential approach rests on the likelihood being separable in the contributions of the choices and outcomes.

Namely, consider the case of an individual i attending college for T_c periods, who participates in the blue-collar (white-collar) labor market for T_B (T_W) periods, and for whom we observe a sequence of T_d decisions. We write the individual contributions to the likelihood of the grades, log wages and choices by integrating out the unobserved ability terms $A = (A_2, A_{4S}, A_{4N}, A_W, A_B)'$. This breaks down the dependence across the grades, log wages, choices, and between these variables. The contribution to the likelihood then writes, denoting by $(G_{i\tau})_\tau$ ($\tau \in \{1, \dots, T_c\}$) the grades, $(w_{iB\tau})_\tau$ ($\tau \in \{1, \dots, T_B\}$) the log wages in the blue-collar sector, $(w_{iW\tau})_\tau$ ($\tau \in \{1, \dots, T_W\}$) the log wages in the white-collar sector, and $(d_{i\tau})_\tau$ the decisions ($\tau \in \{1, \dots, T_d\}$), as a five-dimensional integral:

$$\begin{aligned} & L(d_{i1}, \dots, d_{iT_d}, G_{i1}, \dots, G_{iT_c}, w_{iB1}, \dots, w_{iBT_B}, w_{iW1}, \dots, w_{iWT_W}) \\ = & \int L(d_{i1}, \dots, d_{iT_d}, G_{i1}, \dots, G_{iT_c}, w_{iB1}, \dots, w_{iBT_B}, w_{iW1}, \dots, w_{iWT_W} | A) \varphi(A) dA \end{aligned} \tag{27}$$

where $\varphi(\cdot)$ denotes the pdf of the unobserved ability distribution, which is $\mathcal{N}(0, \Delta)$.

From the law of successive conditioning, and using the fact that schooling and work choices depend on ability A only through the observed sequence of signals, we obtain the following partially separable expression (using y as a shorthand for the vector of grades and

[37]Note that throughout this section, we keep the conditioning on the observed covariates implicit to save on notation.

[38]We refer to the first stage as the estimation steps that are necessary to complete before estimating the structural parameters. Note that this stage also includes estimating the graduation and search friction parameters, consumption-related inputs (i.e., loans, grants, and transfers), flexible conditional choice probabilities (CCPs), and a measurement system to account for unobserved heterogeneity. We will discuss these remaining elements later to ease the exposition. See Appendix Table B1 for a full summary of the estimation steps.

log wages):

$$L(d_{i1},\ldots,d_{iT_d},G_{i1},\ldots,G_{iT_c},w_{iB1},\ldots,w_{iBT_B},w_{iW1},\ldots,w_{iWT_W}) = L_{id}\times L_{iy} \tag{28}$$

where the contribution of the sequence of schooling and work decisions is given by:

$$\begin{aligned} L_{id} &= L(d_{i1})L(d_{i2}|d_{i1},G_{i1})\cdots \\ &\quad \times L(d_{iT_d}|d_{i1},d_{i2},\ldots,d_{iT_d-1},G_{i1},G_{i2},\ldots w_{iB1},w_{iB2},\ldots,w_{iW1},w_{iW2},\ldots) \end{aligned} \tag{29}$$

This corresponds to the product over T_d periods of the Type 1 extreme value choice probabilities obtained from the dynamic discrete choice model.

Finally, the contribution of the sequence of grades, blue-collar and white-collar log wages is given by:

$$\begin{aligned} L_{iy} \;=\; & \int L(G_{i1}|d_{i1},A)\cdots L(G_{iT_c}|d_{i1},d_{i2},\ldots,A)L(w_{iB1}|d_{i1},A)\cdots L(w_{iBT_B}|d_{i1},d_{i2},\ldots,A) \\ & \times L(w_{iW1}|d_{i1},A)\cdots L(w_{iWT_W}|d_{i1},d_{i2},\ldots,A)\varphi(A)dA \end{aligned} \tag{30}$$

where $(L(w_{iB\tau}|d_{i1},\ldots,A))_\tau$, $(L(w_{iW\tau}|d_{i1},\ldots,A))_\tau$, and $(L(G_{i\tau}|d_{i1},\ldots,A))_\tau$ respectively denote the normal pdf's of the blue- and white-collar log wages as well as the college grade distributions, all conditional on the ability A and the sequence of choices. Taking logs of (28) results in the choice part of the log-likelihood being additively separable from the outcome part of the log-likelihood.[39]

5.2 Estimation of grade and wage parameters

Estimation of the parameters of the grade and wage equations proceeds as follows. Instead of directly maximizing the likelihood of the outcomes, which would be computationally costly because of the ability integration, we compute the parameter estimates using the EM algorithm (Dempster, Laird, and Rubin, 1977). The estimation procedure iterates over the following two steps until convergence:[40]

- E-step: update the posterior ability distribution from all the observed outcome data (log wages and grades), using the outcome equation parameters obtained from the

[39] With type-specific unobserved heterogeneity, the log-likelihood is no longer additively separable. However, applying the EM algorithm restores the additive separability at the maximization step (Arcidiacono and Jones, 2003). See Section 5.5 for more discussion.

[40] In this context, the EM algorithm is guaranteed to converge to a local optimum.

previous iteration. This follows from the Bayesian updating formulas for the posterior ability mean and covariance given in Section 3.4.1 (Equation 11). The (population) covariance matrix of the ability distribution is then updated as follows for each iteration n of the EM estimation:

$$\Delta^{(n)} = \frac{1}{N} \sum_{i=1}^{N} \left(\Lambda_i^{(n)}(A) + E_i^{(n)}(A) E_i^{(n)}(A)' \right) \tag{31}$$

where N denotes the number of individuals in the sample, $E_i^{(n)}(A)$ the posterior ability mean ($E_i^{(n)}(A)'$ its transpose) and $\Lambda_i^{(n)}(A)$ the posterior ability covariance computed at the beginning of the E-step.

- M-step: given the posterior ability distribution obtained at the E-step, maximize the expected complete log-likelihood of the outcome data, which is separable across sectors (two-year college, four-year college science major, four-year college non-science major, blue-collar and white-collar labor).

Namely, at the M-step of each iteration n of the EM estimation, denoting by $\varphi_i^{(n)}(\cdot)$ the pdf of the posterior ability distribution computed at the E-step, we maximize the expected complete log-likelihood $E\ell_i^{(n)}$:

$$\begin{aligned} E\ell_i^{(n)} &= \int \ln \left[L(G_{i1}|d_{i1}, A) \cdots L(G_{iT_c}|d_{i1}, d_{i2}, \ldots, A) L(w_{iB1}|d_{i1}, A) \cdots \right. \\ &\qquad \left. \times\ L(w_{iBT_B}|d_{i1}, d_{i2}, \ldots, A) \right] \varphi_i^{(n)}(A) dA \\ &= E\ell_{i,2}^{(n)} + E\ell_{i,4S}^{(n)} + E\ell_{i,4N}^{(n)} + E\ell_{i,W}^{(n)} + E\ell_{i,B}^{(n)} \end{aligned} \tag{32}$$

For instance, the parameters of the grade equation for four-year college sector j are updated by maximizing the contribution $E\ell_{i,j}^{(n)}$, $j \in \{4S, 4N\}$, which writes, denoting by $\varphi_{ij}^{(n)}(\cdot)$ the pdf of the posterior distribution of A_j:

$$E\ell_{i,j}^{(n)} = \int \left\{ \ln \left[L(G_{i,j,1}|d_{i1}, A_j) \right] + \cdots + \ln \left[L(G_{i,j,T_j}|d_{i1}, d_{i2}, \ldots, A_j) \right] \right\} \varphi_{ij}^{(n)}(A_j) dA_j \tag{33}$$

Note that this term is additively separable over time. For any given period τ of participation in college sector j, it follows from the normality assumptions on the idiosyncratic grade shocks and the unobserved ability that:

$$
\int \ln(L(G_{ij\tau}|d_{i1}, d_{i2}, \ldots, A_j))\varphi_{ij}^{(n)}(A_j)dA_j = \\
-\frac{1}{2}\ln(2\pi\sigma_{j\tau}^2) - \frac{1}{2\sigma_{j\tau}^2}\left(\lambda_{1j}^2\Lambda_{ijj}^{(n)}(A) + (G_{ij\tau} - \lambda_{0j} - \lambda_{1j}AI_{ij\tau}^{(n)})^2\right) \tag{34}
$$

where $j \in \{4S, 4N\}$, $\Lambda_{ijj}^{(n)}(A)$ denotes the posterior variance of the college-j ability (computed at the E-step), $AI_{ij\tau}^{(n)} = \gamma_{0j} + X_{ij\tau}\gamma_{1j} + E_{ij}^{(n)}(A)$ is the posterior mean of the ability index in college and major j, and $\varphi_{ij}^{(n)}(\cdot)$ denotes the pdf of the posterior distribution of A_j. It follows that the parameters $(\gamma_{0j}, \gamma_{1j}, \lambda_{0j}, \lambda_{1j}, (\sigma_{j\tau}^2)_\tau)$ are updated by solving the following minimization problem:

$$
\min \sum_{i,\tau}\left(\ln(\sigma_{j\tau}^2) + \frac{1}{\sigma_{j\tau}^2}\left(\lambda_{1j\tau}^2\Lambda_{ijj}^{(n)}(A) + (G_{ij\tau} - \lambda_{0j\tau} - \lambda_{1j\tau}AI_{ij\tau}^{(n)})^2\right)\right) \tag{35}
$$

where $(\lambda_{0j\tau}, \lambda_{1j\tau}) = (0, 1)$ for $\tau \leq 2$ and $j \in \{4S, 4N\}$, and $(\lambda_{0j\tau}, \lambda_{1j\tau}) = (\lambda_{0j}, \lambda_{1j})$ otherwise. The estimation of the two-year college parameters proceeds in a similar fashion, with $(\lambda_{02\tau}, \lambda_{12\tau}) = (0, 1)$ for all τ.

The estimation of the wage parameters in (6) proceeds in a similar manner, with two main differences: (i) the idiosyncratic wage shock variances σ_{sl}^2 do not differ across time periods τ, but instead differ across in-school work status s; and (ii) the aggregate labor market shocks δ_t are common across both l sectors. As a result of (ii), we estimate the wage equation parameters jointly across sectors, adapting the loss function in (35) by taking the sum across sectors, too.

5.3 Estimation of the graduation, search friction parameters and aggregate shock process

Under the assumption that the graduation probabilities take a logit form, we use individual data pooled over time on college graduation and on the set of characteristics $X_{ig\tau}$ and estimate via maximum likelihood the parameters ψ governing the graduation probabilities (see Equation 14, Section 3.4.3).

Turning to the search friction parameters, we treat the arrival of a job offer in the white-collar sector in any given period as a latent variable, and estimate the parameters governing the logit probabilities of receiving an offer (see Equation 15, Section 3.4.4) as discussed in detail in Appendix H.

Finally, under the Normal AR(1) specification discussed in Subsection 3.4.2 for the labor market shocks, we estimate the parameters ϕ and σ_ζ^2 in (13) by maximum likelihood, using the estimated values of δ_t as data.[41]

In the absence of type-specific unobserved heterogeneity, each of these sets of parameters can be consistently estimated separately from all of the other parameters of the model. We discuss in Subsection 5.5 below how the estimation procedure needs to be adjusted to accommodate type-specific unobserved heterogeneity.

5.4 Estimation of the flow payoffs

With the estimates of the grade, wage, search friction, and graduation parameters taken as given, we estimate the flow payoffs in a second stage. Estimation relies on the finite dependence property of our model (see Subsection 3.7).

Specifically, estimation of the flow utility parameters involves the following steps:

1. Estimate the CCPs via a flexible multinomial logit model in a first stage.
2. Calculate the expected differenced future value terms along the finite dependence paths.
3. Estimate the flow utility parameters after expressing the future value function as a function of the CCPs. Having estimated the CCPs in a first step, this simply amounts to estimating a multinomial logit with an offset term.

Applying CCP methods to our model is key to making our model computationally feasible. With five-dimensional unobserved ability, plus the integration over the aggregate labor market shocks, graduation, and job offer arrival events, solving this type of multi-armed bandit model by backward recursion would be computationally prohibitive.[42] By using the finite dependence property of our model, we only need to integrate out over the future shocks for two periods.[43]

[41] Because of the finite dependence property of the model we only need to use the AR(1) assumption three periods ahead when computing the future value terms.

[42] Additionally, estimation requires integration over the stochastic components of the CRRA utility of consumption. We provide complete details in Appendix D.4.

[43] Another advantage of applying this approach is that we do not have to make assumptions about beliefs far out into the future: everything about the future is captured in the conditional choice probabilities. Note, however, that conducting counterfactuals requires more assumptions as in this case we do not have counterfactual data and hence do not observe the conditional choice probabilities.

5.5 Estimation with type-specific unobserved heterogeneity

We account for permanent heterogeneity, unobserved to the econometrician but known to the individuals, by assuming that individuals belong to one of R heterogeneity types, where the heterogeneity type is orthogonal to the covariates at $t = 1$. To this end, we augment our model with a measurement system that is used to define and identify three unobserved binary heterogeneity factors—one for schooling ability, one for schooling preferences, and one for work motivation—resulting in $R = 8$ heterogeneity types. Accounting for type-specific unobserved heterogeneity breaks down the separability between the choice and outcome components of the likelihood described above. The full log-likelihood function when accounting for unobserved heterogeneity is:

$$\ell = \sum_i \ln \left[\sum_{r=1}^{R} \pi_r L_{imr} L_{idr} L_{ibr} L_{iyr} \right] \tag{36}$$

where π_r denotes the population probability of being of type r, and L_{imr}, L_{idr}, L_{ibr}, and L_{iyr} respectively denote individual i's contribution to the likelihood of (i) the measurement system, (ii) the choices, aggregate market shocks, and white-collar offer arrival, (iii) four-year college graduation outcomes, and (iv) grade and wage outcomes $y = (G, w_B, w_W)'$, all conditional on the unobserved heterogeneity type r.[44]

Estimation in the presence of unobserved heterogeneity proceeds as follows. We estimate in a first step the measurement system that we combine with a flexible parametric specification for the outcomes that enter the structural model, namely choices, log-wages, grades, and four-year college graduation (see Appendix C for a detailed discussion). This first estimation stage yields estimates of the posterior probabilities of being of each type (q_{ir}) as well as estimates of the π_r's.

After recovering the posterior probabilities of being of each type, we then use them as weights when estimating the learning parameters, graduation probabilities, the aggregate labor market time series process, CCPs (including white collar job offer arrival parameters), and the structural flow utility parameters. Note that this part is identical to the case without unobserved heterogeneity discussed above, except that, instead of a simple log-likelihood,

[44] In practice, in our data, grades and majors are each missing at a non-trivial rate. We address this issue by treating the first instance of missing grades or major as another discrete latent variable, effectively approximating the distribution of missing grades with a distribution with four support points (see Appendix C.1 for more details). The distribution of unobserved grades and majors (conditional on each heterogeneity type) are also estimated in the first stage of our procedure. This results in a finite mixture model with $R \times 2 \times 4 = 64$ points of support.

we are now maximizing the expectation (over the posterior distribution of heterogeneity types) of the type-specific log-likelihood. This two-step estimation procedure follows in spirit Arcidiacono and Miller (2011), with the use of an EM algorithm allowing us to restore the additive separability of the log-likelihood function despite the presence of unobserved heterogeneity. We summarize our complete estimation procedure in Table B1.

Finally, standard errors are estimated using a parametric bootstrap procedure with 150 replications.[45] We discuss this procedure in detail in Appendix I.

6 Results

In this section, we present our estimation results, show how our model fits the data, and discuss the model-implied sorting patterns on unobserved ability. We focus our attention on the parameters of the grade and wage equations as well as the parameters of the flow payoffs. We estimate the model allowing for eight unobserved types. Estimation results of the measurement system for the unobserved types are given in Tables B5–B7 of Appendix B. Table B3 shows the sign and significance of all unobserved type coefficients across all equations of the model. The parameters governing the probability of graduating college and the probability of receiving a white collar offer are also reported in Appendix B, respectively in Tables B9 and B10.

6.1 Grade parameters

The parameter estimates for the grade equations are presented in Table 6. High school grades are positively associated with college grades for each of the schooling options—strongest for 2-year college, less so for non-science majors, and least so for science majors. Working full-time hinders performance, and this is especially true for the four-year options.

Turning to the type-specific unobserved ability (known to the agent), those with a high schooling preference have significantly higher grades for both 4-year options. Those with a high schooling ability also see higher grades in the 4-year options but the effects are smaller. The fact that unobserved ability (of which test scores are a noisy measure) loads less on college grades in non-science fields (where high school grades load more) is consistent with

[45] For finite mixture models like the one estimated in this paper, the asymptotic variance tends to be a particularly poor approximation with typical sample sizes. See McLachlan and Peel (2004, Section 2.16) who recommend the use of parametric bootstrap in this context.

Ahn et al. (2023).

6.2 Wage parameters

Estimates of the wage equations are given in Table 7.[46] Each year of work experience in the blue-collar sector corresponds to roughly a three-and-a-half-percent increase in earnings in both the blue-collar and white-collar sectors. Returns to white-collar experience are higher—1 percentage point higher in the blue-collar sector and 2 percentage points higher in the white-collar sector.

Returns to schooling are also higher in the white-collar sector. For each year of schooling (up to a maximum of four), workers see 3.8 (6.3) percent higher earnings in the blue (white) collar sector. On top of that, workers who graduate college in the non-sciences see an earnings premium of 6.1 (11.3) percent in the blue (white) collar sector. Graduating in the sciences is even more lucrative with a wage increase of 9.1 (13.2) percent in the blue (white) collar sector. All else equal, the total premium for a non-science graduate relative to someone with no college, given the same experience, is then respectively 21 and 37 percent in the blue- and white-collar sectors; the equivalent premia for science graduates are 30 and 50 percent. A first takeaway from these estimates is the existence of sizable returns to graduating from a science relative to non-science major, consistent with recent empirical evidence on this question (see, e.g., Mountjoy and Hickman, 2021, and Altonji, Arcidiacono, and Maurel, 2016 for a survey). Another takeaway is the existence of a large penalty for college graduates working in the blue- rather than in the white-collar sector, a finding in line with the overeducation literature (see, e.g., Clark, Joubert, and Maurel, 2017; Shephard and Sidibe, 2019). Our results further point to a noticeable interaction between college major and labor market sector, with the earnings advantage of science relative to non-science majors being higher in the white-collar sector (13 vs. 9 percent).

Our estimation results also indicate that Black workers see lower earnings in both sectors, though the penalty is lower in the white collar sector, consistent with Bjerk (2007). The effects of other background measures (HS grades, parent graduated from college, and Hispanic) are smaller and insignificant in both sectors.

Turning next to the portion of the unobserved heterogeneity that is known by the individuals, type-specific heterogeneity parameters suggest that the low-work-motivation as well as low-schooling-ability types earn less in both sectors. More unexpected perhaps, those with a high schooling preference also tend to earn less.

[46]Estimates of the AR(1) process governing the aggregate shocks are reported in Table B8.

Finally, the returns we have described apply when the individual is not in school. Returns to all characteristics are dampened when the individual is also a student. Note that this implies that the information content of wage signals will be lower, all else equal. The magnitude of the estimated coefficients is around 0.67 in both sectors, implying that in-school work is associated with about a substantial 33% reduction in skill returns and informativeness of the signal.

6.3 Learning

Table 8 presents the estimated correlation matrix for the unobserved abilities (initially unknown to the individual) in each sector, along with their variances. The first key takeaway from the correlation matrix is that it clearly supports the idea that skills are multidimensional: all the correlation coefficients are significantly different from one at the 1% level. The data unambiguously rejects a unidimensional model, or even a model with two imperfectly correlated skills (schooling ability and labor market productivity). As such, these results add to a large and growing empirical literature providing evidence that skills are multidimensional in nature (see Heckman and Mosso, 2014, and multiple references therein).

Four-year schooling ability is highly correlated across majors, with a correlation coefficient of 0.83. Positive, though weaker, correlations are also seen with 2-year college ability, ranging from 0.47 with 4-year science ability to 0.67 with 4-year non-science ability. Work abilities are also strongly correlated, with an estimated correlation coefficient of 0.62.

The correlations between schooling abilities and labor market productivity are positive but markedly lower than the correlations across college types and majors. Among the schooling abilities, science ability has the highest correlation with both blue- and white-collar ability, with correlation coefficients between 0.29 and 0.23. The other schooling abilities have correlation coefficients of less than 0.2 with each of the work abilities, with the correlation coefficient between non-science ability and blue collar ability being especially low.

The variances of each of the unobserved ability measures, and of the corresponding outcomes, are given in the bottom two rows of Table 8. Individuals have a significant and substantial amount of uncertainty about their own abilities by the end of high school. Namely, a one-standard-deviation increase in ability in the white-collar (blue-collar) sector translates into a 36% (27%) increase in wages. On the schooling front, a one-standard-deviation increase in ability accounts for around a half of a grade point for each of the schooling options.

While the unknown ability component is large, learning may still take time due to the noise of the signals. Table 9 gives the estimated variances of the idiosyncratic components of wages and grades, respectively. It shows that, even though we account for both types of unobserved ability (known and unknown to the individuals), residual variation in log wages and grades remains sizable. For instance, in the first year of college the signal-to-noise ratios range from 0.33 to 0.37 depending on the schooling option.[47] Turning to the white- and blue-collar abilities, in-school employment yields substantially less informative ability signals than working out of school. This difference is especially salient in the white-collar sector, where the signal-to-noise ratios are 0.20 and 0.49 for in-school and out-of-school employment, respectively. Besides, conditional on being out of college, working in the blue-collar sector is less informative than in the white-collar sector, with a signal-to-noise ratio of 0.34. However, wages earned while in college have similar informational content in both sectors.

6.4 Flow payoffs

Table 10 reports the structural parameter estimates of the flow utility parameters. Recall that we specified the utility of consumption as CRRA and calibrated the risk aversion parameter to 0.4 (see Appendix D.5). The marginal utility of consumption given this risk aversion parameter is positive and significant. Work abilities do not enter the utility directly except through wages, which in turn affect consumption.

The coefficients on prior academic ability—with the variables here referring to two-year, four-year science, and four-year non-science, respectively—indicate that academic ability is particularly important to the utility of the four-year college options. Similarly, the coefficient on high school grades in four-year college options is also large and positive. These positive effects suggest lower costs of effort when prior abilities and, in the case of the four-year options, high school grades are high.

The estimated coefficients on previous activities point to the existence of large switching costs across types of colleges and majors, as well as large costs to changing one's work status. The parameters on working full-time in the college options indicate negative complementarities between school and full-time work. The coefficients on the unobserved types indicate that those that have a high schooling preference value the schooling options more, with those with high schooling ability placing a greater value on the four-year options.

[47] Signal-to-noise ratios are defined and computed as the share of the variance of the signal that is attributable to the latent ability (as opposed to the noise).

6.5 Model fit and ability sorting

We now discuss the fit of the model as well as the (predicted) sorting patterns by forward simulating the model. Model comparisons are computed through forward simulation, using the structural parameter estimates discussed above along with the reduced-form CCPs for the formation of the future value terms. Specifically, we begin by drawing an ability vector for each individual from the population distribution (a multivariate normal with mean zero and covariance $\hat{\Delta}$).[48] We then draw an unobserved type for each individual from a categorical distribution with parameter $\hat{\pi}$ (estimated vector of unobserved type probabilities reported in Table B4). Next, we draw white-collar job offers, preference shocks, and compute choice probabilities using the observed states (i.e., the demographic characteristics and heterogeneity type and ability drawn at the beginning of the simulation), the structural flow utility estimates and the reduced-form CCPs to represent the future value term.[49] We then draw idiosyncratic shocks for the outcome equations (wages and grades) corresponding to the choice that was made. Finally, we compute the implied ability beliefs using the idiosyncratic shock draws and the ability draws, and then update the state space and repeat for $T = 10$ periods.[50] We perform this forward simulation 10 times for each individual in the estimation sample.

Tables B12 and B13 show how the model matches the choice probabilities in the data for all individuals and college graduates, respectively. The model-predicted choice probabilities and the data choice probabilities are pooled across the first ten periods. In each case, the choice probabilities match well. While this is somewhat to be expected given that this is what the estimation procedure is designed to match, the model could still fail to capture dynamic selection.

In Figure 1, we assess how well the model captures dynamic selection by showing the fit on aspects that were not directly targeted by the estimation algorithm. The figure shows how well our model matches educational decisions over time, focusing in particular on college entry rates, college attrition, and graduating in either type of major. Here, too, the model predictions are consistent with the data, except for slightly overestimating the number of students who drop out of college or graduate in later time periods.

[48] The parameter values of the correlation matrix associated with $\hat{\Delta}$ are listed in Table 8.

[49] In the forward simulation after $t = 1$, the choice probabilities are a function of the demographic characteristics, the unobserved type, the beliefs on unobserved ability, and the endogenous state variables such as previous decision and experience.

[50] Updating the state space involves updating the ability beliefs and the choice-dependent state variables. For example, if the person worked full-time in the previous period, then his work experience in the following period is increased by one unit and his previous decision is work full-time.

Turning to the ability sorting patterns, Table 11 shows the posterior mean of each unobserved ability either (i) in the period of last college enrollment (for those who ever enroll) or (ii) in period $T = 10$ (for those who never do). These results are obtained by following the same forward simulation process described in the previous subsection.

Two key patterns stand out from Table 11. First, there is substantial sorting on the basis of college ability. Those who complete college degrees have generally received strongly positive signals regarding their abilities. Further sorting occurs among college graduates, with those who have had high science ability signals choosing science majors. But, reflecting the strong correlation between abilities for science and non-science majors, science graduates have similar non-science posteriors to those of non-science graduates.

The second key pattern from Table 11 is the limited amount of sorting there is across educational paths on the basis of work abilities, especially for those who study in a non-science major.[51] This finding indicates that providing more information about abilities, as we do in the next section, may substantially shift the work ability distribution of college graduates.

7 The importance of information

We now use the structural parameter and learning estimates to investigate the importance of information about one's abilities in two counterfactual scenarios. Because it is computationally infeasible to conduct counterfactuals when individuals are uncertain about their abilities, each of the counterfactuals entails giving individuals full information about their abilities. The first counterfactual does this alone. The second adds to the first by eliminating occupational search frictions, meaning that all individuals have the option of working in the white-collar sector in every period.

To conduct the counterfactuals, we set a retirement date at age 65. We then give all individuals initial draws on the five ability components—two-year, four-year science, four-year non-science, white-collar productivity, and blue-collar productivity—which individuals are now assumed to know when making their educational and labor supply decisions. In the first counterfactual, individuals remain uncertain as to whether they will have the option to work in the white-collar sector in the future. In addition, there are three other main sources of uncertainty that remain in the counterfactuals: the probability of graduating from a four-year

[51]The substantial unresolved uncertainty for college graduates can be seen in Appendix Table B15 which shows the equivalent of Table 11 except the entries are the posterior ability variances.

college conditional on attendance, consumption (in particular through the aggregate labor market shocks and idiosyncratic productivity shocks), and individual preference shocks.[52] We then solve the model backwards to obtain the counterfactual choice probabilities and then forward simulate to obtain the distribution of choices and the average abilities across different choice paths. We include in Appendix J details about additional assumptions we make in our counterfactual exercises.

7.1 Information and educational choices

We first report in Table 12 the four-year college completion status frequencies in the baseline and in the two counterfactual scenarios. Comparing Column 2 (full information) to Column 1 (baseline) shows a 4.4 percentage point increase in the four-year college graduation rate from providing full information about their abilities. These gains in graduation come despite an over 5 percentage point increase in the share of individuals who never go to college. It follows that dropout rates also fall substantially, both for those who drop out and never return (from 32.4% to 26.1%) and for those who stop out and then eventually dropout (from 9% to 6.7%).

These findings are explained by a set of individuals who enroll in college under imperfect information, find out they are not a good match, and then drop out. When ability is known, these individuals do not enroll in the first place. In contrast, there are a set of students who do not enroll under the imperfect information scenario. In the scenario where ability is known, however, these individuals realize they are academically talented and/or have high levels of white-collar productivity and choose to enroll in college continuously. These countervailing forces result in a simultaneous increase in the number of individuals graduating from college and in the number of individuals never attending college.

A second noteworthy difference between the full information counterfactual and the baseline scenario is the shift in majors from non-science to science. This is the result of the wage returns to science degrees being higher than non-science degrees, but especially so in the white-collar sector. As we will show, those who find out that their abilities are high in the white-collar sector are much more likely to pursue a science degree given the higher returns, while those who have especially high blue-collar ability find college less attractive as the returns to education are lower in that sector.

[52] Despite these additional sources of uncertainty, we refer to our counterfactuals as 'full information' where it is implicit that the full information refers to abilities alone.

Comparing the full-information counterfactual to the one where occupational search frictions are also removed (Column 3), we see a further increase in college graduation rates. This increase is, again, the product of two countervailing forces. First, when search frictions are present, college graduates face higher probabilities of having a white-collar option than non-graduates. When search frictions are removed, this channel makes college relatively less attractive. Weighed against this is that the returns to college are higher in white-collar jobs. When search frictions are removed, college graduates are better able to take advantage of these higher returns, making college attendance more attractive. In practice this latter effect dominates, as evidenced by the net increase in college graduation rates.

The patterns in Table 12 mask heterogeneity in the effects of information provision. Table 13 shows counterfactual results broken down by whether the individual's family is above or below the median income in the data.[53] Comparing full information (Column 2) to the baseline (Column 1) shows that information increases the graduation rates of individuals from low-income families by over 6 percentage points. The increase is less than half that for those from high-income families. All told, full information cuts the gap in graduation rates between low and high-income individuals by 3 percentage points, a 17% relative decrease.

The mechanism for this convergence lies in differences in beliefs about the suitability of college between low and high-income individuals. Low-income individuals tend to expect college not to be a good match, as can be seen from the estimated positive marginal utility of family income for four-year colleges (see Table 10). Providing information reveals to some of them that they actually are a good match, increasing their college graduation rates. The reverse holds true for high-income individuals, with full information on ability revealing to some of them that they are not as good of a match for college as what they initially thought. The slow revelation of information in the baseline scenario leads to higher graduation rates for high-income individuals, as a result of switching costs and having already accumulated some years of college experience. Hence, some individuals who discover after a few years that college is not a good match for them opt to finish their degree anyway since they have relatively little of school remaining to do so.

We next investigate how the counterfactual scenarios change the ability compositions across the different choice paths. Table 14 replicates Table 11, but with the abilities now calculated based on the first counterfactual simulation.[54] As in the baseline, Table 14 shows substantial sorting on the basis of college ability. However, there are a couple of significant changes. First, these results point to quantitatively much stronger sorting on college ability.

[53]We refer to those with families above and below median income as high- and low-income, respectively.
[54]Results for counterfactual 2 are in Table B16.

Besides, in contrast to Table 11, there is now also evidence of substantial sorting on the basis of white-collar and blue-collar abilities. Across all the educational paths, those who graduate from college have a comparative advantage in the white-collar sector. Among college graduates, higher abilities in either sector are associated with being a science major. On the other hand, those who never attend college have average blue-collar abilities, but below-average white-collar abilities. Taken together, these results show that imperfect information about one's abilities is an important barrier to the pursuit of comparative advantage.

7.2 Information and labor market outcomes

In order to assess the importance of information about labor market productivity and schooling ability on labor market outcomes, we next study how full information affects wages in each of the labor market sectors. Given that full information results in much more sorting on education based on work abilities, we may expect to see significant changes in the earnings gap between college- and non-college-educated workers, as well as between blue- and white-collar workers.

To conduct this analysis, we focus on wage outcomes at age 28 and examine the sorting patterns among those who are working full-time. Panel (a) of Table 15 shows differences in average earnings for different sectors and education levels for the baseline, the full-information-only counterfactual, and the full-information counterfactual with no search frictions. Average log wages are expressed relative to those working in the blue-collar sector who did not graduate from college.

Comparing the full-information-only counterfactual (Column 2) with the baseline (Column 1) shows that information provision does magnify sorting. Whereas white-collar science graduates in the baseline earned approximately 46 percent more than blue-collar non-graduates, the gap increases to about 90 percent in the counterfactual. The main mechanism underlying this increase is that full information results in workers with high white-collar ability choosing to get degrees in science as the returns to a science degree are especially high in the white-collar sector. The source is not a drop in earnings for blue-collar workers without a college degree as their earnings also increase, though only slightly so. Information provision, therefore, results in significantly stronger matching of abilities and occupations in the labor market. With full information, more individuals also choose a full-time work option, as seen in the second set of columns in Panel (a) of Table 15. Full-time work is chosen by less than 68% individuals in the baseline but rises to over 76% in the counterfactual.[55]

[55]These figures are calculated as 100 minus the entries in the "Remainder" row of each column, since each

Relative to the full information counterfactual, additionally removing occupational search frictions (Column 3) results in declines in the gap between each of the white-collar combinations and their blue-collar counterparts, coupled with shifts into white-collar jobs. This shift into white-collar jobs is driven by those individuals who have high white-collar ability, just not as high as those who would work in the white-collar sector when frictions are present.

The first set of columns of Panel (b) of Table 15 shows the overall premium for college degrees and for working in the white-collar sector. In the full information counterfactual, the premium increases by about 25 percentage points for science graduates but by only 5 percentage points for non-science graduates. These changes are driven by those with high white collar ability now being especially likely to pursue science degrees. This is also reflected in the increased premium for working in the white-collar sector. Those who work in the white collar sector have much higher white collar ability in the counterfactual and are now more likely to have graduated from college with a science degree.

The second set of columns shows how much of these changes in premia are due to changes in ability sorting. Here, we are comparing average abilities in the sector of work for the comparison groups in the counterfactual to those in the baseline. For example, to compute the change in the white-collar premium due to ability, we first consider the counterfactual average ability for white-collar workers and subtract off the average ability for blue-collar workers. Second, we subtract off the corresponding difference in baseline abilities. Importantly, across both counterfactuals, changes in work abilities account for at least 74% of each of the changes in premia; the remainder is due to changes in the composition of who enters each sector-education combination, changes in sector-specific work experience, combined with, for non-graduates, changes in years of college.[56]

To conclude, our results show that informational frictions play a major role in shaping labor market outcomes. Providing full information to students about their own abilities by the end of high school results in substantial changes in the average work abilities of those who graduate from college. Moreover, beyond the college graduation margin, providing more information to students would also result in significant changes in the average productivity levels of workers within each sector, increasing wage gaps between college- and non-college-educated workers, as well as between white- and blue-collar workers.

of the other rows corresponds to full-time work.

[56] Table B14 in Appendix B shows ability sorting for all five dimensions of ability across the baseline and counterfactuals 1 and 2.

8 Conclusion

In this paper, we examine the role played by imperfect information about own schooling ability and labor market productivity in the context of college enrollment decisions, and the transitions between school and work. Using data from the NLSY97, we estimate a dynamic model of college attendance, major choice and work decisions, with the latter over both full- and part-time work as well as whether to work in the blue- or white-collar sector. At the end of each year, individuals update their ability and productivity beliefs using college grades and wages as noisy signals. A central feature of our framework is to allow the different kinds of schooling and workplace abilities to be arbitrarily correlated, implying that signals in one area may be informative about abilities in another area.

Estimation results show that a sizable fraction of dispersion in college grades as well as log wages is attributable to the ability components which are gradually revealed to individuals as they accumulate more signals. These ability components are highly correlated across college types and majors, and across blue- and white-collar sectors. In contrast, grades earned in college, while positively associated with future wages, are less informative about future labor market performance. To the extent that an important part of the mission of higher education is to help prepare students for the labor market, this finding suggests that there may be room for improvement in the screening mechanisms in place in college.

Counterfactual simulations indicate that the four-year college graduation rate would increase by 4.4 percentage points if students perfectly knew ahead of time their own abilities, a share that would go up by 2 more points in the absence of occupational search frictions. Information provision amplifies sorting into college not only based on previously unknown academic abilities, but also sorting on white collar ability. Because the college premium is higher in the white collar sector, and especially so for science majors, those with high white-collar ability now find graduating from college particularly attractive. As a result, providing information significantly increases the wage gap between college and non-college graduates as well as between those working in the white-collar sector and those working in the blue-collar sector. Taken together, our findings indicate that imperfect information about one's own abilities not only plays an important role in accounting for college enrollment and attrition decisions, but also in shaping future labor market outcomes through ability sorting.

References

Adda, Jérôme, Christian Dustmann, Costas Meghir, and Jean-Marc Robin. 2010. "Career Progression and Formal versus On-the-Job Training." Working paper.

Ahn, Thomas, Peter Arcidiacono, Amy Hopson, and James R. Thomas. 2023. "Equilibrium Grade Inflation with Implications for Female Interest in STEM Majors." Working paper, Duke University.

Allman, Elizabeth S., Catherine Matias, and John A. Rhodes. 2009. "Identifiability of Parameters in Latent Structure Models with Many Observed Variables." *Annals of Statistics* 37 (6A):3099–3132.

Altonji, Joseph, Peter Arcidiacono, and Arnaud Maurel. 2016. "The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects." In *Handbook of the Economics of Education*, vol. 5, edited by Eric Hanushek, Stephen Machin, and Ludger Wößmann. Elsevier.

Altonji, Joseph, Erica Blom, and Costas Meghir. 2012. "Heterogeneity in Human Capital Investments: High School Curriculum, College Majors, and Careers." *Annual Review of Economics* 4:185–223.

Altonji, Joseph G. 1993. "The Demand for and Return to Education When Education Outcomes are Uncertain." *Journal of Labor Economics* 11:48–83.

Antonovics, Kate and Limor Golan. 2012. "Experimentation and Job Choice." *Journal of Labor Economics* 30:333–366.

Arcidiacono, Peter. 2004. "Ability sorting and the returns to college major." *Journal of Econometrics* 121:343–375.

Arcidiacono, Peter, Esteban M. Aucejo, and Ken Spenner. 2012. "What Happens after Enrollment? An Analysis of the Time Path of Racial Differences in GPA and Major Choice." *IZA Journal of Labor Economics* 1 (1):1–24.

Arcidiacono, Peter and John B. Jones. 2003. "Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm." *Econometrica* 71:933–946.

Arcidiacono, Peter and Robert Miller. 2011. "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity." *Econometrica* 79:1823–1867.

———. 2015. "Identifying Dynamic Discrete Choice Models off Short Panels." Working paper.

———. 2019. "Nonstationary Dynamic Models with Finite Dependence." *Quantitative Economics* 10 (3):853–890.

———. 2020. "Identifying Dynamic Discrete Choice Models off Short Panels." *Journal of Econometrics* 215 (2):473–85.

Ashworth, Jared, Joseph V. Hotz, Arnaud Maurel, and Tyler Ransom. 2021. "Changes across cohorts in wage returns to schooling and early work experiences." *Journal of Labor Economics* 39 (4):931–64.

Bjerk, David. 2007. "The Differing Nature of Black-White Wage Inequality across Occupational Sectors." *Journal of Human Resources* 42 (2):398–434.

Bound, John and Sarah Turner. 2011. "Dropouts and Diplomas: The Divergence in Collegiate Outcomes." In *Handbook of the Economics of Education*, vol. 4, edited by Eric Hanushek, Stephen Machin, and Ludger Wößmann. Elsevier.

Carneiro, Pedro, Karsten Hansen, and James J. Heckman. 2003. "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice." *International Economic Review* 44 (2):361–422.

Ching, Andrew, Tulin Erdem, and Michael Keane. 2013. "Learning Models: an Assessment of Progress, Challenges, and New Developments." *Marketing Science* 36 (2):913–938.

Chiong, Khai X., Alfred Galichon, and Matt Shum. 2013. "Duality in Dynamic Discrete Choice Models." *Quantitative Economics* 7 (1):83–115.

Clark, Brian, Clement Joubert, and Arnaud Maurel. 2017. "The Career Prospects of Overeducated Americans." *IZA Journal of Labor Economics* 6 (1).

DeGroot, Morris. 1970. *Optimal Statistical Decisions*. New York: McGraw Hill.

Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. 1977. "Maximum Likelihood from Incomplete Data with the EM Algorithm." *Journal of the Royal Statistical Society, Series B* 39:1–38.

Erdem, Tulin and Michael P. Keane. 1996. "Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets." *Marketing Science* 15 (1):1–20.

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, and Michael Westberry. 2022. "Integrated Public Use Microdata Series, Current Population Survey: Version 10.0 [dataset]." https://doi.org/10.18128/D030.V10.0.

Goldin, Claudia and Lawrence F. Katz. 2008. *The Race Between Education and Technology*. Cambridge, Mass.: Belknap/Harvard University Press.

Guner, Nezih, Remzi Kaygusuz, and Gustavo Ventura. 2014. "Income Taxation of U.S. Households: Facts and Parametric Estimates." *Review of Economic Dynamics* 17 (4):559–581.

Hai, Rong and James J. Heckman. 2017. "Inequality in Human Capital and Endogenous Credit Constraints." *Review of Economic Dynamics* 25:4–36.

Hastings, Justine, Christopher A. Neilson, Anely Ramirez, and Seth Zimmerman. 2016. "(Un)informed College and Major Choice: Evidence from Linked Survey and Administrative Data." *Economics of Education Review* 51 (1):136–151.

Heckman, James J., John E. Humphries, and Gregory Veramendi. 2018. "The Nonmarket Benefits of Education and Ability." *Journal of Human Capital* 12 (2):282–304.

Heckman, James J., Anne Layne-Farrar, and Petra Todd. 1996. "Human Capital Pricing Equations with an Application to Estimating the Effect of Schooling Quality on Earnings." *Review of Economics and Statistics* 78:562–610.

Heckman, James J., Lance J. Lochner, and Petra E. Todd. 2006. "Earnings Functions, Rates of Return and Treament Effects: the Mincer Equation and Beyond." In *Handbook of the Economics of Education*, vol. 1, edited by Eric Hanushek and Finis Welch. Elsevier.

Heckman, James J. and Stefano Mosso. 2014. "The Economics of Human Development and Social Mobility." *Annual Review of Economics* 6:689–733.

Heckman, James J. and Salvador Navarro. 2007. "Dynamic Discrete Choice and Dynamic Treatment Effects." *Journal of Econometrics* 136 (2):341–396.

Heckman, James J. and Burt Singer. 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica* 52:271–320.

Heckman, James J. and Sergio Urzúa. 2009. "The Option Value of Educational Choices and the Rate of Return to Educational Choices." Working paper.

Hendricks, Lutz and Oksana Leukhina. 2017. "How Risky is College Investment?" *Review of Economic Dynamics* 26:140–63.

Henry, Marc, Yuichi Kitamura, and Bernard Salanie. 2014. "Partial Identification of Finite Mixtures in Econometric Models." *Quantitative Economics* 5 (1):123–144.

Hotz, V. Joseph and Robert A. Miller. 1993. "Conditional Choice Probabilities and the Estimation of Dynamic Models." *Review of Economic Studies* 60:497–529.

James, Jonathan. 2011. "Ability Matching and Occupational Choice." Working paper.

Joensen, Juanna Schrøter. 2009. "Academic and Labor Market Success: The Impact of Student Employment, Abilities and Preferences." Working paper.

Johnson, Matthew T. 2013. "Borrowing Constraints, College Enrollment, and Delayed Entry." *Journal of Labor Economics* 31 (4):669–725.

Keane, Michael P. and Kenneth I. Wolpin. 1997. "The Career Decisions of Young Men." *The Journal of Political Economy* 105:473–522.

———. 2000. "Eliminating Race Differences in School Attainment and Labor Market Success." *Journal of Labor Economics* 18:614–652.

———. 2001. "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment." *International Economic Review* 42:1051–1103.

Larroucau, Tómas and Ignacio Rios. 2022. "Dynamic College Admissions and the Determinants of Students' College Retention." Working paper.

Manski, Charles F. 1989. "Schooling as Experimentation: A Reappraisal of the Postsecondary Dropout Phenomenon." *Economics of Education Review* 8:305–312.

Manski, Charles F. and David A. Wise. 1983. *College Choice in America*. Harvard University Press.

McLachlan, Geoffrey and David Peel. 2004. *Finite Mixture Models*. John Wiley and Sons.

Miller, Robert. 1984. "Job Matching and Occupational Choice." *The Journal of Political Economy* 92:1086–1120.

Mountjoy, Jack and Brent R. Hickman. 2021. "The Returns to College(s): Relative Value-Added and Match Effects in Higher Education." National Bureau of Economic Research Working Paper No. 29276.

National Center for Education Statistics. 2021a. "Table 326.10. Graduation rate from first institution attended for first-time, full-time bachelor's degree-seeking students at 4-year postsecondary institutions, by race/ethnicity, time to completion, sex, control of institution, and acceptance rate: Selected cohort entry years, 1996 through 2014." https://nces.ed.gov/programs/digest/d21/tables/dt21_326.10.asp. Accessed Apr 6 2023.

———. 2021b. "Table 326.40. Percentage Distribution of First-Time Postsecondary Students Starting at 2- and 4-Year Institutions During the 2011–12 Academic Year, by Highest Degree Attained as of Spring 2017 and Enrollment Status in Spring 2017: Spring 2017." https://nces.ed.gov/programs/digest/d21/tables/dt21_326.40.asp. Accessed Apr 6 2023.

Papageorgiou, Theodore. 2014. "Learning your Comparative Advantages." *Review of Economic Studies* 81 (3):1263–1295.

Proctor, Andrew. 2022. "Did the Apple Fall Far from the Tree? Uncertainty and Learning about Ability with Family-Informed Priors." Working paper.

Pugatch, Todd. 2018. "Bumpy Rides: School to Work Transitions in South Africa." *Labour* 32 (2):205–242.

Robin, Jean-Marc. 2011. "On the Dynamics of Unemployment and Wage Distributions." *Econometrica* 79:1327–1355.

Sanders, Carl. 2014. "Skill Accumulation, Skill Uncertainty, and Occupational Choice." Working paper.

Shephard, Andrew and Modibo Sidibe. 2019. "Schooling investment, mismatch and wage inequality." Working paper.

Stange, Kevin M. 2012. "An Empirical Investigation of the Option Value of College Enrollment." *American Economic Journal: Applied Economics* 4:49–84.

Stinebrickner, Todd and Ralph Stinebrickner. 2012. "Learning about Academic Ability and the College Drop-Out Decision." *Journal of Labor Economics* 30:707–748.

———. 2014. "Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model." *Journal of Labor Economics* 32:601–644.

Tauchen, George. 1986. "Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions." *Economics Letters* 20 (2):177–181.

Trachter, Nicholas. 2015. "Stepping Stone and Option Value in a Model of Postsecondary Education." *Quantitative Economics* 6 (1):223–256.

Table 1: Background characteristics of estimation sample by college enrollment status

	Starting College Type					
	Two-year	Four-year Sci	Four-year Non-Sci	Four-year Missing Major	No college	Total
Black	0.207	0.138	0.208	0.156	0.243	0.201
	(0.406)	(0.345)	(0.406)	(0.363)	(0.429)	(0.401)
Hispanic	0.207	0.127	0.119	0.121	0.189	0.167
	(0.406)	(0.334)	(0.325)	(0.327)	(0.392)	(0.373)
SAT Math	468	572	519	531	439	488
	(94)	(115)	(104)	(111)	(91)	(109)
SAT Verbal	431	525	482	486	408	451
	(104)	(146)	(134)	(147)	(99)	(126)
HS GPA	-0.041	0.694	0.444	0.547	-0.355	0.121
	(0.752)	(0.728)	(0.774)	(0.748)	(0.792)	(0.853)
Parent Graduated College	0.267	0.587	0.531	0.475	0.101	0.327
	(0.443)	(0.494)	(0.500)	(0.500)	(0.302)	(0.469)
Family Income ($1996) (000's)	49.047	69.541	69.542	71.168	39.218	55.379
	(38.979)	(50.279)	(55.817)	(56.207)	(28.094)	(46.146)
Observations	719	189	318	461	613	2,300

Notes: This table reports summary statistics for the subsample of the NLSY97 that is used to estimate our structural model. The sample corresponds to the first observation in which an individual has enrolled in college. Grades are standardized to the NLSY97 male population. Standard deviations are listed directly below the mean (in parentheses) for each entry. See Table A4 for complete details on sample selection.

Table 2: Background characteristics of estimation sample by occupational sector and college completion status

	Non-graduates		Graduates	
	Blue Collar	White Collar	Blue Collar	White Collar
Age (years)	24.813	26.116	27.982	28.399
	(4.434)	(4.319)	(3.281)	(3.184)
Black	0.210	0.169	0.135	0.138
	(0.407)	(0.375)	(0.342)	(0.345)
Hispanic	0.197	0.193	0.112	0.086
	(0.398)	(0.395)	(0.316)	(0.281)
SAT Math	456	488	519	569
	(95)	(96)	(110)	(105)
SAT Verbal	422	465	474	531
	(111)	(114)	(151)	(148)
High School GPA (z-score)	-0.164	0.098	0.545	0.855
	(0.784)	(0.785)	(0.745)	(0.686)
Parent Graduated College	0.188	0.378	0.589	0.619
	(0.391)	(0.485)	(0.492)	(0.486)
Family Income ($1996) (000's)	44.446	54.656	63.502	78.175
	(36.038)	(40.788)	(43.021)	(52.497)
Observations	11,631	1,168	953	1,185
Share Conditional on Graduation Outcome	0.910	0.090	0.446	0.554

Notes: This table reports summary statistics for the subsample of the NLSY97 that is used to estimate our structural model. The sample corresponds to all individual-year observations in a work activity ($N = 14,937$). Grades are standardized to the NLSY97 male population. Standard deviations are listed directly below the mean (in parentheses) for each entry. See Table A4 for complete details on sample selection.

Table 3: Completion outcomes of college enrollees (%)

	Starting College Type				
	Two-Year	Four-Year Sci	Four-year Non-Sci	Four-year Missing Major	Total
Continuous completion (CC), grad. Sci	2.23	44.44	3.46	13.67	10.31
Continuous completion (CC), grad. Non-Sci	8.34	16.40	48.74	36.88	24.66
Stopped out (SO), grad. Sci	2.64	4.23	1.26	0.87	2.07
Stopped out (SO), grad. Non-Sci	7.23	4.76	9.75	8.24	7.71
Stopped out (SO) then dropped out	21.14	6.88	10.38	12.15	15.06
Dropped out (DO)	53.55	20.11	23.27	23.86	35.98
CC right censored	0.28	0.53	0.31	1.30	0.59
SO right censored	4.59	2.65	2.83	3.04	3.62
Total N	719	189	318	461	1,687

Notes: This table reports college completion status statistics for the subsample of the NLSY97 that is used to estimate our structural model, conditional on ever attending college. Completion status is computed using the full available data regardless of missing outcomes. "Right censored" refers to those who are still enrolled in college in the last period of the survey. Students who begin two-year college but never enroll in a four-year college are considered as dropouts. See Table A4 for complete details on sample selection.

Table 4: Period-t GPA outcomes (by $t+1$ period college decision)

(a) 4-year Students, GPA levels

	Mean GPA	Std Dev	N	$\|t$-stat$\|$
Leave 4-year college	2.171	1.003	179	11.561
Stay	2.911	0.793	1720	

(b) 2-year Students, GPA levels

	Mean GPA	Std Dev	N	$\|t$-stat$\|$
Leave 2-year college	2.304	1.098	229	6.650
Stay	2.798	0.866	539	

(c) 4-year Students, GPA *residuals*

	Mean residual	Std Dev	N	$\|t$-stat$\|$
Leave 4-year college	-0.548	0.979	140	8.909
Stay	0.053	0.739	1444	

(d) 2-year Students, GPA *residuals*

	Mean residual	Std Dev	N	$\|t$-stat$\|$
Leave 2-year college	-0.310	1.057	229	6.266
Stay	0.132	0.815	539	

Note: Each t-statistic tests for difference in means between the specified activity and its complement. For residual outcomes, regression covariates include race dummies, SAT scores, parental education, high school GPA, age dummies, birth year, and work intensity dummies.

Table 5: Period t log wage outcomes for stopouts (by $t+1$ decision)

(a) Log wage levels

	Mean log wage	Std Dev	N	$\|t\text{-stat}\|$
Stay in work	2.364	0.501	1,350	3.567
Return to school	2.190	0.432	113	
Total	2.350	0.498	1,463	

(b) Log wage *residuals*

	Mean residual	Std Dev	N	$\|t\text{-stat}\|$
Stay in work	0.068	0.477	1,350	2.001
Return to school	-0.025	0.413	113	
Total	0.060	0.473	1,463	

Note: Results are conditional on having attended at least one year of college, currently working, and not yet having graduated from college. As a result, the residuals do not average to zero here because the relevant population is all wage observations in the estimation subsample of the data. Regression covariates include levels and interactions of the following variables: race and year dummies; SAT scores; graduation outcomes, experience in different sectors; field of study; birth year; age; in-school work dummies; and work intensity dummies.

Table 6: Estimates of 2- and 4-year GPA Parameters

	4 year Science		4 year Non-Science		2 year	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Black	-0.352	(0.074)	-0.469	(0.051)	-0.314	(0.065)
Hispanic	-0.192	(0.086)	-0.151	(0.060)	-0.038	(0.065)
Parent graduated college	0.087	(0.052)	0.163	(0.044)	0.032	(0.064)
HS Grades (z-score)	0.177	(0.032)	0.211	(0.025)	0.234	(0.026)
Work full-time	-0.173	(0.061)	-0.156	(0.041)	-0.103	(0.048)
Work part-time	-0.029	(0.051)	0.036	(0.029)	-0.002	(0.054)
Year 2 or higher in college					-0.008	(0.039)
Unobserved type						
Schooling ability type H	0.151	(0.075)	0.078	(0.050)	0.021	(0.064)
Schooling preference type H	0.649	(0.122)	0.467	(0.085)	-0.039	(0.111)
Work motivation type H	-0.003	(0.080)	0.010	(0.059)	-0.007	(0.079)
λ_0 (ability index intercept)	-0.261	(0.227)	0.302	(0.108)	0.000	(—)
λ_1 (ability index loading)	1.044	(0.082)	0.907	(0.042)	1.000	(—)
Mean of dependent variable	2.778		2.683		2.468	
Person-year obs.	867		1,653		1,272	

Notes: Bootstrap standard errors in parentheses. Estimates for the constant, birth cohort dummies, and age group dummies (18 and under, 19, 20) are suppressed. Reference categories for multinomial variables are as follows: "White" for race/ethnicity, "Not working while in school" for work intensity, and "L" for each unobserved type.

Table 7: Estimates of White- and Blue-collar Wage Parameters

	White Collar		Blue Collar	
	Coeff.	Std. Error	Coeff.	Std. Error
Black	-0.059	(0.024)	-0.098	(0.020)
Hispanic	0.049	(0.028)	0.018	(0.021)
Parent graduated college	0.003	(0.021)	0.004	(0.019)
HS Grades (z-score)	-0.019	(0.015)	-0.017	(0.009)
Work experience (any sector)	0.035	(0.003)	0.039	(0.001)
Work experience (white collar sector)	0.020	(0.003)	0.010	(0.003)
Years of college completed	0.063	(0.010)	0.038	(0.005)
College graduate	0.113	(0.024)	0.061	(0.018)
College graduate × science major	0.132	(0.035)	0.091	(0.026)
Work part-time	0.011	(0.020)	-0.044	(0.009)
Unobserved type				
Schooling ability type H	0.087	(0.026)	0.086	(0.024)
Schooling preference type H	-0.411	(0.038)	-0.255	(0.024)
Work motivation type H	0.248	(0.032)	0.189	(0.025)
λ_0 (in-school work index intercept)	0.718	(0.007)	0.659	(0.009)
λ_1 (in-school work index loading)	0.633	(0.009)	0.667	(0.005)
Mean of dependent variable	2.669		2.332	
Person-year obs.	2,373		12,755	

Notes: Bootstrap standard errors in parentheses. Estimates for the constant, birth cohort dummies, age group dummies (18 and under, 19, 20), and calendar year dummies are suppressed. Reference categories for multinomial variables are as follows: "White" for race/ethnicity, "Work full-time" for work intensity, and "L" for each unobserved type.

The return to a college degree is equal to four times the coefficient on years of college completed, plus the relevant coefficients on the college graduate dummies. For example, the return to a science major in the white-collar sector is equal to $4 \times 0.063 + 0.113 + 0.132 = 0.497$.

Table 8: Ability Correlation Matrix and Variances of Unobserved Abilities and Raw Outcomes

	White Collar	Blue Collar	Science	Non-Science	2-year
White Collar	1.000				
	(—)				
Blue Collar	0.615	1.000			
	(0.037)	(—)			
Science	0.283	0.239	1.000		
	(0.085)	(0.096)	(—)		
Non-Science	0.194	0.045	0.831	1.000	
	(0.062)	(0.062)	(0.052)	(—)	
2-year	0.153	0.162	0.470	0.669	1.000
	(0.082)	(0.079)	(0.106)	(0.076)	(—)
Variance of Unobserved Abilities	0.130	0.072	0.207	0.274	0.353
	(0.010)	(0.005)	(0.039)	(0.027)	(0.034)
Variance of Raw Outcomes	0.360	0.272	0.842	0.828	1.029

Notes: Bootstrap standard errors in parentheses. "Variance of Unobserved Abilities" refers to the diagonal elements of the covariance matrix corresponding to the correlation matrix presented here. "Variance of Raw Outcomes" refers to the variance of the corresponding outcome variables (log wages, college GPA). Each cell of the correlation matrix contains at least 127 unique individuals and at most 1,810 unique individuals. Our estimation sample contains 2,300 unique individuals.

Table 9: Idiosyncratic Variances

Employment			Schooling			
Work Type	White Collar	Blue Collar	Schooling Period	Science	Non-Science	2-year
In-school	0.212	0.120	1	0.431	0.519	0.609
	(0.013)	(0.003)		(0.041)	(0.033)	(0.047)
Out-of-school	0.136	0.143	2	0.364	0.322	0.495
	(0.003)	(0.001)		(0.036)	(0.021)	(0.037)
			3	0.414	0.385	0.624
				(0.045)	(0.024)	(0.037)
			4	0.607	0.317	
				(0.066)	(0.023)	
			5+	0.403	0.472	
				(0.036)	(0.026)	

Notes: Bootstrap standard errors in parentheses. The period-3 variance in 2-year college is the same for all periods after period 3.

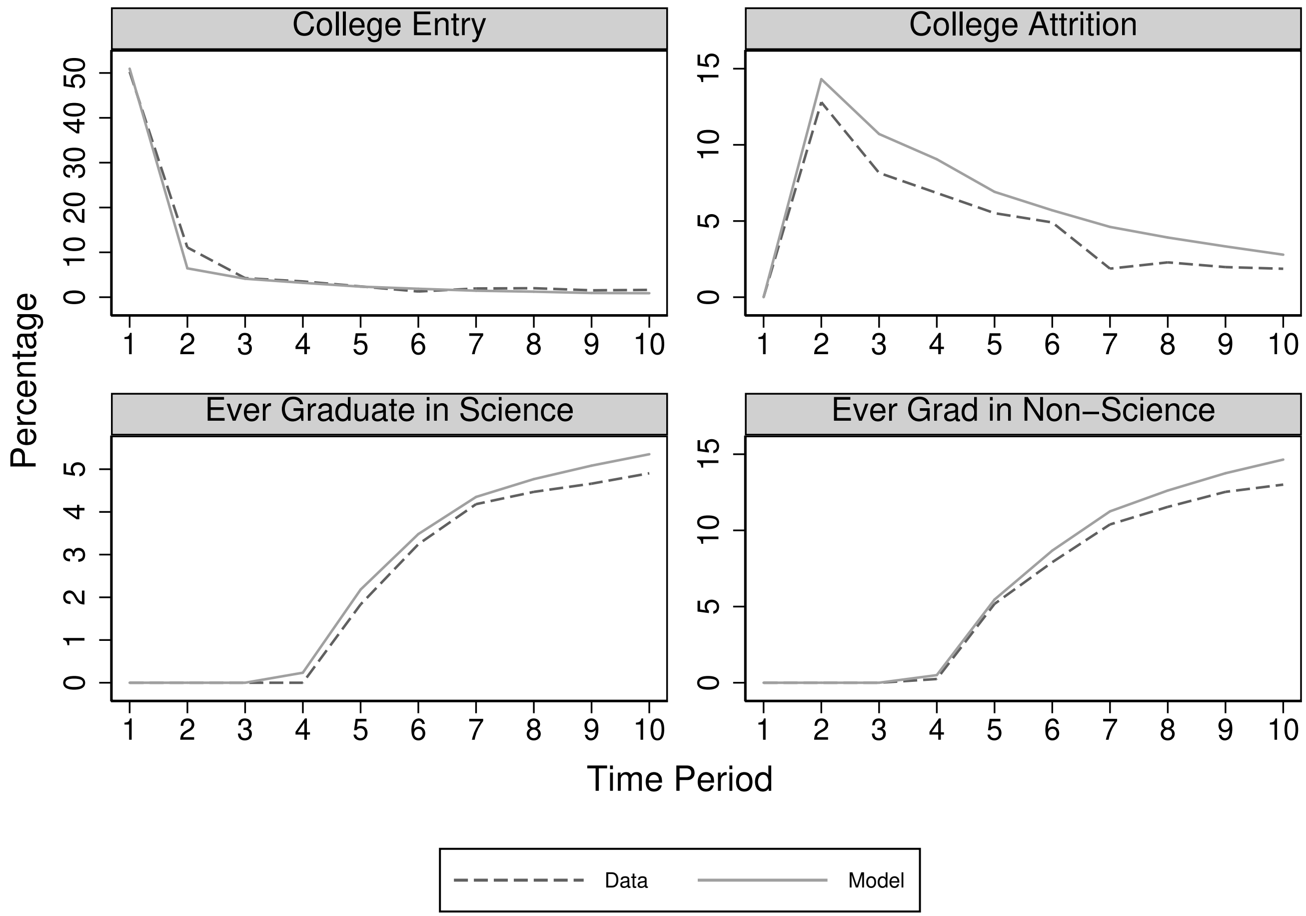
Table 10: Flow Utility Parameter Estimates

	College			Work		
Variable	2-year	Science	Non-Science	Part-time	Full-time	White Collar
Black	-0.036	-0.034	-0.013	0.025	0.014	-0.001
	(0.055)	(0.097)	(0.109)	(0.046)	(0.033)	(0.051)
Hispanic	0.053	-0.013	0.056	-0.070	-0.030	0.027
	(0.051)	(0.109)	(0.116)	(0.052)	(0.040)	(0.053)
HS Grades (z-score)	0.032	0.373	0.255	0.007	0.020	0.055
	(0.028)	(0.056)	(0.057)	(0.021)	(0.015)	(0.020)
Parent graduated college	0.168	0.290	0.220	0.063	-0.044	0.206
	(0.057)	(0.088)	(0.098)	(0.040)	(0.031)	(0.039)
Family Income ($10,000)	0.011	0.051	0.058	-0.014	-0.004	-0.002
	(0.007)	(0.009)	(0.010)	(0.004)	(0.003)	(0.004)
Prior academic ability	0.148	1.357	1.109			
	(0.074)	(0.280)	(0.214)			
Previous high school	1.033	2.746	1.813	1.169	0.855	-0.578
	(0.104)	(0.174)	(0.117)	(0.089)	(0.082)	(0.194)
Previous 2-year college	2.455	1.274	0.711	0.049	0.281	0.118
	(0.082)	(0.190)	(0.094)	(0.085)	(0.079)	(0.140)
Previous 4-year science	1.254	4.796	2.189	0.718	0.378	-0.157
	(0.172)	(0.165)	(0.131)	(0.116)	(0.094)	(0.153)
Previous 4-year non-science	-0.068	2.035	3.409	0.598	0.559	0.046
	(0.164)	(0.187)	(0.091)	(0.088)	(0.077)	(0.106)
Previous work part-time	0.030	0.577	0.376	2.186	1.357	-0.966
	(0.092)	(0.127)	(0.101)	(0.058)	(0.054)	(0.093)
Previous work full-time	0.105	0.261	0.557	0.963	2.270	-0.923
	(0.097)	(0.135)	(0.098)	(0.062)	(0.042)	(0.095)
Previous work white-collar	-0.032	-0.513	-0.115	-1.284	-1.464	2.692
	(0.161)	(0.235)	(0.159)	(0.088)	(0.079)	(0.184)
College graduate				-0.215	-0.023	0.483
				(0.092)	(0.103)	(0.052)
Currently work white-collar				-0.144		
				(0.115)		
Currently work part-time	0.647	-0.698	-0.366			
	(0.099)	(0.142)	(0.122)			
Currently work full-time	-0.715	-1.434	-1.851			
	(0.111)	(0.147)	(0.123)			
Unobserved type						
Schooling ability type H	-0.066	0.336	0.088	-0.024	-0.036	0.064
	(0.048)	(0.076)	(0.065)	(0.046)	(0.039)	(0.045)
Schooling preference type H	0.438	0.277	0.411	0.033	0.013	0.060
	(0.103)	(0.142)	(0.143)	(0.068)	(0.091)	(0.064)
Work motivation type H	-0.012	-0.026	0.064	-0.273	-0.073	0.030
	(0.057)	(0.077)	(0.071)	(0.045)	(0.056)	(0.046)
$\mathbb{E}[u(\text{consumption})] \div 1,000$	2.510	(0.467)				
Pr(graduate in $t+1$)	2.116	(1.233)				
Constant Relative Risk Aversion parameter (θ)	0.4					
Log likelihood	-26,105					
Person-year obs.	22,398					

Notes: Home production is the reference alternative. Bootstrap standard errors are listed below or to the side of each coefficient in parentheses. Beliefs on labor market productivity are included in the expected utility of consumption term. Consumption is evaluated in terms of yearly consumption flow in 1996 dollars. Missing majors are estimated to be science with probability 0.35. Missing GPAs are estimated to be ≤ 2.5 w.p. 0.48, 2.5–3.0 w.p. 0.21, 3.0–3.6 w.p. 0.19, and 3.6–4.0 w.p. 0.13.

Reference categories for multinomial variables are as follows: “White” for race/ethnicity, “Previous home production” for previous decision, “Not working” for in-college work intensity, and “L” for each unobserved type. We omit the following coefficients: the constant (for each choice); birth year dummies (for each choice); and the interactions between currently working full- or part-time and currently working in the white-collar sector (for each of the schooling choices).

Figure 1: Model fit of untargeted moments



Notes: This figure plots rates of college entry, attrition, and graduation by time period separately for the data and model. Model frequencies are constructed using 10 simulations of the structural model for each individual included in the estimation.

Table 11: Average posterior abilities after last period of college for different choice paths in baseline model

Choice Path	White Collar	Blue Collar	Science	Non-Science	2-year	Share(%)
Continuous enrollment, graduate in science with x years of in-school work experience						
$x = 0$	0.10	0.09	0.33	0.30	0.19	1.39
$x > 0$, white collar only	0.02	0.04	0.29	0.24	0.13	0.52
$x > 0$, blue collar only	0.12	0.12	0.34	0.30	0.20	3.63
$x > 0$, mixture	0.13	0.10	0.24	0.20	0.13	0.85
Continuous enrollment, graduate in non-science with x years of in-school work experience						
$x = 0$	0.05	0.01	0.22	0.28	0.20	3.03
$x > 0$, white collar only	0.08	0.03	0.19	0.22	0.15	1.04
$x > 0$, blue collar only	0.04	-0.01	0.26	0.33	0.24	8.88
$x > 0$, mixture	0.03	-0.03	0.26	0.32	0.24	1.90
Stop out (SO)						
SO, graduate in science	0.16	0.21	0.18	0.14	0.14	0.97
SO, graduate in non-science	-0.01	-0.06	0.17	0.25	0.21	3.45
SO then DO, start in 2yr	-0.02	-0.01	-0.09	-0.09	-0.04	4.98
SO then DO, start in science	-0.01	0.05	-0.36	-0.38	-0.26	1.37
SO then DO, start in non-science	-0.07	-0.03	-0.27	-0.32	-0.25	2.67
Truncated	-0.03	-0.04	0.02	0.05	0.09	5.76
Drop out (DO) after x years of school						
$x = 1$	-0.03	-0.02	-0.11	-0.13	-0.13	16.59
$x = 2$	-0.05	-0.03	-0.18	-0.21	-0.19	7.85
$x = 3$	-0.05	-0.03	-0.19	-0.21	-0.15	4.22
$x = 4$	-0.04	-0.00	-0.24	-0.27	-0.19	1.93
$x \geq 5$	-0.06	-0.01	-0.29	-0.29	-0.11	1.85
Never attended college						
Never attend college	-0.00	-0.00	-0.00	-0.00	-0.00	27.11

Notes: Abilities are reported in standard deviation units. This table is constructed using 10 simulations of the baseline model for each individual included in the estimation.

"Truncated" refers to those who were enrolled in period 10.

Table 12: College completion status frequencies: baseline and counterfactual

Status	Baseline model	Counterfactuals: Full info. alone	Counterfactuals: Full info. & no search frictions
Continuous completion (CC), Science	6.39	10.30	11.09
Continuous completion (CC), Non-Science	14.86	13.74	14.31
Stop out (SO) but graduated Science	0.97	1.74	1.94
Stop out (SO) but graduated Non-Science	3.45	4.33	4.72
Stop out (SO) then drop out	9.02	6.71	6.50
Truncated	5.76	4.73	5.07
Drop out (DO)	32.44	26.10	25.17
Never went to college	27.11	32.35	31.20
Graduate from 4-year college	25.67	30.11	32.07
Ever Switch Major	25.37	25.13	26.31
Time to degree	5.16	4.99	5.02

Notes: Model frequencies are constructed using 10 simulations of the structural model for each individual included in the estimation. Counterfactual frequencies use 10 simulations of each counterfactual model. "Full info. alone" refers to our counterfactual where individuals have complete information about their abilities. "Full info. & no search frictions" maintains full information and sets to 1 the white collar job offer arrival rate for everyone in every period.

We set the panel length in all columns to be 10 periods. Completion status is computed on the first 10 periods of data (i.e. assuming that college is not an option after period 10).

"Truncated" refers to those who were enrolled in period 10.

Table 13: College completion status in model and counterfactuals: heterogeneity by level of family income

Status	Baseline model	Counterfactuals: Full info. alone	Counterfactuals: Full info. & no search frictions
Panel A: Above-median family income in high school			
Dropout	28.83	23.73	22.34
Never went to college	20.85	27.01	25.60
Graduate from 4-year college	35.40	38.22	40.43
Ever Switch Major	27.53	26.87	28.09
Panel B: Below-median family income in high school			
Dropout	36.04	28.46	27.99
Never went to college	33.37	37.69	36.81
Graduate from 4-year college	15.93	22.01	23.70
Ever Switch Major	22.19	22.75	23.87

Notes: Model frequencies are constructed using 10 simulations of the structural model for each individual included in the estimation. Counterfactual frequencies use 10 simulations of each counterfactual model. "Full info. alone" refers to our counterfactual where individuals have complete information about their abilities. "Full info. & no search frictions" maintains full information and sets to 1 the white collar job offer arrival rate for everyone in every period.

We set the panel length in all columns to be 10 periods. Completion status is computed on the first 10 periods of data (i.e. assuming that college is not an option after period 10).

Table 14: Average abilities for different choice paths in full-information counterfactual scenario

Choice Path	White Collar	Blue Collar	Science	Non-Science	2-year	Share(%)
Continuous enrollment, graduate in science with x years of in-school work experience						
$x = 0$	0.36	0.06	1.04	0.69	0.28	0.98
$x > 0$, white collar only	1.41	0.69	1.09	0.83	0.49	0.48
$x > 0$, blue collar only	0.57	0.42	1.02	0.70	0.31	7.59
$x > 0$, mixture	1.16	0.59	0.93	0.65	0.33	1.25
Continuous enrollment, graduate in non-science with x years of in-school work experience						
$x = 0$	0.02	-0.38	0.59	0.89	0.55	1.86
$x > 0$, white collar only	0.80	-0.03	0.62	0.91	0.49	0.50
$x > 0$, blue collar only	0.19	-0.07	0.59	0.86	0.58	10.27
$x > 0$, mixture	0.65	0.12	0.53	0.77	0.48	1.10
Stop out (SO)						
SO, graduate in science	0.36	0.25	0.87	0.60	0.30	1.74
SO, graduate in non-science	0.06	-0.12	0.51	0.74	0.54	4.33
SO then DO, start in 2yr	-0.20	-0.13	-0.33	-0.32	-0.13	3.17
SO then DO, start in science	-0.19	-0.05	0.06	-0.05	-0.05	0.98
SO then DO, start in non-science	-0.24	-0.18	-0.10	0.02	0.06	2.57
Truncated	-0.09	-0.12	0.04	0.14	0.15	4.73
Drop out (DO) after x years of school						
$x = 1$	-0.18	-0.03	-0.33	-0.33	-0.17	15.73
$x = 2$	-0.21	-0.09	-0.25	-0.24	-0.13	5.77
$x = 3$	-0.09	-0.03	-0.10	-0.08	-0.03	2.58
$x = 4$	-0.20	-0.01	-0.06	-0.05	0.06	1.26
$x \geq 5$	-0.06	-0.02	0.04	-0.06	-0.11	0.76
Never attended college						
Never attend college	-0.13	-0.01	-0.43	-0.48	-0.30	32.35

Notes: Abilities are reported in standard deviation units. This table is constructed using 10 simulations of the counterfactual model described in the title for each individual included in the estimation.

"Truncated" refers to those who were enrolled in period 10.

Table 15: Wage Decompositions

(a) Average full-time log wage and choice share by employment sector and education level at age 28 in baseline and counterfactual models

	Average full-time log wage, relative to blue-collar non-graduates in baseline			Choice shares (%)		
Sector and Education Level	Baseline	Counterfactual	No Frictions Cfl	Baseline	Counterfactual	No Frictions Cfl
White collar, Science graduate	0.458	0.904	0.868	3.44	5.54	6.87
White collar, Non-Science graduate	0.317	0.632	0.557	8.15	5.28	7.25
White collar, Non-graduate	0.110	0.345	0.146	5.46	0.96	4.72
Blue collar, Science graduate	0.197	0.298	0.301	2.45	5.38	4.87
Blue collar, Non-Science graduate	0.111	0.082	0.078	6.62	9.62	8.46
Blue collar, Non-graduate	0.000	0.009	0.007	41.4	49.6	43.8
Remainder	—	—	—	32.5	23.7	24.0

Notes: "No Frictions Cfl" refers to the counterfactual where white-collar work is always an option. Columns in the "choice shares" panel sum to 100.

(b) Full-time log wage premia at age 28 in baseline and counterfactual models

	Full-time log wage premium			Change in premium (relative to baseline) due to better sorting on abilities	
Sector	Baseline	Counterfactual	No Frictions Cfl	Counterfactual	No Frictions Cfl
College wage premium	0.247	0.400	0.421	0.137	0.140
Science college premium	0.336	0.590	0.612	0.213	0.217
Non-science college premium	0.212	0.262	0.279	0.075	0.077
White-collar wage premium	0.255	0.692	0.525	0.321	0.229

Notes: "College wage premium" is the difference in average log wages between college graduates (regardless of major) and non-graduates. "Science college premium" is the difference in average log wages between science graduates and non-graduates. "Non-science college premium" is the difference in average log wages between non-science graduates and non-graduates. "White collar premium" is the difference in average log wages between white-collar and blue-collar workers.

For the panel on changes in premia, numbers represent differences in differences in average abilities (in log dollar units). The first difference is between sector groups (e.g. college graduates vs. non-graduates) and the second difference is between counterfactual and baseline. We compress the bivariate work ability distribution into a single ability index based on which sector each full-time worker is working in.

College Attrition and the Dynamics of Information Revelation

Online Appendix

Contents

A Data details

This Appendix section details our construction of three sets of key variables used in our analysis: choices, college grades, and wages. We also detail how we select the subsample used in estimation.

A.1 Choices

From the NLSY97 data we classify individuals based on their labor force participation, occupational sector, and educational choices. Specifically, we classify individuals in each period using the following rules:

1. Any individual attending a college in the month of October is classified as being in college for this year (either in a two- or a four-year college). For four-year colleges, our definition of "Science" majors includes majors in Science, Technology, Engineering, and Mathematics (STEM). See Table A1 for details on the exact majors in each category. We use "Science" and "STEM" interchangeably.

2. Any individual reporting college attendance who also reports working at least four weeks in October and at least 10 hours per week is classified as working part-time while in school, with full-time work requiring at least 35 hours per week and four weeks worked in October.

3. Any individual not in college (according to the criterion above) is classified as working part-time or full-time according to the criteria above.[A1]

4. Individuals who report part- or full-time employment are also classified as working in the blue- or white-collar sector. Using data from the March Current Population Survey (CPS) accessed from Flood et al. (2022) for years 2007–2014, we classify each three-digit 2010 Census occupation code as being white collar if 50% or more of male workers aged 18–64 in that occupation hold at least a bachelor's degree. Individuals in the NLSY97 are then classified as working in a blue- or white-collar sector if the three-digit 2010 Census occupation code associated with their October job has a majority of college graduates as reported in the CPS.[A2] See Table A2 for the most common occupations by sector and education level.

[A1] These criteria for labor force participation resemble those of Keane and Wolpin (1997).

[A2] For a similar approach, see Clark, Joubert, and Maurel (2017).

5. Finally, all other cases are classified as home production.[A3]

A.2 Grades and Majors

We use a four-point scale to measure college GPA (grades), which is the average GPA over all enrollment periods in a calendar year. We use self-reported survey responses to construct college major (see Table A1) and transcript data from the NLSY97 to construct grades.[A4] Out of 8,984 individuals in the NLSY97 (men and women), 2,830 never enrolled in post-secondary education and 1,445 did not permit collection of their transcripts. The NLSY received transcripts for 3,818 youths (men and women).

In our male-only subsample, grades are missing for 54% and majors are missing or unknown for 26% of college students. We deal with these missing grades and majors by integrating over the first missing period (see Appendix Section C.4).

A.3 Wages

Wages are calculated as follows:

1. We compute the hourly compensation (i.e. wage plus tips and bonuses) for the self-reported main job, converted to 1996 dollars.

2. Of the 22,631 person-year observations that report full- or part-time employment, about one-fourth have missing wage observations. Many of these are coincident with the NLSY97's shift to a biennial frequency beginning in 2012. Since our data go until 2015, the years 2012 and 2014 tend to have larger numbers of missing wage observations than the other years in our sample.

3. We take the following steps to reduce the number of missing wages and thus increase our sample size. Table A3 contains complete details.

    - If a person reports working for the same employer across a 3-year span, but earnings are only observed in the first and third years, we linearly interpolate their second-year earnings as the average between the first and third years. This

[A3] Following this criterion, any individual who is unemployed (and not enrolled in college) in October is classified in the home production sector.

[A4] See `https://www.nlsinfo.org/content/cohorts/nlsy97/other-documentation/codebook-supplement/appendix-12-post-secondary-transcript` for a complete discussion of the transcript data, which we summarize in the following sentence.

is a common situation for calendar years 2012 and 2014. It is also a common situation if someone skipped a round of the survey. This method allows us to reduce the rate of missingness by 12 percentage points.

- If the missing wage observation comes at the end of a job spell or comprises the entirety of a single job spell, we use the reported wage for the following year. This is about 3% of all employment observations.
- If the missing wage observation comes at the beginning of a job spell, we use a flexible model with individual fixed effects to fill in the missing wage observation. This is about 6% of all employment observations.
- Finally, we impute any remaining missing earnings as annual income divided by annual hours worked. This is about 2% of all employment observations.

4. As a final step, we top- and bottom-code the resulting earnings distribution at the 99.5 percentile and 2.5 percentile, respectively.

After following these steps, we are left with 1.6% of employment observations that are missing. These missing observations then affect our sample in the same way as second or later instances of missing grades or majors. In the end, our sample contains 14,937 work observations with no missing wages.

A.4 Sample selection

To conclude, Table A4 shows all the sample restrictions that we imposed for the analysis. In particular, we limit our analysis to men who have completed high school. We also drop all current and future observations for any respondents missing a wage while choosing a work activity. We integrate out over the first missing grade and/or college major (see Section C.4), but drop current and future observations when a second missing grade or college major is observed. Our final estimation sample includes 22,398 person-year observations for 2,300 men.

Table A1: Major Definitions

Science (STEM) Majors	Non-Science Majors
Agriculture and natural resource sciences	All other majors
Biological sciences	
Computer/Information science	
Engineering	
Mathematics	
Physical sciences	
Nutrition/Dietetics/Food Science	

Table A2: Most common occupations by sector in the NLSY97

Sector	Education level	Occupation title	Frequency (%)
Blue Collar	Non-College Graduate	Laborers and Freight, Stock, and Material Movers	4.55
		Food Preparation Workers	4.13
		Driver/Sales Workers and Truck Drivers	3.87
		Retail Salespersons	3.76
	College Graduate	First-Line Supervisors of Sales Workers	6.51
		Sheriffs, Bailiffs, Correctional Officers	5.14
		Customer Service Representatives	3.46
		Laborers and Freight, Stock, and Material Movers	3.36
White Collar	Non-College Graduate	Managers, nec (including Postmasters)	10.69
		Network and Computer Systems Administrators	6.20
		Sales Representatives and Services	5.57
		Human Resources Managers	4.49
	College Graduate	Managers, nec (including Postmasters)	6.26
		Software Developers	6.18
		Postsecondary Teachers	4.48
		Secondary School Teachers	3.98

Table A3: Steps taken to mitigate number of missing wage observations

Description	Person-years	Percentage missing
Employed part- or full-time in preliminary sample[a]	22,631	—
Initial number with missing wages	5,384	23.79
Interpolation and imputation:		
Remainder missing after interpolating missing wages within the same job spell[b]	2,631	11.63
Remainder missing after using next-period reported wage for some of the missing wages[c]	2,107	9.31
Remainder missing after imputing (via FE regression) prior-period wage for missing current-period wage[d]	836	3.69
Remainder missing after imputing wages as annual income / annual hours worked	372	1.64
Employed part- or full-time in final sample	14,937	0.00

Notes: Each row of the table lists the remaining number and percentage of employment observations that have missing wages after cumulatively taking the corresponding action described in the row and all rows above it.

[a] Preliminary sample refers to our estimation subsample prior to dropping missing wages, college grades, or college majors.

[b] We linearly interpolate missing wages within the same job spell. This occurs most frequently in waves after the survey switched to a biennial frequency (i.e. years after 2011).

[c] We replace missing current-period wages with the next-period wage in years 2012 and 2014 when the job spell ended in 2012 and 2014.

[d] We use a regression model with individual fixed effects to fill in missing wage observations within the same employment spell that cannot be interpolated due to not having two endpoints. This occurs most frequently in years 2012 and 2014 that are not directly covered by the survey due to being in the biennially administered phase.

Table A4: Sample Selection

Selection criterion	Resultant persons	Resultant person-years
Full NLSY97 sample	8,984	170,696
Drop women	4,599	87,381
Drop other race	4,559	86,621
Drop missing AFQT and SAT test scores	3,789	71,991
Drop missing HS grades, Parental education, or Parental income	3,059	58,121
Drop HS Dropouts and GED recipients	2,411	45,809
Drop observations before HS graduation	2,331	36,003
Drop right-censored missing interview spells	2,331	35,548
Drop any who attend college at a young age or graduate college in 2 or fewer years	2,331	34,278
Drop any who are not in HS at age 15 or under or have other outlying data	2,331	34,186
Drop any who graduate HS after age 20	2,301	33,825
Drop observations after and including the first instance of missing a wage while working, or after the first instance of a missing college major or GPA	2,300	22,398
Final structural estimation subsample[a]	2,300	22,398

[a] Our structural estimation procedure incorporates integration of missing GPA and major observations, as discussed in Section C.4.

B Supporting Tables

Table B1: Summary of estimation steps

Estimation Stage	Description	Inputs	Outputs	Notes
–	Loans, grants, and transfers	- NLSY97 data on demographics and parental transfers - NPSAS data on grants and loans - SIPP data on family assets	- Function mappings for predicting loans, grants, and transfers	- Use NLSY97, NPSAS, and SIPP data - See Appendix D for complete details
1	Measurement system and missing data parameters	- measurements - outcomes (wages, GPA) - choices - demographics - other state variables (experiences, educational degrees)	- Unobserved type probabilities - Missing outcome (major or GPA) type probabilities	- Flexible model to approximate nonlinear relationships - Find global optimum - See Appendix C for complete details
2	Learning parameters	- outcomes (wages, GPA), - unobserved type probabilities, - missing outcome type probabilities, - demographics, - other state variables (experiences, educational degrees)	- learning parameter estimates	
3	Parameters for offer arrival logit, CCP logit, Graduation logit, Wage AR(1) process	- choices, - wages & consumption, - ability priors, - unobserved type probabilities, - missing outcome type probabilities, - demographics - college graduation outcome - other state variables	- job offer arrival parameter estimates (including individual offer probabilities) - CCP logit coefficients - Graduation logit coefficients - Wage AR(1) coefficients	- Choice model in this stage includes CRRA consumption - Integrate over future consumption in CCPs - See Appendices D and H for complete details
4	Compute future value (FV) terms	All data, all parameter estimates	Future value terms	See Appendix F for complete details
5	Structural flow utility parameters	All data, all parameter estimates, FV terms	Structural flow utility parameter estimates	Estimation is a McFadden logit model with alternative-specific offset terms from stage 4

Table B2: Variables Included in Each Component of the Model, by Estimation Stage

	Stage 2		Stage 3						Stage 5
Variable	Log Wages	College GPAs	Parental Transfers	Expected Grants	Expected Loans	Static Choice	White Collar Offer	College Graduation	Dynamic Choice
Individual background									
Race dummies	✓	✓	✓			✓		✓	✓
Birth cohort dummies	✓	✓				✓		✓	✓
High school GPA	✓	✓				✓		✓	✓
SAT Math (dummies)				✓	✓				
SAT Verbal (dummies)				✓	✓				
Family background									
Parent completed college	✓	✓				✓		✓	✓
Family income ($10,000)						✓		✓	✓
Family income (log)			✓						
Family income (dummies)				✓	✓				
EFC (dummies)				✓	✓				
Aggregate labor market									
Year dummies	✓								
Individual characteristics									
Age dummies ($\leq$18, 19, 20, $\geq$21)	✓	✓							
Age (linear)			✓				✓		
Expected prior work ability	✓								
Expected prior acad. ability						✓			✓
Education									
Bachelor's degree	✓					✓	✓		✓
Bachelor's × Science major	✓								
Expected prior acad. ability × Science major						✓		✓	✓
Experiences (in years)									
College (either type, linear)	✓		✓						
College (dummies)	✓								
Graduate school (dummies)	✓								
4-year college (dummies)								✓	
2-year college (dummies)								✓	
4-year dummies × 2-year dummies								✓	
Overall work experience	✓								
White-collar work experience	✓								
Work/Study characteristics									
Work white-collar dummy						✓			✓
Work full-time dummy		✓				✓		✓	✓
Work part-time dummy	✓	✓				✓		✓	✓
Work in-school signal adjustment	✓								
Upperclassman signal adjustment		✓							
Expected utility of consumption						✓			✓
Accumulated debt (quadratic)						✓			
Previous decision (dummies)						✓			✓
CCP adjustment terms									✓
Unobserved Types									
Type dummies	✓	✓				✓	✓	✓	✓

Notes: There are no columns in this table for Stage 1 or Stage 4. This is because Stage 1 is a flexible model that includes many more covariates and interactions than we can display here, while Stage 4 consists of calculating the future value terms and does not involve any estimation. See Appendices C (flexible Stage 1) and F (Stage 4 formulas) for complete details.

Table B3: Unobserved Type Coefficient Signs and Significance Across All Equations of the Model

	Unobserved Type Identity		
Model Equation	Sch. Abil. H	Sch. Pref. H	Work Motivation H
Panel A: Measurement System			
ASVAB Arithmetic Reasoning	+**		
ASVAB Coding Speed	+**		
ASVAB Mathematical Knowledge	+**		
ASVAB Numerical Operations	+**		
ASVAB Paragraph Comprehension	+**		
ASVAB Word Knowledge	+**		
SAT Math	+**		
SAT Verbal	+**		
Late for Classes		-**	
Regularly Break Rules		-**	
Took Extra Classes/Lessons		+**	
Ever Took Classes During School Break		-	
Reason Took Classes During Break		+**	
Have High Standards at Work			+**
Make Every Effort to Do What is Expected			+**
Percent Chance Work at Age 30			+**
Parental Assessment of Age-30 Work Pr.			+
Panel B: Learning Outcomes			
White Collar Log Wages	+**	-**	+**
Blue Collar Log Wages	+**	-**	+**
4-year Science Grades	+**	+**	-
4-year Non-Science Grades	+	+**	+
2-year Grades	+	-	-
Panel C: White Collar Offer Arrival			
Receive White Collar Offer	+**	+	-
Panel D: Logit for Graduation in $t+1$			
Graduate in $t+1$	-	+**	-
Panel E: Flow Utilities			
2-year College	-	+**	-
4-year Science	+**	+*	-
4-year Non-Science	+	+**	+
Work Part-Time	-	+	-**
Work Full-Time	-	+	-
Work White Collar	+	+	+
Home Production	(Ref.)	(Ref.)	(Ref.)

Notes: * indicates statistical significance at the 10% level. ** indicates statistical significance at the 5% level. "(Ref.)" indicates that the coefficient is normalized as the reference category. See Tables B5–B7 (measurement system), 7 (log wages), 6 (grades), B10 (offer arrival), B9 (graduation), and 10 (flow utilities) for exact parameter estimates.

Table B4: Estimates of Probability Mass for Each Unobserved Type

Type Identity	Mass Probability	Std. Error
(H, H, H)	0.205	(0.013)
(H, H, L)	0.215	(0.015)
(H, L, H)	0.029	(0.009)
(H, L, L)	0.135	(0.014)
(L, H, H)	0.153	(0.011)
(L, H, L)	0.041	(0.008)
(L, L, H)	0.074	(0.012)
(L, L, L)	0.147	(0.010)

Notes: Bootstrap standard errors in parentheses. Type dummy labels are as follows: "H" signifies "high type"; "L" signifies "low type". Labels are ordered as { Schooling ability, Schooling preferences, Work motivation }. e.g. "Unobserved type (H, L, H)" corresponds to a worker with high schooling ability, low schooling preferences, and high work motivation. Labels are identified through the measurement system detailed in Appendix C.

Table B5: Measurement System Estimates for Schooling Ability Measurements

	ASVAB						SAT	
	Arithmetic	Coding	Mathematical	Numerical	Paragraph	Word		
Variable	Reasoning	Speed	Knowledge	Operations	Comprehension	Knowledge	Math	Verbal
Constant	-0.448	-0.556	-0.593	-0.532	-0.600	-0.343	-0.474	-0.481
	(0.040)	(0.056)	(0.037)	(0.054)	(0.043)	(0.044)	(0.046)	(0.062)
Black	-0.671	-0.326	-0.519	-0.217	-0.514	-0.553	-0.671	-0.246
	(0.035)	(0.052)	(0.032)	(0.046)	(0.034)	(0.035)	(0.045)	(0.054)
Hispanic	-0.355	-0.117	-0.237	-0.155	-0.223	-0.362	-0.244	0.027
	(0.039)	(0.050)	(0.037)	(0.049)	(0.038)	(0.035)	(0.050)	(0.056)
Born in 1980	-0.094	-0.048	-0.027	-0.040	-0.094	-0.018	-0.263	-0.228
	(0.045)	(0.055)	(0.039)	(0.054)	(0.046)	(0.046)	(0.053)	(0.065)
Born in 1981	-0.019	0.008	0.014	-0.008	0.011	0.010	-0.160	-0.219
	(0.040)	(0.052)	(0.040)	(0.056)	(0.040)	(0.039)	(0.056)	(0.063)
Born in 1982	-0.087	-0.138	-0.079	-0.091	-0.143	-0.040	-0.147	-0.305
	(0.041)	(0.058)	(0.041)	(0.050)	(0.043)	(0.040)	(0.056)	(0.064)
Born in 1983	-0.042	-0.010	-0.006	-0.038	-0.025	0.039	-0.068	-0.194
	(0.041)	(0.055)	(0.039)	(0.049)	(0.043)	(0.043)	(0.050)	(0.063)
Parent graduated college	0.340	0.261	0.367	0.188	0.336	0.278	0.299	0.182
	(0.032)	(0.039)	(0.030)	(0.041)	(0.030)	(0.031)	(0.039)	(0.048)
Family Income ($10,000)	0.016	0.008	0.016	0.010	0.015	0.021	0.020	0.024
	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.005)
Schooling ability type H	1.349	0.966	1.315	1.100	1.263	0.974	0.930	0.719
	(0.029)	(0.036)	(0.028)	(0.034)	(0.029)	(0.027)	(0.036)	(0.038)
Std. Dev. of noise	0.621	0.847	0.595	0.769	0.647	0.657	0.810	0.954
	(0.010)	(0.013)	(0.009)	(0.012)	(0.009)	(0.010)	(0.011)	(0.015)
Observations	2,136	2,122	2,134	2,122	2,135	2,136	1,232	1,223

Notes: Bootstrap standard errors are listed below each coefficient in parentheses. Each column represents estimates of a linear regression model with normally distributed errors, estimated by maximum likelihood.

Table B6: Measurement System Estimates for Schooling Preferences Measurements

Variable	No. times late for school	Break rules regularly	Hours per week extra classes	Ever took classes during school break	Reason took classes during school break
Constant			-7.165	-3.005	-0.593
			(0.646)	(0.223)	(0.139)
Black	0.488	-0.391	1.461	0.627	-0.603
	(0.104)	(0.099)	(0.580)	(0.172)	(0.128)
Hispanic	0.607	-0.137	0.308	0.753	-0.266
	(0.128)	(0.100)	(0.577)	(0.192)	(0.150)
Born in 1980	1.284	0.148		1.193	-0.788
	(0.138)	(0.126)		(0.202)	(0.159)
Born in 1981	0.908	0.092		0.739	-0.389
	(0.157)	(0.119)		(0.212)	(0.123)
Born in 1982	0.534	0.324	-1.498	0.604	
	(0.154)	(0.125)	(0.528)	(0.211)	
Born in 1983	0.193	-0.063	-2.263		
	(0.146)	(0.137)	(0.542)		
Parent graduated college	0.029	-0.120	1.980	0.329	-0.010
	(0.097)	(0.088)	(0.457)	(0.145)	(0.125)
Family Income ($10,000)	0.010	0.000	0.236	-0.011	0.028
	(0.010)	(0.009)	(0.044)	(0.018)	(0.013)
Schooling preference type H	-0.423	-0.644	1.310	-0.119	2.334
	(0.111)	(0.084)	(0.540)	(0.156)	(0.138)
Cut point 1	1.245	-1.613			
	(0.146)	(0.131)			
Cut point 2	1.736	-0.648			
	(0.151)	(0.124)			
Cut point 3	2.160	-0.112			
	(0.151)	(0.128)			
Cut point 4	2.560	0.292			
	(0.152)	(0.129)			
Cut point 5	2.813	1.022			
	(0.153)	(0.127)			
Cut point 6	3.163	1.802			
	(0.160)	(0.129)			
Cut point 7	4.072				
	(0.173)				
Std. Dev. of noise			7.877		
			(0.261)		
Observations	2,303	2,088	1,386	1,141	151

Notes: Bootstrap standard errors are listed below each coefficient in parentheses. The first two columns are estimates of ordered logit models, where "Break rules regulary" is on a Likert scale with seven levels. "Hours per week in extra classes" is a Type II Tobit model left-censored at zero hours. "Ever took classes during school break" and "Reason took classes during school break" are binary logit models with respective positive categories of "Yes" and "In order to accelerate, for fun, or for enrichment" and respective reference categories "No" and "To make up classes or for other reasons."

Table B7: Measurement System Estimates for Work Motivation Measurements

Variable	High standards at work	Try to do what is expected	Individual's subjective likelihood of working at age 30	Parent's subjective likelihood of individual working at age 30
Black	0.776	0.151	-0.251	-0.206
	(0.149)	(0.129)	(0.136)	(0.194)
Hispanic	-0.079	0.129	-0.366	0.107
	(0.143)	(0.134)	(0.132)	(0.223)
Born in 1980	0.056	0.061	-0.073	0.037
	(0.167)	(0.150)	(0.140)	(0.188)
Born in 1981	0.389	0.470		
	(0.174)	(0.149)		
Born in 1982	0.138	0.186		
	(0.169)	(0.156)		
Born in 1983	-0.073	0.098		
	(0.164)	(0.151)		
Parent graduated college	0.001	-0.284	-0.100	-0.149
	(0.113)	(0.104)	(0.123)	(0.191)
Family Income ($10,000)	0.050	0.012	0.013	0.075
	(0.014)	(0.011)	(0.013)	(0.023)
Work motivation type H	3.472	3.619	0.544	0.197
	(0.187)	(0.233)	(0.134)	(0.154)
Cut point 1	-4.500	-4.298	-2.247	-2.969
	(0.416)	(0.320)	(0.114)	(0.204)
Cut point 2	-3.738	-3.027	-1.148	-1.949
	(0.280)	(0.183)	(0.109)	(0.168)
Cut point 3	-2.586	-2.073		
	(0.182)	(0.168)		
Cut point 4	-1.693	-0.835		
	(0.157)	(0.144)		
Cut point 5	-0.080	0.879		
	(0.160)	(0.157)		
Cut point 6	2.410	3.400		
	(0.214)	(0.234)		
Observations	2,085	2,087	915	849

Notes: Bootstrap standard errors are listed below each coefficient in parentheses. Each column represents estimates of an ordered logit model. The first two columns are Likert scales with seven levels. The latter two columns are on a scale of 0%–100% that has been discretized into three bins: 0%–75%, 76%–90%, and 91%+

Table B8: Labor market shock forecasting estimates

Parameter	Estimate	Std. Error
Autocorrelation	0.504	(0.115)
Std. Dev. of shock	0.020	(0.002)
Observations	16	

Notes: Bootstrap standard errors in parentheses. We estimate a single AR1 process for both labor market sectors.

Table B9: Estimates of Probability of Graduation

Variable	Coeff.	Std. Error
Constant	-3.557	(0.331)
Black	-0.834	(0.146)
Hispanic	-0.485	(0.151)
HS Grades (z-score)	0.253	(0.056)
Parent graduated college	-0.024	(0.103)
Born in 1980	0.605	(0.155)
Born in 1981	0.329	(0.137)
Born in 1982	0.353	(0.130)
Born in 1983	0.499	(0.142)
Family Income ($10,000)	0.017	(0.010)
College experience completion profiles:		
0 years of 2yr	0.458	(0.216)
2+ years of 2yr	0.520	(0.190)
2 years of 4yr	1.355	(0.213)
3 years of 4yr	1.738	(0.219)
4 years of 4yr	2.728	(0.306)
5 years of 4yr	0.768	(14.847)
6+ years of 4yr	0.905	(0.395)
2 years of 4yr and 0 years of 2yr	-2.805	(0.318)
4 years of 4yr and 0 years of 2yr	-0.574	(0.327)
5 years of 4yr and 0 years of 2yr	2.163	(14.831)
6+ years of 4yr and 0 years of 2yr	1.360	(0.455)
Science major	-0.336	(0.141)
Prior ability science × Science major	1.848	(0.298)
Prior ability non-sci. × Non-Sci. major	1.351	(0.179)
Work part-time	-0.063	(0.128)
Work full-time	0.361	(0.142)
Schooling ability type H	-0.096	(0.155)
Schooling preference type H	0.783	(0.322)
Work motivation type H	-0.043	(0.189)
Person-year observations	1,115	

Notes: Parameter estimates from a logit predicting probability of graduating in the following period. Estimated only on four-year college students in their junior year and above. Bootstrap standard errors in parentheses. Reference categories for multinomial variables are as follows: "White" for race/ethnicity, "Born in 1984" for birth year, "1 year of 2yr college" and "3 years of 4yr college and 0 years of 2yr college" for college experience, "Not working" for work intensity, and "L" for each unobserved type.

Table B10: Estimates of White Collar Offer Arrival Parameters

Variable	Coeff.	Std. Error
Constant	-0.811	(0.268)
Age	-0.164	(0.015)
College graduate	1.550	(0.110)
Schooling ability type H	0.753	(0.114)
Schooling preference type H	0.119	(0.099)
Work motivation type H	-0.106	(0.091)
Person-year observations	22,398	

Notes: Estimates of the δ_λ parameters in Equation (H.4). Bootstrap standard errors in parentheses. Age is normalized to be zero at 18 years old. Reference category is "L" for each unobserved type. We restrict the offer arrival probability to equal 1 for those who worked in the white-collar sector in the previous period.

Table B11: Parameter Estimates of Static Choice Model

Variable	2-year	4-year Sci	4-year Non-Sci	Work PT	Work FT	White Collar
Constant	-2.887	-5.653	-4.181	-1.848	-1.034	-2.280
	(0.146)	(0.252)	(0.166)	(0.103)	(0.107)	(0.241)
Black	-0.117	0.054	0.189	-0.044	-0.106	0.027
	(0.079)	(0.099)	(0.082)	(0.057)	(0.053)	(0.090)
Hispanic	0.121	-0.062	-0.022	-0.086	0.022	0.114
	(0.071)	(0.110)	(0.083)	(0.065)	(0.056)	(0.089)
HS Grades (z-score)	0.191	0.839	0.591	-0.029	-0.052	0.099
	(0.032)	(0.048)	(0.033)	(0.027)	(0.022)	(0.034)
Parent graduated college	0.485	0.948	0.871	-0.033	-0.331	0.532
	(0.064)	(0.091)	(0.068)	(0.055)	(0.042)	(0.093)
Born in 1980	0.052	0.040	0.079	0.196	0.244	0.402
	(0.090)	(0.139)	(0.103)	(0.076)	(0.062)	(0.107)
Born in 1981	-0.095	-0.135	-0.015	0.190	0.190	0.299
	(0.088)	(0.123)	(0.093)	(0.065)	(0.058)	(0.112)
Born in 1982	-0.105	0.205	-0.166	0.181	0.125	0.282
	(0.088)	(0.125)	(0.092)	(0.066)	(0.061)	(0.113)
Born in 1983	-0.053	-0.029	-0.024	0.117	0.018	0.165
	(0.083)	(0.121)	(0.087)	(0.064)	(0.057)	(0.108)
Family Income ($10,000)	0.017	0.054	0.063	-0.029	-0.013	0.006
	(0.008)	(0.009)	(0.008)	(0.006)	(0.005)	(0.009)

continued

Variable	2-year	4-year Sci	4-year Non-Sci	Work PT	Work FT	White Collar
Age	-0.180	-0.189	-0.226	-0.134	-0.147	0.181
	(0.017)	(0.023)	(0.019)	(0.013)	(0.013)	(0.021)
Age squared	0.005	0.002	0.001	0.001	-0.004	-0.004
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Experience	0.016	-0.116	-0.017	0.303	0.383	-0.042
	(0.038)	(0.060)	(0.042)	(0.024)	(0.017)	(0.029)
Experience squared	-0.006	0.016	-0.002	-0.019	-0.012	0.003
	(0.003)	(0.004)	(0.003)	(0.002)	(0.001)	(0.002)
Years of college	-0.063	0.534	0.405	0.308	0.236	0.345
	(0.055)	(0.075)	(0.046)	(0.036)	(0.028)	(0.040)
Years of college squared	0.021	-0.015	-0.009	-0.024	-0.004	-0.034
	(0.007)	(0.008)	(0.005)	(0.004)	(0.003)	(0.004)
Prior academic ability	0.846	2.429	2.020			
	(0.110)	(0.290)	(0.137)			
Accumulated debt ($1,000)	-0.003	-0.003	0.010	-0.006	0.006	
	(0.004)	(0.006)	(0.004)	(0.003)	(0.002)	
Accumulated debt squared $\div$100	-0.011	-0.010	-0.007	0.005	-0.005	
	(0.002)	(0.004)	(0.003)	(0.002)	(0.001)	
Non-grad $\times \mathbb{E}[u(\text{consumption})] \div 1,000$	1.183	1.183	1.183	1.183	1.183	
	(0.199)	(0.199)	(0.199)	(0.199)	(0.199)	
Previous high school	1.009	2.803	1.845	1.157	0.873	-0.501
	(0.107)	(0.183)	(0.120)	(0.093)	(0.085)	(0.216)
Previous 2-year college	2.456	1.282	0.705	0.031	0.289	0.120

continued

Variable	2-year	4-year Sci	4-year Non-Sci	Work PT	Work FT	White Collar
	(0.085)	(0.196)	(0.094)	(0.086)	(0.080)	(0.142)
Previous 4-year science	1.267	4.782	2.159	0.701	0.398	-0.159
	(0.172)	(0.170)	(0.132)	(0.117)	(0.093)	(0.156)
Previous 4-year non-science	-0.047	2.028	3.384	0.580	0.573	0.053
	(0.165)	(0.191)	(0.092)	(0.090)	(0.077)	(0.109)
Previous work part-time	0.041	0.592	0.391	2.177	1.349	-0.976
	(0.092)	(0.128)	(0.102)	(0.059)	(0.055)	(0.099)
Previous work full-time	0.129	0.279	0.573	0.951	2.241	-0.941
	(0.099)	(0.138)	(0.100)	(0.063)	(0.042)	(0.103)
Previous work white-collar	-0.040	-0.590	-0.164	-1.278	-1.465	2.683
	(0.164)	(0.243)	(0.161)	(0.088)	(0.080)	(0.185)
Currently work white-collar				-0.139		
				(0.082)		
Currently work part-time	0.921	-0.266	-0.004			
	(0.072)	(0.090)	(0.073)			
Currently work full-time	-0.420	-1.783	-1.843			
	(0.070)	(0.097)	(0.066)			
Currently work part-time in white collar	-0.286	0.289	0.073			
	(0.188)	(0.182)	(0.157)			
Currently work full-time in white collar	-0.058	0.741	0.240			
	(0.165)	(0.212)	(0.137)			
Schooling ability type H	0.209	1.274	0.750	0.004	0.013	0.002
	(0.067)	(0.097)	(0.069)	(0.046)	(0.040)	(0.086)

continued

Variable	2-year	4-year Sci	4-year Non-Sci	Work PT	Work FT	White Collar
Schooling preference type H	0.408	0.154	0.535	-0.319	-0.664	-0.192
	(0.121)	(0.169)	(0.145)	(0.077)	(0.077)	(0.081)
Work motivation type H	0.055	0.111	0.214	-0.231	0.212	0.228
	(0.081)	(0.116)	(0.086)	(0.054)	(0.053)	(0.069)
College graduate				0.485	0.953	0.667
				(0.216)	(0.213)	(0.169)
Black × col. grad.				0.264	-0.055	0.665
				(0.174)	(0.166)	(0.139)
Hispanic × col. grad.				-0.065	-0.172	-0.060
				(0.197)	(0.151)	(0.140)
HS Grades (z-score) × col. grad				-0.203	-0.312	0.335
				(0.070)	(0.055)	(0.061)
Parent grad. col. × col. grad.				0.363	0.429	-0.584
				(0.130)	(0.103)	(0.105)
Born in 1980 × col. grad.				0.153	0.455	-1.102
				(0.197)	(0.150)	(0.137)
Born in 1981 × col. grad.				-0.100	0.202	-0.766
				(0.200)	(0.143)	(0.152)
Born in 1982 × col. grad.				-0.826	-0.137	-0.758
				(0.213)	(0.153)	(0.155)
Born in 1983 × col. grad.				-0.690	-0.184	-0.524
				(0.193)	(0.143)	(0.169)
Family Income ($10,000) × col. grad.				-0.024	-0.033	0.037

continued

Variable	2-year	4-year Sci	4-year Non-Sci	Work PT	Work FT	White Collar
				(0.010)	(0.010)	(0.012)
Col. grad $\times \mathbb{E}[u(\text{consumption})] \div 1,000$				-0.169	-0.169	
				(0.233)	(0.233)	
Constant Relative Risk Aversion parameter (θ)	0.4					
Log likelihood	-26,109					
Person-year obs.	22,398					

Notes: Home production is the reference alternative. Bootstrap standard errors are listed below each coefficient in parentheses. Beliefs on labor market productivity are included in the expected utility of consumption term. Consumption is evaluated in terms of yearly consumption flow in 1996 dollars. Missing majors are estimated to be science with probability 0.35. Missing GPAs are estimated to be ≤ 2.5 w.p. 0.48, 2.5–3.0 w.p. 0.21, 3.0–3.6 w.p. 0.19, and 3.6–4.0 w.p. 0.13.

Reference categories for multinomial variables are as follows: “White” for race/ethnicity, “Born in 1984” for birth year, “Previous home production” for previous decision, “Not working” for in-college work intensity, and “L” for each unobserved type.

Table B12: Model fit: Overall choice frequencies

Choice alternative	Data Frequency (%)	Model Frequency (%)
2-year & work FT blue collar	1.70	1.86
2-year & work FT white collar	0.15	0.16
2-year & work PT blue collar	1.82	1.95
2-year & work PT white collar	0.10	0.10
2-year only	1.92	2.01
4-year Science & work FT blue collar	0.43	0.52
4-year Science & work FT white collar	0.15	0.16
4-year Science & work PT blue collar	0.94	0.97
4-year Science & work PT white collar	0.14	0.16
4-year Science only	2.27	2.37
4-year Non-Science & work FT blue collar	0.90	1.10
4-year Non-Science & work FT white collar	0.20	0.24
4-year Non-Science & work PT blue collar	1.88	2.29
4-year Non-Science & work PT white collar	0.29	0.29
4-year Non-Science only	4.06	4.72
Work PT blue collar	6.57	6.45
Work PT white collar	0.96	0.99
Work FT blue collar	42.71	41.95
Work FT white collar	8.61	8.93
Home production	24.22	22.77

Note: Model frequencies are constructed using 10 simulations of the structural model for each individual included in the estimation. We set the panel length in the model to be the same as the panel length in the data. This is because the model assumes random attrition conditional on all observables and unobservables. White collar offer probability in simulation is 0.3183 and in estimation is 0.3077.

Table B13: Model fit: Graduate choice frequencies

Choice alternative	Data Frequency (%)	Model Frequency (%)
Work PT blue collar	4.09	4.07
Work PT white collar	4.38	3.72
Work FT blue collar	34.54	37.25
Work FT white collar	43.66	41.35
Home production	13.34	13.61

Note: Model frequencies are constructed using 10 simulations of the structural model for each individual included in the estimation.
White collar offer probability in simulation is 0.6675 and in estimation is 0.7044.

Table B14: Average abilities by employment sector and education level at age 28 in baseline and counterfactual models

			White Collar			Blue Collar			4-year Science			4-year Non-Science			2-year		
Sector and education level			Baseline	Cfl	N.F. Cfl	Baseline	Cfl	N.F. Cfl	Baseline	Cfl	N.F. Cfl	Baseline	Cfl	N.F. Cfl	Baseline	Cfl	N.F. Cfl
Full-time	White Collar	Science	0.03	0.42	0.38	0.03	0.13	0.12	0.16	0.46	0.46	0.16	0.37	0.38	0.09	0.16	0.20
		Non-Science	0.01	0.30	0.24	-0.01	0.00	-0.01	0.10	0.26	0.23	0.15	0.45	0.43	0.13	0.31	0.32
		Non-graduate	-0.01	0.27	0.10	-0.00	0.04	-0.01	-0.07	-0.09	-0.13	-0.10	-0.13	-0.16	-0.09	-0.09	-0.14
	Blue Collar	Science	0.03	0.07	0.06	0.01	0.11	0.11	0.12	0.46	0.44	0.14	0.36	0.34	0.11	0.22	0.19
		Non-Science	0.01	-0.01	-0.03	-0.02	-0.01	-0.02	0.11	0.27	0.25	0.16	0.43	0.40	0.15	0.35	0.33
		Non-graduate	-0.00	-0.03	-0.05	0.01	0.03	0.02	-0.03	-0.13	-0.15	-0.04	-0.18	-0.19	-0.03	-0.11	-0.12
Part-time	White Collar	Science	0.06	0.31	0.28	0.06	0.04	0.08	0.16	0.49	0.48	0.18	0.44	0.34	0.13	0.18	0.18
		Non-Science	0.02	0.14	0.10	-0.02	-0.12	-0.10	0.07	0.29	0.16	0.12	0.51	0.30	0.08	0.43	0.18
		Non-graduate	0.00	0.10	-0.02	-0.03	-0.06	-0.05	-0.06	-0.04	-0.20	-0.06	0.10	-0.22	-0.09	0.07	-0.17
	Blue Collar	Science	0.02	0.01	-0.02	0.01	0.04	0.01	0.17	0.35	0.34	0.11	0.30	0.25	0.05	0.18	0.18
		Non-Science	-0.06	-0.10	-0.12	-0.05	-0.08	-0.08	0.07	0.23	0.21	0.11	0.39	0.36	0.13	0.31	0.23
		Non-graduate	-0.02	-0.09	-0.10	-0.01	-0.05	-0.04	-0.04	-0.14	-0.15	-0.05	-0.15	-0.17	-0.01	-0.08	-0.11
Home		Science	0.07	-0.09	-0.01	0.06	-0.10	-0.04	0.15	0.36	0.40	0.14	0.30	0.31	0.14	0.11	0.13
		Non-Science	0.00	-0.11	-0.12	-0.01	-0.16	-0.16	0.07	0.22	0.21	0.12	0.42	0.41	0.09	0.29	0.31
		Non-graduate	-0.02	-0.17	-0.16	-0.02	-0.15	-0.13	-0.04	-0.19	-0.19	-0.05	-0.16	-0.19	-0.06	-0.13	-0.17

Notes: “Cfl” refers to the counterfactual while “N.F. Cfl” refers to the counterfactual with no search frictions.

Table B15: Average posterior variances after last period of college for different choice paths in baseline model

Choice Path	White Collar	Blue Collar	Science	Non-Science	2-year	Share(%)
Continuous enrollment, graduate in science with x years of in-school work experience						
$x = 0$	0.12	0.07	0.07	0.13	0.27	1.39
$x > 0$, white collar only	0.06	0.06	0.07	0.13	0.27	0.52
$x > 0$, blue collar only	0.10	0.03	0.07	0.12	0.26	3.63
$x > 0$, mixture	0.06	0.03	0.07	0.12	0.25	0.85
Continuous enrollment, graduate in non-science with x years of in-school work experience						
$x = 0$	0.12	0.07	0.09	0.07	0.23	3.03
$x > 0$, white collar only	0.06	0.06	0.09	0.07	0.23	1.04
$x > 0$, blue collar only	0.10	0.03	0.09	0.07	0.22	8.88
$x > 0$, mixture	0.06	0.03	0.08	0.07	0.21	1.90
Stop out (SO)						
SO, graduate in science	0.09	0.03	0.08	0.10	0.20	0.97
SO, graduate in non-science	0.09	0.03	0.09	0.08	0.19	3.45
SO then DO, start in 2yr	0.09	0.03	0.15	0.16	0.14	4.98
SO then DO, start in science	0.09	0.03	0.10	0.14	0.20	1.37
SO then DO, start in non-science	0.09	0.03	0.12	0.11	0.19	2.67
Truncated	0.08	0.02	0.11	0.12	0.17	5.76
Drop out (DO) after x years of school						
$x = 1$	0.11	0.05	0.17	0.21	0.25	16.59
$x = 2$	0.11	0.04	0.15	0.17	0.20	7.85
$x = 3$	0.10	0.04	0.13	0.15	0.19	4.22
$x = 4$	0.10	0.03	0.11	0.12	0.18	1.93
$x \geq 5$	0.09	0.03	0.10	0.10	0.16	1.85
Never attended college						
Never attend college	0.09	0.02	0.20	0.27	0.35	27.11
Time 0 population variance	0.13	0.07	0.21	0.27	0.35	

Notes: Average posterior variances of ability across individuals are reported in each cell. This table is constructed using 10 simulations of the baseline model for each individual included in the estimation.

"Truncated" refers to those who were enrolled in period 10.

Table B16: Average abilities for different choice paths in full-information no-search-frictions counterfactual scenario

Choice Path	White Collar	Blue Collar	Science	Non-Science	2-year	Share(%)
Continuous enrollment, graduate in science with x years of in-school work experience						
$x = 0$	0.27	-0.01	1.04	0.76	0.28	0.87
$x > 0$, white collar only	1.09	0.39	1.03	0.75	0.40	0.98
$x > 0$, blue collar only	0.50	0.39	0.98	0.67	0.32	6.33
$x > 0$, mixture	1.01	0.53	0.98	0.69	0.30	2.92
Continuous enrollment, graduate in non-science with x years of in-school work experience						
$x = 0$	-0.00	-0.38	0.54	0.87	0.58	1.67
$x > 0$, white collar only	0.40	-0.08	0.50	0.78	0.50	1.05
$x > 0$, blue collar only	0.12	-0.11	0.55	0.81	0.54	8.67
$x > 0$, mixture	0.45	-0.02	0.44	0.74	0.56	2.93
Stop out (SO)						
SO, graduate in science	0.46	0.28	0.87	0.62	0.34	1.94
SO, graduate in non-science	0.10	-0.14	0.47	0.72	0.49	4.72
SO then DO, start in 2yr	-0.19	-0.05	-0.34	-0.34	-0.12	2.87
SO then DO, start in science	-0.10	0.03	-0.07	-0.16	-0.10	1.03
SO then DO, start in non-science	-0.28	-0.18	-0.07	-0.00	-0.02	2.60
Truncated	-0.05	-0.08	0.03	0.07	0.04	5.07
Drop out (DO) after x years of school						
$x = 1$	-0.17	-0.07	-0.37	-0.40	-0.25	14.84
$x = 2$	-0.21	-0.09	-0.26	-0.28	-0.14	5.80
$x = 3$	-0.23	-0.14	-0.25	-0.20	-0.05	2.60
$x = 4$	-0.20	-0.04	-0.06	-0.03	0.04	1.24
$x \geq 5$	-0.11	-0.08	0.02	-0.05	-0.06	0.68
Never attended college						
Never attend college	-0.17	-0.02	-0.46	-0.50	-0.32	31.20

Notes: Abilities are reported in standard deviation units. This table is constructed using 10 simulations of the counterfactual model described in the title for each individual included in the estimation.

"Truncated" refers to those who were enrolled in period 10.

C Recovering unobserved types and missing data using an EM algorithm

This appendix section provides an overview of how we recover the unobserved heterogeneity types and missing data on college GPA and major using an EM algorithm. Aside from standard unobserved heterogeneity, we treat the first missing major or GPA observation as a permanent unobserved type, which we integrate out using a modified version of the EM algorithm detailed in Subsection 5.5.

This first stage of our estimation incorporates a measurement system to aid in identification of the unobserved heterogeneity types, as well as flexible specifications for all outcomes that enter the structural model (i.e. choices, wages, grades, and college graduation). The goal of this first stage is to obtain estimates of the posterior unobserved type probabilities for each individual as well as the probability mass on each type. All other parameters are treated as nuisance parameters at this stage of the estimation.

Appendix Subsection C.6 briefly describes how we use the first stage to subsequently estimate the structural parameters.

C.1 Overview of the unobserved heterogeneity and missing data types

C.1.1 Unobserved heterogeneity

We assume that there are three different dimensions of discrete unobserved heterogeneity types. Each of the dimensions is associated with S_d^r heterogeneity types, for a total of $S^r = \prod_{d=1}^{3} S_d^r$ types:

1. Schooling ability
2. Schooling preferences
3. Work motivation

In practice, for our baseline specification we set each $S_d^r = 2$ for a total number of $S^r = 8$ types.

C.1.2 Missing college GPA

We also assume that there are four latent types for the missing college GPA observations, where the types are characterized by quartiles of the observed GPAs g:

1. $g \in [0, 2.5]$
2. $g \in (2.5, 3.0]$
3. $g \in (3.0, 3.6]$
4. $g \in (3.6, 4.0]$

C.1.3 Missing college major

Finally, as we model college majors in a binary fashion, the latent majors that are missing follow the same structure:

1. Science majors
2. All other majors

C.2 Dependent variables used in the first stage

C.2.1 Description of measurements used

Here we outline the measurements used for each of the dimensions of unobserved types described in C.1.1.

Schooling ability The schooling ability type is measured from cognitive test scores (each of which has been z-scored relative to the entire NLSY97 sample) taken from the Armed Services Vocational Aptitude Battery (ASVAB) and the SAT I exam:

- ASVAB Arithmetic Reasoning
- ASVAB Coding Speed
- ASVAB Mathematical Knowledge
- ASVAB Numerical Operations
- ASVAB Paragraph Comprehension
- ASVAB Word Knowledge
- SAT I Math
- SAT I Verbal

Each of these cognitive test scores has a continuous distribution.

Schooling preferences The schooling preferences type is measured from various discrete outcomes, such as Likert-scaled questions:

- Number of times the individual reported being late for school without excuse
- How strongly the individual agrees with the following statement: "When I was in school, I broke the rules regularly"
- How many hours per week the individual spent taking extra classes (such as music lessons, etc.)
- If the individual ever took classes during a school break (this could either be for remedial or accelerative reasons)
- If the individual took classes during break, the reason for doing so (e.g. "To accelerate, for fun, for enrichment" or "To make up classes")

Each of these measures is discrete, taking on various numbers of effective categories. For the number of hours spent in extra classes, we treat the distribution as being censored at zero.

Work motivation We have no measures that separately inform us about the individual's work ability and preferences, so the final dimension of unobserved heterogeneity types is combined into a factor we call "work motivation." As in the case of schooling preferences, the measurements of work motivation are discrete:

- How strongly the individual agrees with the following statement: "I have high standards at work"
- How strongly the individual agrees with the following statement: "I make every effort to do what is expected of me"
- The individual's subjective likelihood of working part- or full-time at age 30 (reported as a percent chance on a scale from 0-100)
- The parent's subjective likelihood of the individual working part- or full-time at age 30 (reported on the same scale as above)

C.2.2 Description of other dependent variables in the first stage

Aside from the measurement system, our first-stage estimation requires data on wages in each occupation, college grades in each sector, college graduation outcomes, and choices about schooling, work intensity, and occupations. We include these as follows:

- Each of the five learning outcomes (white-collar log wages, blue-collar log wages, 4-year science grades, 4-year non-science grades, and 2-year college grades)
- Graduation in $t+1$ conditional on being eligible to graduate (i.e., at least a junior in college).
- Binary variables for the following choices:
  - Enrolling in any type of college
  - Enrolling in a 4-year college, conditional on having enrolled in any type of college
  - Choosing a 4-year science major, conditional on being in 4-year college
  - Choosing any work option
  - Working full-time, conditional on choosing any work option
  - Working white-collar, conditional on choosing any work option

C.3 Overall likelihood

We now characterize the overall likelihood for the first-stage auxiliary model.

C.3.1 Measurement System

To estimate the measurement system, we use a variety of likelihood-based models. Each of the left-hand side variables is a measurement described in the previous section, and the right-hand side variables consist of demographic and family background variables, as well as dummies for each unobserved type (i.e. whether the type identity is "High" in the first, second, or third dimension). As demographic and family background variables, we use indicators for race/ethnicity, birth year dummies, whether either of the individual's parents is a college graduate, and the individual's family income when he was a teenager (in thousands of dollars).

Indexing individuals by i, we estimate from the measurement data, $(y_{i,b})_{i,b}$, the parameters of the following equation for each measurement b and heterogeneity type r:

$$y_{ib} = X_i \tilde{\beta}_b + \omega_{br} + \varepsilon_{ib} \tag{C.1}$$

where $y_{i,b} = \sum_r y_{i,b,r} 1_{\{R_i = r\}}$, X_i consists of the demographic and family background variables, ω_{br} is a measurement- and type-specific intercept, and ε_{ib} denotes measurement errors.

Types are interpreted through normalizations on the ω_{br} coefficients. For example, only the schooling ability type determines the measurements of schooling ability, while the other type coefficients are normalized to zero.

Likelihood for continuous measurements For measurements that are continuous, we assume that $\varepsilon_{ib} \sim N\left(0, \sigma_b^2\right)$ which yields the following likelihood:

$$L_{ibr}^{co}\left(\Theta; y_{ib}, X_i\right) = \frac{1}{\sigma_b} \varphi\left(\frac{y_{ib} - X_i \tilde{\beta}_b - \omega_{br}}{\sigma_b}\right) \tag{C.2}$$

where $\varphi\left(\cdot\right)$ is the density of the standard normal distribution.

Likelihood for censored measurements For measurements that are censored below at value $\underline{y}$, we use the Type I Tobit likelihood. We modify the likelihood to allow for a third case where we know that the value of y is above $\underline{y}$ but do not know its exact value:

$$\begin{aligned} L_{ibr}^{ce}\left(\Theta; y_{ib}, X_i, \underline{y}\right) = & \left\{\frac{1}{\sigma_b} \varphi\left(\frac{y_{ib} - X_i \tilde{\beta}_b - \omega_{br}}{\sigma_b}\right)\right\}^{I(y_{ib} > \underline{y},\, y_{ib} \text{ observed})} \times \\ & \left\{1 - \Phi\left(\frac{X_i \tilde{\beta}_b + \omega_{br} - \underline{y}}{\sigma_b}\right)\right\}^{I(y_{ib} = \underline{y})} \times \\ & \left\{\Phi\left(\frac{X_i \tilde{\beta}_b + \omega_{br} - \underline{y}}{\sigma_b}\right)\right\}^{I(y_{ib} > \underline{y},\, y_{ib} \text{ unobserved})} \end{aligned} \tag{C.3}$$

where $I\left(\cdot\right)$ is the indicator function and $\Phi\left(\cdot\right)$ is the CDF of the standard normal distribution.

Likelihood for discrete ordered measurements For measurements that are discrete but have an inherent ordering (e.g. degree to which individual agrees with various statements; etc.), we assume that ε_{ib} is consistent with the ordered logit model. This yields the following likelihood:

$$L^{o}_{ibr}\left(\Theta; y_{ib}, X_{i},\right) = \prod_{j=1}^{J^{b}} P_{o,ijbr}^{1[y_{ib}=j]} \tag{C.4}$$

where $1\left[\cdot\right]$ is the indicator function, and where

$$\begin{aligned} P_{o,ijbr} &= \Pr\left(y_{ib} = j\right) \\ &= \Pr\left(\kappa_{j-1,b} < X_{i}\tilde{\beta}_{b} + \omega_{br} \leq \kappa_{jb}\right) \\ &= \frac{1}{1 + \exp\left(X_{i}\tilde{\beta}_{b} + \omega_{br} - \kappa_{jb}\right)} - \frac{1}{1 + \exp\left(X_{i}\tilde{\beta}_{b} + \omega_{br} - \kappa_{j-1,b}\right)} \end{aligned} \tag{C.5}$$

where $\kappa_{0b} = -\infty$ and $\kappa_{Jb} = \infty$.

Likelihood for discrete unordered measurements For measurements that are discrete but have no inherent ordering (e.g. did the individual take extra classes; what was the reason the individual took extra classes; etc.), we assume that ε_{ib} is consistent with the multinomial logit model. This yields the following likelihood:

$$L^{u}_{ibr}\left(\Theta; y_{ib}, X_{i}\right) = \prod_{j=1}^{J^{b}} P_{u,ijbr}^{1[y_{ib}=j]} \tag{C.6}$$

where

$$P_{u,ijbr} = \frac{\exp\left(X_{i}\tilde{\beta}_{jb} + \gamma_{jbr}\right)}{\sum_{k}\exp\left(X_{i}\beta_{kb} + \gamma_{kbr}\right)} \tag{C.7}$$

Measurement system likelihood The joint likelihood function for the entire measurement system (conditional on unobserved type $R = r$) is then

$$L^{(1,s)}_{ir} = \prod_{b\in\text{cont}} L^{co}_{ibr} \prod_{b\in\text{censo}} L^{ce}_{ibr} \prod_{b\in\text{ord}} L^{o}_{ibr} \prod_{b\in\text{unord}} L^{u}_{ibr} \tag{C.8}$$

where cont stands for "continuous"; censo stands for "censored"; ord stands for "ordered categorical"; and unord stands for "unordered categorical."

Summary of measurement system specification assumptions Table C1 summarizes the assumptions we make about the measurements.

Table C1: Summary of assumptions about measurements

Measurement	Latent type	Distribution	Estimator	Categories
ASVAB Arithmetic Reasoning	school ability	continuous	normal MLE	—
ASVAB Coding Speed	school ability	continuous	normal MLE	—
ASVAB Mathematical Knowledge	school ability	continuous	normal MLE	—
ASVAB Numerical Operations	school ability	continuous	normal MLE	—
ASVAB Paragraph Comprehension	school ability	continuous	normal MLE	—
ASVAB Word Knowledge	school ability	continuous	normal MLE	—
SAT I Math	school ability	continuous	normal MLE	—
SAT I Verbal	school ability	continuous	normal MLE	—
Number of times late for school	school prefs.	ordered categorical	ordered logit	0, 1, 2, 3, 4, 5, 6-10, or 11+ times
Broke rules regularly	school prefs.	ordered categorical	ordered logit	1-7 Likert scale
Hours per week took extra classes	school prefs.	censored ($\underline{y} = 0$)	normal MLE	—
Took class during break?	school prefs.	binary	binary logit	yes or no
Reason took class during break	school prefs.	binary	binary logit	1) for enrichment; 2) to catch up
Have high standards at work	work motivation	ordered categorical	ordered logit	1-7 Likert scale
Make every effort to do what is expected	work motivation	ordered categorical	ordered logit	1-7 Likert scale
Percent chance work at age 30	work motivation	ordered categorical	ordered logit	0-75%, 76-90%, 91-100%
Parent: percent chance i works at age 30	work motivation	ordered categorical	ordered logit	0-75%, 76-90%, 91-100%

C.3.2 Other first-stage equations

Estimation of the other first-stage equations proceeds in a similar fashion as the measurement system. The primary difference is that each of these equations uses panel data, whereas each of the measurements in the measurement system is purely cross-sectional. As right-hand side variables, we employ the same X_i as in the measurement system, but we add time-varying schooling and work experiences, interactions between the two, previous decision dummies, college graduation and graduated major dummies (in the equations for wages and work decisions only), and calendar year dummies (in the wage equations only).

Of particular importance to this part of the first-stage estimation is our inclusion of the (standardized) cumulative sum of lagged learning outcomes in each equation. We denote this as a vector $\overline{\mathbf{y}}_{it} \equiv (\overline{y}_{i1t} \quad \cdots \quad \overline{y}_{iJt})$ which is formally defined as follows for each element:

$$\overline{y}_{ijt} = \frac{1}{\sum_{\tau=1}^{t-1} d_{ij\tau}} \sum_{\tau=1}^{t-1} \left\{ d_{ij\tau} \left[\frac{y_{ij\tau} - \mu_j}{\sigma_j} \right] \right\} \tag{C.9}$$

where μ_j is the mean of y_{ijt} over all i and t, σ_j is the standard deviation of y_{ijt} over all i and t, and $d_{ij\tau}$ indicates that $y_{ij\tau}$ is observed (i.e. the person chose to participate in sector j in period τ)

The inclusion of $\overline{\mathbf{y}}_{it}$ is crucial to separately identifying the ability that is learned about—which is unknown to both the individual and the researcher—from the permanent unobserved heterogeneity type ω_{br}—which is private to the individual.

We flexibly include $\overline{\mathbf{y}}_{it}$ in our auxiliary equations by interacting it with work experience variables, previous decision dummies, and unobserved type. However, to maintain estimation precision, we only interact the $\overline{y}_{ijt}$ (corresponding to the current equation's outcome) with the unobserved type. For example, in the estimation of white-collar wages, we only include the interaction between the $\overline{y}_{ijt}$ for white collar with the unobserved types. We also place some structure on the interaction between work experiences and $\overline{\mathbf{y}}_{it}$. For example, in the grades equations, we only allow the $\overline{\mathbf{y}}_{it}$ associated with college grades to be interacted with the work experiences and previous decision dummies. However, the entire set of components of $\overline{\mathbf{y}}_{it}$ (not interacted with anything) always enters each equation.

For individual i in time period t, we estimate the following equation for each outcome b:

$$y_{ibt} = X_{it}\tilde{\beta}_{1b} + \overline{\mathbf{y}}_{it}\tilde{\beta}_{2b} + X_{it}\overline{\mathbf{y}}_{it}\tilde{\beta}_{3b} + \omega_{1br} + \overline{y}_{ijt}\omega_{2br} + \varepsilon_{ibt} \tag{C.10}$$

where X_{it} consists of demographic and family background variables as well as previous decision dummies and full quadratic interactions for each of the work and schooling experiences, college graduation and graduated major dummies, and calendar year dummies (where applicable). $\overline{\mathbf{y}}_{it}$ is as defined above, and ω_{1br} and ε_{ibt} are as defined in Equation (C.1), albeit with slightly different notation which has been adapted to the panel data setting.

Unlike in the measurement system, we allow each type to enter each of the equations in this part of the model.

Likelihood for grades and wages The grades and wages are continuous; hence, when we assume ε_{ibt} is normally distributed, the likelihood for these equations closely mirrors that of Equation (C.2). Let b index the college sector or job occupation:

$$L_{ibr}^{\text{co}}\left(\Theta; y_{ibt}, X_{it}, \overline{\mathbf{y}}_{it}\right) = \frac{1}{\sigma_b}\varphi\left(\frac{y_{ibt} - X_{it}\tilde{\beta}_{1b} - \overline{\mathbf{y}}_{it}\tilde{\beta}_{2b} - X_{it}\overline{\mathbf{y}}_{it}\tilde{\beta}_{3b} - \omega_{1br} - \overline{y}_{ijt}\omega_{2br}}{\sigma_b}\right) \quad \text{(C.11)}$$

Likelihood for choices and graduation We model each of the choices and the graduation outcome as binary and assume that ε_{ib} is consistent with the logit model. This yields the following likelihood:

$$L_{ibr}^{u}\left(\Theta; y_{ibt}, X_{it}, \overline{\mathbf{y}}_{it}\right) = P_{u,ibrt}^{1[y_{ibt}=1]}\left(1 - P_{u,ibrt}\right)^{1[y_{ibt}=0]} \quad \text{(C.12)}$$

where

$$P_{u,ibrt} = \frac{\exp\left(X_{it}\tilde{\beta}_{1b} + \overline{\mathbf{y}}_{it}\tilde{\beta}_{2b} + X_{it}\overline{\mathbf{y}}_{it}\tilde{\beta}_{3b} + \omega_{1br} + \overline{y}_{ijt}\omega_{2br}\right)}{1 + \exp\left(X_{it}\tilde{\beta}_{1b} + \overline{\mathbf{y}}_{it}\tilde{\beta}_{2b} + X_{it}\overline{\mathbf{y}}_{it}\tilde{\beta}_{3b} + \omega_{1br} + \overline{y}_{ijt}\omega_{2br}\right)} \quad \text{(C.13)}$$

Likelihood for all non-measurement-system first-stage equations The likelihood for all non-measurement-system first-stage equations mimics Equation (C.8), but combines each equation characterized by (C.11) and (C.12):

$$L_{ir}^{(1,c)} = \prod_{b \in \text{cont}} L_{ibr}^{co} \prod_{b \in \text{binary}} L_{ibr}^{u} \quad \text{(C.14)}$$

where cont stands for "continuous" and binary stands for "binary."

C.3.3 Overall first-stage likelihood

The overall first-stage likelihood—conditional on unobserved type r—is then the product of the measurement system likelihood in Equation (C.8) and the other likelihood in Equation

(C.14):

$$L_{ir}^{(1)} = L_{ir}^{(1,s)} L_{ir}^{(1,c)} \tag{C.15}$$

where we superscript with (1) to emphasize that this is an auxiliary likelihood for the first-stage and should not be confused with the structural likelihood in Equation (36).

The overall log-likelihood for the first stage is then

$$\ell^{(1)} = \sum_i \ln \left[\sum_{r=1}^{R} \pi_r L_{ir}^{(1,s)} L_{ir}^{(1,c)} \right] \tag{C.16}$$

C.4 EM algorithm

We now describe how we use the EM algorithm to estimate the posterior unobserved type probabilities, the missing major and missing GPA type probabilities, and the probability masses on the unobserved types. The notation used throughout this subsection mirrors that which is used in Subsection 5.5. Additionally, we introduce two time-invariant indices: (*i*) $m \in \{\text{science}, \text{non-science}\}$, which indexes missing major; and (*ii*) $g \in \{1, \ldots 4\}$, which indexes missing GPA quartile.

C.4.1 E-step

At the E-step of our algorithm, we need to take appropriate likelihood contributions for each individual's observations. The key idea is that the entire string of future likelihood contributions depends on the missing choice that is being integrated over. Below, we list the joint probabilities of i being of a particular unobserved type r and unobserved major m or GPA quartile g (if either or both of these outcomes are missing).

$$\Pr(r|i) = q_{ir} = \frac{\pi_r L_{ir}^{(1)}}{\sum_{r'} \pi_{r'} L_{ir'}^{(1)}} \tag{C.17}$$

$$\Pr(r, m|i) = q_{irm} = \frac{\pi_r L_{ir}^{(1,m)}}{\sum_{m'} \sum_{r'} \pi_{r'} L_{ir'}^{(1,m')}} \tag{C.18}$$

$$\Pr(r, g|i) = q_{irg} = \frac{\pi_r L_{ir}^{(1,g)}}{\sum_{g'} \sum_{r'} \pi_{r'} L_{ir'}^{(1,g')}} \tag{C.19}$$

$$\Pr(r, m, g|i) = q_{irmg} = \frac{\pi_r L_{ir}^{(1,m,g)}}{\sum_{g'} \sum_{m'} \sum_{r'} \pi_{r'} L_{ir'}^{(1,m,g)}} \tag{C.20}$$

where $L_{ir}^{(1,m)}$ indicates that missing majors in $L_{ir}^{(1)}$ are evaluated major m. $L_{ir}^{(1,g)}$ and $L_{ir}^{(1,m,g)}$ are likewise defined for missing grades g and the combination of missing majors m and grades g. Thus, equation (C.17) holds for those who have no missing data, (C.18) holds for those who only have a missing major, (C.19) holds for those who only have a missing GPA, and (C.20) holds for those who have both outcomes missing. Note that, for the cases with missing grades, the contribution to the likelihood of the grades that are missing is computed by integrating over the relevant portion of the continuous grade distribution. For instance, if $g = g_1$ and $g = g_2$ denote the first and second quartiles of the grade distribution, the contribution is given by:

$$\Phi(g_2 - X'\beta_r) - \Phi(g_1 - X'\beta_r)$$

where Φ is the cdf. of a normal distribution with mean zero and variance σ^2 (variance of the residuals of the auxiliary grade equation) and β_r are the (type-specific) parameter estimates of the auxiliary grade equation. The contribution corresponding to the first and last quartiles are given by $\Phi(g_1 - X'\beta_r)$ and $1 - \Phi(g_3 - X'\beta_r)$, respectively.

C.4.2 M-step

At the M-step of our algorithm, we treat the three different unobserved types as observed and assign each individual a corresponding weight depending on his missing outcome status.

- all who *never* have a missing major or missing GPA: weight by q_{ir}
- those who have a missing major but no missing GPA: weight by q_{irm} for all observations of the individual
- those who have a missing GPA but no missing major: weight by q_{irg} for all observations of the individual
- those who have both a missing major and missing GPA: weight by q_{irmg} for all observations of the individual

We then estimate the system of logit and regression models that make up $L_{ir}^{(1)}$ using the q_{irmg}, q_{irm}, q_{irg}, or q_{ir} as the weights—depending on which of the four cases in Equations (C.17)–(C.20) applies.

C.5 First-stage estimation results

We report the estimates of the type mass probabilities in Table B4. Estimates of the missing major and GPA probabilities are included in the note to Table 10. The estimates of the measurement system are in Tables B5–B7. We do not report the estimates of the other first-stage equations, but these are available upon request.

C.6 Subsequent stage estimation

The estimation steps described above pertain only to the first stage of our model, in which we estimate the measurement system and recover the distribution of unobserved types, missing majors, missing GPAs, and individual-specific posterior type probabilities (the q's above).

In the subsequent stages of our estimation procedure, we take as given the q's listed above and then estimate the learning parameters (i.e. the γ_j's, γ_l's, λ_j's, λ_l's, Δ and σ^2) by weighted m-estimation.

Once we obtain estimates of the learning parameters, we take these estimates as given and then estimate the graduation probabilities, the aggregate labor market time series process, the CCPs (including white collar job offer arrival parameters), and the structural flow utility parameters. In each subsequent stage, we continue to use the weights described above for each individual, depending on his missing outcome status and first-stage auxiliary likelihood value.

D Details on consumption

This appendix details how we model consumption. Individuals make decisions in part based on their expected utility of consumption, where the utility of consumption follows a constant relative risk aversion form. Consumption is supported by some combination of labor income, parental transfers, educational grants or loans, or the social safety net. We detail below each of these components.

D.1 Wages

Our model includes log (hourly) wage equations for the blue- and white-collar sectors. We also allow individuals to choose between part-time and full-time employment. Denoting by w the log wage and by h the annual hours worked, we compute $E\left(W\right)$ using the work intensity and number of weeks worked, as follows (assuming log wages are normally distributed as in (6)):

$$E\left(W \mid X\right) = h \exp\left(\delta_t + \lambda_{0l} + \lambda_{1l}\left(\gamma_{0l} + X_{ilt}\gamma_{1l} + A_{il}\right) + \frac{\hat{\sigma}_{sl}^2}{2}\right) \tag{D.1}$$

where

$$h = \begin{cases} 40 \cdot 52 & \text{if working full-time in October} \\ 20 \cdot 52 & \text{if working part-time in October} \end{cases}$$

As we only model employment decisions in the month of October, we verify, using the individuals in our data, that the values chosen above correspond to the median individual choosing each work alternative in October. Specifically, the median full-time October worker reports working 40 hours per week and 52 weeks in the year. Likewise, the median part-time October worker reports working 20 hours per week and 52 weeks in the year. As a reassurance, the median among those who are non-employed in October is 0 weeks worked in the year.

D.2 Parental transfers

For those who are enrolled in college, we impute (expected) parental transfers separately by college sector. Our specification is in the spirit of Johnson (2013). Parental transfers have two components: (*i*) a probability of receiving any transfer at all; and (*ii*) a transfer amount conditional on receiving a transfer at all. Following Johnson, we assume that parental transfers are taken as given, but that actual transfer amounts depend on the individual's school enrollment decision (i.e. 2-year college is less expensive than 4-year college, so the individual would expect a lesser amount of parental transfers if enrolling in the former).

D.2.1 Logit model of receiving any transfers at all

We use a logit model to compute the probability of receiving a parental transfer while enrolled in college, where we estimate the parameters separately by sector of college (2-year vs. 4-year):

$$\Pr\left(PT > 0 \,|\, X\right) = \frac{\exp\left(X\tilde{\beta}_{PT}\right)}{1+\exp\left(X\tilde{\beta}_{PT}\right)} \tag{D.2}$$

We include the following in X:

- age
- log family income
- race/ethnicity dummies

D.2.2 Log transfers conditional on receiving any

We estimate log transfers conditional on receiving positive transfers using a log-linear regression model (again, separately estimating the parameters by sector of college). We include the following as covariates in the regression:

- age
- log family income
- years of college completed
- race/ethnicity dummies

D.2.3 Expected parental transfers

Using the two components of the transfer function, we then compute expected parental transfers as follows:

$$\begin{aligned} E\left(PT \mid X\right) &= \Pr\left(PT > 0 \mid X\right) \times E\left[\exp\left(\log\left(PT\right)\right) \mid X, PT > 0\right] \\ &= \frac{\exp\left(X\hat{\hat{\beta}}_{pr}\right)}{1+\exp\left(X\hat{\hat{\beta}}_{pr}\right)} \times \exp\left(X\hat{\hat{\beta}}_{ln} + \frac{\hat{\hat{\sigma}}_{ln}^2}{2}\right) \end{aligned} \tag{D.3}$$

where $\hat{\hat{\beta}}_{pr}$ are the estimated logit parameters, and $\hat{\hat{\beta}}_{ln}$ and $\hat{\hat{\sigma}}_{ln}$ are the estimated OLS parameters ($\hat{\hat{\sigma}}_{ln}$ is the root mean squared error of the regression model).

D.3 Tuition, Grants, Loans, and EFC

For those who are enrolled in college, we impute the out-of-pocket cost of attending college. In order to capture realistic gradients in need- and merit-based grants, we make use of the following three variables: Expected Family Contribution (EFC), parental income, and SAT math and verbal scores.

Data on grants and loans are included in the NLSY97, but are missing at sufficiently high rates as to render them unusable for our analysis. Moreover, it is difficult to separate need-based from merit-based grants. To combat these limitations, we use an outside survey (the National Center for Education Statistics 2008 National Postsecondary Student Aid Survey or NPSAS) to compute the mapping between loans (or grants) and EFC, family income, and SAT score. It is crucial to capture both family income and SAT score, as students with lower income will tend to receive more generous grants, as also will students with greater academic preparation (measured by SAT or ACT-equivalent score), through academic scholarships.

D.3.1 Tuition

We impute tuition as the average annual tuition among men in the 2008 NPSAS. For four-year colleges, we look only at public institutions.[D1] This value is $6,394. For two-year colleges, the value is $1,380.

[D1] Many students at elite private universities receive substantial grants to offset the higher tuition sticker price. We discuss how we handle this in the ensuing paragraphs.

D.3.2 Grants

We impute grants using data in the 2008 NPSAS. We follow a similar procedure as with parental transfers: we first impute the probability of having positive levels of loans or grants, and then impute the amount conditional on having any. The only difference is that the NPSAS interface we use does not allow us to specify log grants as the dependent variable.[D2]

We then obtain

$$\begin{aligned} E\left(GR \mid X\right) &= \Pr\left(GR > 0 \mid X\right) \times E\left[GR \mid X, GR > 0\right] \\ &= \frac{\exp\left(X\hat{\hat{\beta}}_{pr,GR}\right)}{1 + \exp\left(X\hat{\hat{\beta}}_{pr,GR}\right)} \times \left(X\hat{\hat{\beta}}_{GR}\right) \end{aligned} \tag{D.4}$$

We estimate these models separately by college sector (2-year vs. 4-year). Because there is little merit-based aid in 2-year colleges, and because 2-year colleges do not require prospective students to take the SAT or ACT, we exclude entrance exam scores from the 2-year college specification.

As right-hand-side variables, we include deciles of the EFC distribution, deciles of the family income distribution, and deciles of the SAT (or ACT-equivalent) math and verbal distributions.

We include as additional controls deciles of the tuition level of the university that the NPSAS respondent is attending. This nets out unobservable college quality (which we do not model) from the imputed levels of grants.

D.3.3 Loans

To impute loans, we also use data from the 2008 NPSAS. However, we model loans differently than grants because loans eventually need to be paid back. The main difference with loans is that we treat loans as deterministic. That is, an individual has an expectation regarding the amount of loans he will be required to take up if he decides to attend 2-year or 4-year college. This expected loan amount has no uncertain component with regards to consumption. We obtain expected loans in a similar way as we impute grants:

[D2] The NPSAS data is confidential, but the NCES allows researchers to access it through its PowerStats interface. NCES PowerStats has limited flexibility in terms of regression specifications, but it has enough flexibility to serve our purposes.

$$
\begin{aligned}
E(L \mid X) &= \Pr(L > 0 \mid X) \times E[L \mid X, L > 0] \\
&= \frac{\exp\left(X\hat{\hat{\beta}}_{pr,L}\right)}{1+\exp\left(X\hat{\hat{\beta}}_{pr,L}\right)} \times \left(X\hat{\hat{\beta}}_{L}\right)
\end{aligned}
\tag{D.5}
$$

As with grants, we estimate these models separately by college sector (2-year vs. 4-year). The only difference is that we estimate loans on the age-18 subsample of the NPSAS, whereas with grants, we include all enrolled students. The right-hand side variables are identical to the model specification for grants. See Appendix G for more details on how we treat loans in our dynamic model.

D.3.4 EFC

The EFC is an important input to our method of imputing grants and loans. Here, we describe our procedure for computing the EFC using observable components of the NLSY97, along with calibration of other components that are not observed in the NLS97.

The EFC takes as inputs information on parental taxable income and assets and student taxable income and assets.[D3] The goal of the EFC is to summarize a student's eligibility for need-based financial aid. Students with lower EFC values may be eligible for more generous financial aid. The level of generosity is specific to the institution.

The EFC is a highly non-linear function of assets and income for both parents and students, as well as characteristics of the family. We impute each individual's EFC using a simplified calculator available at http://www.collegegold.com/calculatecost/efcworksheets. We briefly cover below how each component contributes to the EFC.

$$
EFC = f\left(A^P, I^P, X, A^C, I^C\right) \tag{D.6}
$$

where A denotes assets, I denotes income, and P and C superscripts denote parent and child, respectively. X represents characteristics of the parents or family. These include the parent's age and marital status, the size of the family, and the number of children currently in college.

We now outline the contributions for each component of the EFC. We denote by $\sim$

[D3] The following commonly held assets are not considered in the EFC: equity of primary residence, retirement funds, and life insurance.

the EFC contribution from each component, as opposed to the amount of each component (which is used in (D.6)). The contributions are also, of course, functions of the amounts themselves. We suppress the amounts to conserve notation.

$$\begin{aligned}
\widetilde{A}^P(X) &= 0.12\left\{A^P - D^{A,P}(X) - 1732\left[\left(\frac{\text{age}-23}{1[\text{single}]\,2.3 + 1[\text{married}]}\right)\right]\right\} \\
\widetilde{I}^P(X) &= I^P - D^{I,P}(X) \\
\widetilde{A}^C(X) &= 0.35\left\{A^C - D^{A,C}(X)\right\} \\
\widetilde{I}^C(X) &= 0.5\left(I^C - D^{I,C}(X)\right)
\end{aligned} \tag{D.7}$$

where $1[\cdot]$ is the indicator function, and where

$$\begin{aligned}
D^{A,P}(X) &= \min\left\{\$250k, 0.5A^{B,P}\right\}1\left[A^{B,P} > 0\right] \\
D^{I,P}(X) &= FICA^P + Tax^{F,P} + 0.06I^P + \$10k + (H\$3.46k - N\$2.46k) + \\
&\qquad \min\left\{\$3.1k, 0.35I^P\right\} \\
D^{A,C}(X) &= \min\left\{\$250k, 0.5A^{B,C}\right\}1\left[A^{B,C} > 0\right] \\
D^{I,C}(X) &= FICA^C + Tax^{F,C} + 0.03I^C + \$2.55k
\end{aligned} \tag{D.8}$$

where $A^{B,P}$ denotes business/farm assets of the parent and $A^{B,C}$ for the child; $FICA$ denotes federal payroll tax (for either parent or child); Tax^F denotes federal income tax paid (for either parent or child); H denotes number of people in household; N denotes number of children in college.

Combining (D.7) and (D.8), we have

$$EFC = \begin{cases} 0 & \text{if } I^P \leq \$20k \\ \widetilde{I}^P(X) + \widetilde{I}^C(X) & \text{if } I^P \in (\$20k, \$50k) \\ \widetilde{A}^P(X) + \widetilde{I}^P(X) + \widetilde{A}^C(X) + \widetilde{I}^C(X) & \text{if } I^P \geq \$50k \end{cases} \tag{D.9}$$

D.3.5 EFC components observed in the NLSY97

We use the following NLSY97 variables as inputs into the EFC function:

- Household size in 1997
- Parent age

- Parent marital status
- Parent income at age 17
- Parent net worth at age 17
  - We use the 2004 Survey of Income and Program Participation (SIPP) to estimate the function mapping household net worth to the subset of household assets that enter into the EFC formula. We then use this function to impute parental EFC assets as a function of parental net worth.[D4]
- Child income in every year of survey
  - We assume that past income has no stochastic element, and that it does not vary except by previous work status. This assumption is innocuous, and is invoked in order to allow us to preserve the properties of our dynamic model for ease of estimation. In practice, EFC varies little with income, holding fixed work intensity and all other EFC inputs.

D.3.6 EFC components not observed in the NLSY97

While the NLSY97 collects detailed data, there are a number of EFC inputs that are not collected. Here we briefly discuss how we handle these.

- Income taxes paid. For this, we impute income tax using the average tax rates reported in Guner, Kaygusuz, and Ventura (2014)
- Number of children in college. We calibrate this number to 1, which is the median in the NLSY97
- Business assets. We calibrate these to $0.
- Child assets. We calibrate these to $0.

D.4 Flow utility of consumption

We assume that individuals have CRRA preferences over their consumption, with a risk aversion parameter set equal to θ. It follows that the flow utility of consumption is given by:

[D4] EFC assets are allowed to be a function of the following variables: a quadratic in net worth, log family income, and the full set of interactions between parental education and race/ethnicity.

$$u(C) = \frac{C^{1-\theta}}{1-\theta} \tag{D.10}$$

where the consumption level C is given by Equations 18 and 19. We distinguish three cases:

1. Not in school and not working:

$$E(U(C)) = \frac{\underline{C}^{1-\theta}}{1-\theta} \tag{D.11}$$

2. Working and not in school:

$$\begin{aligned} E(U(C)) &= \frac{E(\max^{1-\theta}(W, \underline{C}))}{1-\theta} \\ &= \frac{1}{1-\theta} \times \Big(F_w(\ln \underline{C})\underline{C}^{1-\theta} + (1 - F_w(\ln \underline{C}))E(Z|Z \leq \underline{C}^{1-\theta})\Big) \end{aligned} \tag{D.12}$$

where (denoting by h the annual hours worked) $Z = \exp((1-\theta)\ln W)$, and $F_w(\cdot)$ is the cdf. of the (normal) distribution (m_w, σ_w) of w, where $m_w = \log(h) + \lambda_0 + \lambda_1(X'\beta + E(A_w|\mathcal{I}_t))$ and $\sigma_w^2 = \sigma_\varepsilon^2 + \lambda_1^2 \text{Var}(A_w|\mathcal{I}_t)$, where λ_0 is the intercept for the productivity index for wages and λ_1 is the loading on the productivity index for wages, with (λ_0, λ_1) normalized to $(0,1)$ for out-of-school wages. Note that Z is log-normally distributed, with parameters $m_z = (1-\theta)m_w$ and $\sigma_z^2 = (1-\theta)^2\sigma_w^2$. It follows that the conditional expectation term $E(Z|Z \leq \underline{C}^{1-\theta})$ is obtained using the formula (and setting $a = \underline{C}^{1-\theta}$):

$$E(Z|Z \leq a) = \exp(m_z + \sigma_z^2/2)\frac{\Phi(\frac{\ln(a)-m_z-\sigma_z^2}{\sigma_z})}{\Phi(\frac{\ln(a)-m_z}{\sigma_z})} \tag{D.13}$$

where $\Phi(\cdot)$ denotes the cdf. of a standard normal distribution.[D5]

[D5] For obtaining the expected utility of consumption in future periods (which is required according to our finite dependence framework), we simply set $m_{w,t+1} = m_w + \phi\delta_t$ and $\sigma_{w,t+1}^2 = \sigma_w^2 + \sigma_\zeta^2$, where ϕ is the autocorrelation coefficient on the year dummies δ_t and σ_ζ^2 is the variance of the residuals of the AR(1) year dummy model.

We similarly augment the mean and variance of wages with the variances mentioned above for step (3) discussed below.

Note that the expression above applies to $\theta > 1$. If $\theta < 1$, the expression becomes:

$$\begin{aligned} E(U(C)) &= \frac{E(\max^{1-\theta}(W, \underline{C}))}{1-\theta} \\ &= \frac{1}{1-\theta} \times \left(F_w(\ln \underline{C})\underline{C}^{1-\theta} + (1 - F_w(\ln \underline{C}))E(Z|Z \geq \underline{C}^{1-\theta})\right) \end{aligned} \tag{D.14}$$

where the conditional expectation term $E(Z|Z \geq \underline{C}^{1-\theta})$ is obtained using the formula (and setting $a = \underline{C}^{1-\theta}$):

$$E(Z|Z \geq a) = \exp(m_z + \sigma_z^2/2)\frac{1 - \Phi(\frac{\ln(a) - m_z - \sigma_z^2}{\sigma_z})}{1 - \Phi(\frac{\ln(a) - m_z}{\sigma_z})} \tag{D.15}$$

If $\theta = 1$ we have logarithmic preferences and the expected utility of consumption is given by:

$$E(U(C)) = F_w(\ln \underline{C}) \ln \underline{C} + (1 - F_w(\ln \underline{C}))E(w|w \geq \ln \underline{C}) \tag{D.16}$$

where $F_w(\cdot)$ is the cdf. of the normal distribution with moments (m_w, σ_w) given above, and the truncated mean $E(w|w \geq \ln \underline{C})$ is given by:

$$E(w|w \geq \ln \underline{C}) = m_w + \sigma_w \left(\frac{\varphi(\ln \underline{C})}{1 - \Phi(\ln \underline{C})}\right) \tag{D.17}$$

3. In school:

$$E(U(C)) = \frac{E(\max^{1-\theta}(C^*, \underline{C}))}{1-\theta} \tag{D.18}$$

1. and 2. are straightforward to compute and do not require numerical integration. 3. does require numerical integration over the (log-normal) distributions of parental transfers PT, grants G, and (for school and work only) in-school wages W.

Specifically, the expected utility of consumption for the school and no-work alternative is written as follows (keeping the conditioning on the observed characteristics X implicit):

$$E(U(C)) = \frac{1}{1-\theta} \iint \max{}^{1-\theta}(C^*, \underline{C}) f_{PT}(pt) f_{GR}(gr) \mathrm{d}pt \mathrm{d}gr \tag{D.19}$$

where $f_{PT}(\cdot)$ and $f_{GR}(\cdot)$ respectively denote the pdf. of the log-normal distribution of parental transfers and the normal distribution of grants, and C^* is expressed as a function of parental transfers, grants and loans as in (18). Recall that we treat expected loans as deterministic with respect to consumption.

Similarly, the expected utility of consumption for the school and work alternatives is written as follows:

$$E(U(C)) = \frac{1}{1-\theta} \iiint \max{}^{1-\theta}(C^*, \underline{C}) f_{PT}(pt) f_{GR}(gr) f_W(W) \mathrm{d}pt \mathrm{d}gr \mathrm{d}W \tag{D.20}$$

where $f_{PT}(\cdot)$, $f_W(\cdot)$, and $f_{GR}(\cdot)$ respectively denote the pdf. of the log-normal distributions of parental transfers and in-school wages and the normal distribution of grants, and C^* is expressed as a function of parental transfers, loans, grants, and wages (see (18)).

In the following two subsubsections, we provide detailed examples of consumption for the case of not working in college, as well as for the case of working while in college.

D.4.1 Example 1: Consumption in college (and not working)

Expanding out all notation and assuming that consumption in college (no work) is solely a function of parental transfers (i.e. the individual received no grants), we would have (combining (D.3) and (D.19))

$$\begin{aligned} E(U(C)|X=x) =& \frac{1}{1-\theta} \int_{-\infty}^{\infty} \max{}^{1-\theta}(e^{pt} + l - tf, \underline{C}) f_{pt|X=x}(pt) \mathrm{d}pt \\ =& (1 - \Pr(PT > 0|X=x)) \times \frac{\underline{C}^{1-\theta}}{1-\theta} \\ &+ \Pr(PT > 0|X=x) \times \frac{1}{1-\theta} \int_{-\infty}^{\infty} \max{}^{1-\theta}(e^{pt} + l - tf, \underline{C}) \frac{1}{\sigma_{ln}} \varphi(pt - X\beta_{ln}) \mathrm{d}pt \end{aligned} \tag{D.21}$$

where $pt = \log(PT)$ refers to log-parental transfers, l refers to the individual's deterministic amount of expected loans, t refers to tuition, and $\varphi(\cdot)$ denotes the pdf. of a standard normal distribution.

The full version of (D.19) in the case where the student receives grants would then be

$$
\begin{aligned}
E(U(C)|X=x) =& \frac{1}{1-\theta}\iint_{-\infty}^{\infty} \max{}^{1-\theta}(e^{pt}+gr+l-tf,\underline{C})\frac{1}{\sigma_{ln}}\varphi(pt-X\beta_{ln})\frac{1}{\sigma_{gr}}\varphi(gr-X\beta_{gr})\mathrm{d}pt\mathrm{d}gr \\
=& \frac{1}{1-\theta}\iint_{-\infty}^{\infty} \max{}^{1-\theta}(e^{pt}+gr+l-tf,\underline{C})\mathrm{d}F(pt)\mathrm{d}F(gr) \\
=& \Pr(PT=0)\Pr(GR=0)\times\frac{\underline{C}^{1-\theta}}{1-\theta} \\
&+\Pr(PT>0)\Pr(GR=0)\times\frac{1}{1-\theta}\int_{-\infty}^{\infty}\max{}^{1-\theta}(e^{pt}+l-tf,\underline{C})\mathrm{d}F(pt) \\
&+\Pr(PT=0)\Pr(GR>0)\times\frac{1}{1-\theta}\int_{-\infty}^{\infty}\max{}^{1-\theta}(gr+l-tf,\underline{C})\mathrm{d}F(gr) \\
&+\Pr(PT>0)\Pr(GR>0)\times\frac{1}{1-\theta}\iint_{-\infty}^{\infty}\max{}^{1-\theta}(e^{pt}+gr+l-tf,\underline{C})\mathrm{d}F(pt)\mathrm{d}F(gr)
\end{aligned}
\tag{D.22}
$$

where $pt = \log(PT)$ refers to log-parental transfers, g refers to grants (in levels), l refers to deterministic loans (in levels), and $\varphi(\cdot)$ each denote the pdf. of the standard normal distribution.

D.4.2 Example 2: Consumption while working in college

Expanding out all notation and considering the case of working while in college (and denoting h the annual hours worked), we would have

$$
\begin{aligned}
E(U(C)|X=x) =& \frac{1}{1-\theta}\iiint_{-\infty}^{\infty} \max{}^{1-\theta}(e^{\log(h)+w} + e^{pt} + gr + l - tf, \underline{C})\frac{1}{\sigma_w}\varphi(w - \log(h) - \lambda_0 - \lambda_1(X\beta + A))\times \\
& \frac{1}{\sigma_{ln}}\varphi(pt - X\beta_{ln})\frac{1}{\sigma_{gr}}\varphi(gr - X\beta_{gr})\frac{1}{\sigma_l}\varphi(l - X\beta_l)\mathrm{d}w\mathrm{d}pt\mathrm{d}gr \\
=& \frac{1}{1-\theta}\iiint_{-\infty}^{\infty} \max{}^{1-\theta}(e^{\log(h)+w} + e^{pt} + gr + l - tf, \underline{C})\mathrm{d}F(w)\mathrm{d}F(pt)\mathrm{d}F(gr) \\
=& \Pr(PT=0)\Pr(GR=0)\times\frac{1}{1-\theta}\int_{-\infty}^{\infty} \max{}^{1-\theta}(e^{\log(h)+w} + l - tf, \underline{C})\mathrm{d}F(w) \\
&+ \Pr(PT>0)\Pr(GR=0)\times\frac{1}{1-\theta}\iint_{-\infty}^{\infty} \max{}^{1-\theta}(e^{\log(h)+w} + e^{pt} + l - tf, \underline{C})\mathrm{d}F(w)\mathrm{d}F(pt) \\
&+ \Pr(PT=0)\Pr(GR>0)\times\frac{1}{1-\theta}\iint_{-\infty}^{\infty} \max{}^{1-\theta}(e^{\log(h)+w} + gr + l - tf, \underline{C})\mathrm{d}F(w)\mathrm{d}F(gr) \\
&+ \Pr(PT>0)\Pr(GR>0)\times\frac{1}{1-\theta}\iiint_{-\infty}^{\infty} \max{}^{1-\theta}(e^{\log(h)+w} + e^{pt} + gr + l - tf, \underline{C})\mathrm{d}F(w)\mathrm{d}F(pt)\mathrm{d}F(gr)
\end{aligned} \tag{D.23}
$$

where the notation is as on the previous page, and where A denotes the individual's prior ability belief as of time t (i.e. $E(A|\mathcal{I}_t)$). Note that σ_w is as defined above: $\sigma_w = \sqrt{\sigma_\varepsilon^2 + \lambda_1^2 V(A_w|\mathcal{I}_t)}$ where λ_1 is the productivity index loading on in-school wages ($\lambda_1 = 1$ for out-of-school wages).

D.4.3 Computation of expected utility of consumption

In practice, we compute the expected utility of consumption in two different ways, depending on where it enters the model. First, for expected utility of consumption that enters the flow utility terms in periods t and $t+1$ (see Appendix F), we compute the integrals using Monte Carlo integration, or Gaussian quadrature if the dimension of the integral is 1 and the integral does not have a closed form as in Equation (D.15). We do this for the structural estimation as well as the backwards recursion. Second, for expected consumption that enters the conditional choice probability terms, we approximate the integral using a flexible polynomial evaluated at the relevant points of the state space along the finite dependence paths. We do this in order to limit the computational burden. This polynomial is fitted in a preliminary step from a linear regression of the actual expected utility of consumption (computed by Monte Carlo integration), on a "naive" utility of consumption (computed assuming away uncertainty about wage realizations, parental transfers, or educational grants), its square, and several interactions with and squares of parental transfers, educational grants, and expected loans. We estimate these regressions separately for 2-year and 4-year schooling options, where the two 4-year majors are pooled together. The R^2 of this regression always exceeds 0.9, and in many cases exceeds 0.98.

D.5 Risk aversion parameter θ

We set the risk aversion parameter θ equal to 0.4. We choose this parameter as follows. We first implement a grid search procedure for the static choice model (stage 4 in Table B1), using values of θ in the interval $[0, 2]$ in increments of 0.05. We choose this interval because it covers several different values from the literature, ranging from risk neutrality to the value of 2 used by Hai and Heckman (2017) and including the value of 0.48 in Keane and Wolpin (2001). This preliminary step results in a value of θ equal to 0.15. Repeating a similar grid search procedure based on the dynamic choice model, conditional on the static choice $\theta = 0.15$, yields $\theta = 0.4$. While for simplicity we set a common $\theta = 0.4$ for the static and dynamic choice models, our estimation results (available upon request) are quantitatively robust to the use of $\theta = 0.15$ and $\theta = 0.4$ in the static and dynamic choice models, respectively.

E Effort and grades

In this section we show how to reframe our learning model as a setup where students are exerting effort and receiving signals regarding their per-unit study costs. Ahn et al. (2023) show that a particular form for the returns and costs of studying yield linear expressions for grades and utility in A_{ij}. Namely, one could alternatively express our grade equation, Equation (1), as a function of study time, $s_{i\tau}$:

$$G_{ij\tau}(s_{i\tau}) = \gamma_{0j} + X_{ij\tau}\gamma_{1j} + \ln(s_{i\tau}) \tag{E.1}$$

where $A_{ij} + \varepsilon_{ij\tau}$ in the original equation has been replaced by $\ln(s_{i\tau})$, the log of study time.[E1] With the cost of studying in major j given by:[E2]

$$c_j(s_{i\tau}) = s_{i\tau}\exp(\delta_{0j} + \delta_1 A_{ij} + \varepsilon_{ij\tau}), \tag{E.2}$$

and individuals valuing grades at $\alpha_G G_{ij\tau}(s_{i\tau})$, the (interior) optimal choice of study time is given by:

$$s^*_{i\tau} = \frac{\alpha_G}{\exp(\delta_{0j} + \delta_1 A_{ij} + \varepsilon_{ij\tau})} \tag{E.3}$$

Substituting this expression into (E.1) yields:

$$G_{ij\tau}(s^*_{i\tau}) = \tilde{\gamma}_{0j} + X_{ij\tau}\gamma_{1j} + \tilde{A}_{ij} + \tilde{\varepsilon}_{ij\tau} \tag{E.4}$$

where:

$$\tilde{\gamma}_{0j} = \gamma_{0j} + \ln(\alpha_G) - \delta_{0j} \tag{E.5}$$

$$\tilde{A}_{ij} = -\delta_1 A_{ij} \tag{E.6}$$

$$\tilde{\varepsilon}_{ij\tau} = -\varepsilon_{ij\tau} \tag{E.7}$$

Individuals then learn about the permanent unobserved component of the per-unit study costs, A_{ij}, using realized grades as a noisy signal.

On the utility front, substituting in optimal study time into grades and costs of studying yields a net utility of effort $\alpha_G(G_{ij\tau}(s^*_{i\tau}) - 1)$, which also maps directly into our specification.

[E1] That there is no coefficient on $\ln(s_{i\tau})$ is innocuous.

[E2] Note that the model naturally extends to one where $X_{ij\tau}$ affects study costs as well.

F More details on finite dependence

This appendix section details the mathematical derivation of the difference in value functions $v_{jkl} - v_h$. The main goal here, which we accomplish using the finite dependence property of our model, is to obtain an expression of $v_{jkl} - v_h$ that is not recursive. For the v_{jkl} conditional value function, individuals choose home production in both $t+1$ and $t+2$. For the v_h conditional value function, individuals choose (j, k, l) in $t+1$ and home production in $t+2$. Because of the exogenous white collar job offer arrival rate $\tilde{\lambda}^{(h)}_{t+1}$, we respectively weight the acceptance and rejection probabilities in $t+1$ (in the event that i receives an offer) by $\frac{1}{\tilde{\lambda}^{(h)}_{t+1}}$ and $1 - \frac{1}{\tilde{\lambda}^{(h)}_{t+1}}$. This allows us to achieve the cancellation in $t+3$ that is required for estimation of the model.

We make three simplifying assumptions that are crucial to the validity of finite dependence: (i) the utility of rejecting a white-collar offer is same as the utility of not receiving a white-collar offer, i.e. there is no "discouragement effect" to not receiving an offer; (ii) there is no human capital depreciation; and (iii) loans evolve in a specific manner detailed in Appendix G. For expositional purposes, we simplify the states to be work experience (x) and previous decision (d_{t-1}).

It is also helpful to keep in mind that the *ex ante* value function V_t contains multiple dimensions of uncertainty: job offer arrival; aggregate labor market state; college graduation; wage and grade signals; and preference shocks.

Throughout this appendix, we make use of the conditional choice probability mapping of Hotz and Miller (1993) and Arcidiacono and Miller (2011):

$$V_t(Z_{it}) = v_{jkl}(Z_{it}) - \ln p_{jkl}(Z_{it}) + \gamma \tag{F.1}$$

for all i, (j, k, l) and t, where $p_{jkl}(Z_{it})$ denotes the CCP of choosing alternative (j, k, l) and γ is Euler's constant. This holds by virtue of our assumption that the preference shocks are distributed Type I extreme value. Note that, because we are interested in differenced conditional value functions $v_{jkl} - v_h$, c drops out of the expression and we accordingly suppress it for expositional purposes in what follows.

We abbreviate, for all (j, k, l) and t, $u_{jkl}(Z_{it})$ as u_{jklt} and $p_{jkl}(Z_{it})$ as p_{jklt} to conserve on notation. We also omit dependence of v_{jkl} on (Z_{it}) and all E_t operators. Finally, all utilities and probabilities are conditional on unobserved type r, which we also suppress.

F.1 Paths not involving four-year college

F.1.1 Work-home-home path with frictions

$$\begin{aligned}
v_{jkl} &= u_{jklt} + \beta E_t \left[V_{t+1}\left(x_t+1, d_t=(j,k,l)\right)\right] \\
&= u_{jklt} + \beta \left(\begin{array}{c} \tilde{\lambda}_{t+1}^{(jkl)} E_t \left[V_{t+1}\left(x_t+1, d_t=(j,k,l), \text{offer}_{t+1}=1\right)\right] + \\ \left(1-\tilde{\lambda}_{t+1}^{(jkl)}\right) E_t \left[V_{t+1}\left(x_t+1, d_t=(j,k,l), \text{offer}_{t+1}=0\right)\right] \end{array} \right) \\
&= u_{jklt} + \beta \left(\begin{array}{c} \tilde{\lambda}_{t+1}^{(jkl)} \left\{u_{ht+1}^{w} - \ln p_{ht+1}^{w} + \beta V_{t+2}\left(x_t+1, d_{t+1}=h\right)\right\} + \\ \left(1-\tilde{\lambda}_{t+1}^{(jkl)}\right) \left\{u_{ht+1}^{n} - \ln p_{ht+1}^{n} + \beta V_{t+2}\left(x_t+1, d_{t+1}=h\right)\right\} \end{array} \right) \\
&= u_{jklt} + \beta \left(\begin{array}{c} \tilde{\lambda}_{t+1}^{(jkl)} \left\{-\ln p_{ht+1}^{w}\right\} + \\ \left(1-\tilde{\lambda}_{t+1}^{(jkl)}\right) \left\{-\ln p_{ht+1}^{n}\right\} \end{array} \right) + \beta^2 V_{t+2}\left(x_t+1, d_{t+1}=h\right)
\end{aligned} \tag{F.2}$$

where superscript w signifies "received white collar offer" and superscript n signifies "no offer received." Because we assume that "received an offer in the previous period" is *not* a state variable, the two V_{t+2} terms can be combined. Continuing on through period $t+3$:

$$\begin{aligned}
v_{jkl} &= u_{jklt} + \beta \left(\begin{array}{c} \tilde{\lambda}_{t+1}^{(jkl)} \left\{-\ln p_{ht+1}^{w}\right\} + \\ \left(1-\tilde{\lambda}_{t+1}^{(jkl)}\right) \left\{-\ln p_{ht+1}^{n}\right\} \end{array} \right) \\
&\quad + \beta^2 \left(\begin{array}{c} \tilde{\lambda}_{t+2}^{(h)} \left\{u_{ht+2}^{w} - \ln p_{ht+2}^{w} + \beta V_{t+3}\left(x_t+1, d_{t+2}=h\right)\right\} + \\ \left(1-\tilde{\lambda}_{t+2}^{(h)}\right) \left\{u_{ht+2}^{n} - \ln p_{ht+2}^{n} + \beta V_{t+3}\left(x_t+1, d_{t+2}=h\right)\right\} \end{array} \right) \\
&= u_{jklt} + \beta \left(\begin{array}{c} \tilde{\lambda}_{t+1}^{(jkl)} \left\{-\ln p_{ht+1}^{w}\right\} + \\ \left(1-\tilde{\lambda}_{t+1}^{(jkl)}\right) \left\{-\ln p_{ht+1}^{n}\right\} \end{array} \right) + \beta^2 \left(\begin{array}{c} \tilde{\lambda}_{t+2}^{(h)} \left\{-\ln p_{ht+2}^{w}\right\} + \\ \left(1-\tilde{\lambda}_{t+2}^{(h)}\right) \left\{-\ln p_{ht+2}^{n}\right\} \end{array} \right) \\
&\quad + \beta^3 V_{t+3}\left(x_t+1, d_{t+2}=h\right)
\end{aligned} \tag{F.3}$$

F.1.2 Home-work-home path with frictions

Using the same notation as above, we get

$$\begin{aligned}
v_h &= u_{ht} + \beta E_t\left[V_{t+1}\left(x_t, d_t = h\right)\right] \\
&= u_{ht} + \beta \begin{pmatrix} \tilde{\lambda}^{(h)}_{t+1} E_t\left[V_{t+1}\left(x_t, d_t = h, \text{offer}_{t+1} = 1\right)\right] + \\ \left(1 - \tilde{\lambda}^{(h)}_{t+1}\right) E_t\left[V_{t+1}\left(x_t, d_t = h, \text{offer}_{t+1} = 0\right)\right] \end{pmatrix} \\
&= u_{ht} + \beta \begin{pmatrix} \left(\frac{1}{\tilde{\lambda}^{(h)}_{t+1}}\right) \tilde{\lambda}^{(h)}_{t+1} E_t\left[V_{t+1}\left(x_t, d_t = h, \text{offer}_{t+1} = 1, \text{accept}_{t+1} = 1\right)\right] + \\ \left(1 - \frac{1}{\tilde{\lambda}^{(h)}_{t+1}}\right) \tilde{\lambda}^{(h)}_{t+1} E_t\left[V_{t+1}\left(x_t, d_t = h, \text{offer}_{t+1} = 1, \text{accept}_{t+1} = 0\right)\right] + \\ \left(1 - \tilde{\lambda}^{(h)}_{t+1}\right) E_t\left[V_{t+1}\left(x_t, d_t = h, \text{offer}_{t+1} = 0\right)\right] \end{pmatrix} \\
&= u_{ht} + \beta \begin{pmatrix} \left(\frac{1}{\tilde{\lambda}^{(h)}_{t+1}}\right) \tilde{\lambda}^{(h)}_{t+1} \left\{u^w_{jklt+1} - \ln p^w_{jklt+1} + \beta V_{t+2}\left(x_t + 1, d_{t+1} = (j,k,l)\right)\right\} + \\ \left(1 - \frac{1}{\tilde{\lambda}^{(h)}_{t+1}}\right) \tilde{\lambda}^{(h)}_{t+1} \left\{u^w_{ht+1} - \ln p^w_{ht+1} + \beta V_{t+2}\left(x_t, d_{t+1} = h\right)\right\} + \\ \left(1 - \tilde{\lambda}^{(h)}_{t+1}\right) \left\{u^n_{ht+1} - \ln p^n_{ht+1} + \beta V_{t+2}\left(x_t, d_{t+1} = h\right)\right\} \end{pmatrix} \qquad \text{(F.4)} \\
&= u_{ht} + \beta \begin{pmatrix} \left(\frac{1}{\tilde{\lambda}^{(h)}_{t+1}}\right) \tilde{\lambda}^{(h)}_{t+1} \left\{u^w_{jklt+1} - \ln p^w_{jklt+1} + \beta V_{t+2}\left(x_t + 1, d_{t+1} = (j,k,l)\right)\right\} + \\ \left(\tilde{\lambda}^{(h)}_{t+1} - 1\right) \left\{- \ln p^w_{ht+1} + \beta V_{t+2}\left(x_t, d_{t+1} = h\right)\right\} + \\ \left(1 - \tilde{\lambda}^{(h)}_{t+1}\right) \left\{- \ln p^n_{ht+1} + \beta V_{t+2}\left(x_t, d_{t+1} = h\right)\right\} \end{pmatrix} \\
&= u_{ht} + \beta \begin{pmatrix} \left\{u^w_{jklt+1} - \ln p^w_{jklt+1}\right\} + \\ \left(1 - \tilde{\lambda}^{(h)}_{t+1}\right) \left\{- \ln p^n_{ht+1} + \ln p^w_{ht+1}\right\} \end{pmatrix} + \beta^2 V_{t+2}\left(x_t + 1, d_{t+1} = (j,k,l)\right)
\end{aligned}$$

An adequate choice of weighting the offer acceptance and offer rejection future value terms in (F.4) gives us the cancellation that we were looking for. Continuing on through period $t+3$:

$$\begin{aligned}
v_h = u_{ht} &+ \beta \begin{pmatrix} \left\{u^w_{jklt+1} - \ln p^w_{jklt+1}\right\} + \\ \left(1 - \tilde{\lambda}^{(h)}_{t+1}\right) \left\{- \ln p^n_{ht+1} + \ln p^w_{ht+1}\right\} \end{pmatrix} \\
&+ \beta^2 \begin{pmatrix} \tilde{\lambda}^{(jkl)}_{t+2} \left\{u_{ht+2} - \ln p^w_{ht+2} + \beta V_{t+3}\left(x_t + 1, d_{t+2} = h\right)\right\} + \\ \left(1 - \tilde{\lambda}^{(jkl)}_{t+2}\right) \left\{u_{ht+2} - \ln p^n_{ht+2} + \beta V_{t+3}\left(x_t + 1, d_{t+2} = h\right)\right\} \end{pmatrix} \\
= u_{ht} &+ \beta \begin{pmatrix} \left\{u^w_{jklt+1} - \ln p^w_{jklt+1}\right\} + \\ \left(1 - \tilde{\lambda}^{(h)}_{t+1}\right) \left\{- \ln p^n_{ht+1} + \ln p^w_{ht+1}\right\} \end{pmatrix} \\
&+ \beta^2 \begin{pmatrix} \tilde{\lambda}^{(jkl)}_{t+2} \left\{- \ln p^w_{ht+2}\right\} + \\ \left(1 - \tilde{\lambda}^{(jkl)}_{t+2}\right) \left\{- \ln p^n_{ht+2}\right\} \end{pmatrix} + \beta^3 V_{t+3}\left(x_t + 1, d_{t+2} = h\right)
\end{aligned} \qquad \text{(F.5)}$$

F.1.3 Putting it together

Combining (F.3) and (F.5) gives us

$$\begin{aligned} v_{jkl} - v_h = u_{jklt} &+ \beta \begin{pmatrix} \tilde{\lambda}_{t+1}^{(jkl)} \left\{ - \ln p_{ht+1}^w \right\} + \\ \left(1 - \tilde{\lambda}_{t+1}^{(jkl)}\right) \left\{ - \ln p_{ht+1}^n \right\} \end{pmatrix} + \beta^2 \begin{pmatrix} \tilde{\lambda}_{t+2}^{(h)} \left\{ - \ln p_{ht+2}^w \right\} + \\ \left(1 - \tilde{\lambda}_{t+2}^{(h)}\right) \left\{ - \ln p_{ht+2}^n \right\} \end{pmatrix} \\ &- \beta \begin{pmatrix} \left(\frac{1}{\tilde{\lambda}_{t+1}^{(h)}}\right) \tilde{\lambda}_{t+1}^{(h)} \left\{ u_{jklt+1}^w - \ln p_{jklt+1}^w \right\} + \\ \left\{ - \left(1 - \tilde{\lambda}_{t+1}^{(h)}\right) \ln p_{ht+1}^n - \left(1 - \frac{1}{\tilde{\lambda}_{t+1}^{(h)}}\right) \tilde{\lambda}_{t+1}^{(h)} \ln p_{ht+1}^w \right\} \end{pmatrix} \\ &- \beta^2 \begin{pmatrix} \tilde{\lambda}_{t+2}^{(jkl)} \left\{ - \ln p_{ht+2}^w \right\} + \\ \left(1 - \tilde{\lambda}_{t+2}^{(jkl)}\right) \left\{ - \ln p_{ht+2}^n \right\} \end{pmatrix} \end{aligned} \tag{F.6}$$

F.2 Paths involving four-year college

Because graduation from four-year college is stochastic in our model, finite dependence paths involving these options contain additional terms that include graduation probabilities.

F.2.1 College-home-home path with frictions and stochastic graduation

Now we introduce a probability of graduation that four-year college students must forecast over. Let ρ_t be defined as the probability of having graduated with a bachelor's degree *before* period t, and let g_t be a variable equal to 1 if one holds a bachelor's degree at the beginning of period t and 0 otherwise.

$$
\begin{aligned}
v_{jkl} &= u_{jklt} + \beta\rho_{t+1}E_t\left[V_{t+1}\left(x_t+1, d_t=(j,k,l), g_{t+1}=1\right)\right] + \\
&\qquad \beta\left(1-\rho_{t+1}\right)E_t\left[V_{t+1}\left(x_t+1, d_t=(j,k,l), g_{t+1}=0\right)\right] \\
&= u_{jklt} + \beta\rho_{t+1}\left(\begin{array}{c} \tilde{\lambda}_{t+1}^{(jkl,g)}E_t\left[V_{t+1}\left(x_t+1, d_t=(j,k,l), g_{t+1}=1, \text{offer}_{t+1}=1\right)\right] + \\ \left(1-\tilde{\lambda}_{t+1}^{(jkl,g)}\right)E_t\left[V_{t+1}\left(x_t+1, d_t=(j,k,l), g_{t+1}=1, \text{offer}_{t+1}=0\right)\right] \end{array}\right) + \\
&\qquad \beta\left(1-\rho_{t+1}\right)\left(\begin{array}{c} \tilde{\lambda}_{t+1}^{(jkl,ng)}E_t\left[V_{t+1}\left(x_t+1, d_t=(j,k,l), g_{t+1}=0, \text{offer}_{t+1}=1\right)\right] + \\ \left(1-\tilde{\lambda}_{t+1}^{(jkl,ng)}\right)E_t\left[V_{t+1}\left(x_t+1, d_t=(j,k,l), g_{t+1}=0, \text{offer}_{t+1}=0\right)\right] \end{array}\right) \\
&= u_{jklt} + \beta\rho_{t+1}\left(\begin{array}{c} \tilde{\lambda}_{t+1}^{(jkl,g)}\left\{u_{ht+1}^{w,g} - \ln p_{ht+1}^{w,g} + \beta V_{t+2}\left(x_t+1, d_{t+1}=h, g_{t+1}=1\right)\right\} + \\ \left(1-\tilde{\lambda}_{t+1}^{(jkl,g)}\right)\left\{u_{ht+1}^{n,g} - \ln p_{ht+1}^{n,g} + \beta V_{t+2}\left(x_t+1, d_{t+1}=h, g_{t+1}=1\right)\right\} \end{array}\right) + \\
&\qquad \beta\left(1-\rho_{t+1}\right)\left(\begin{array}{c} \tilde{\lambda}_{t+1}^{(jkl,ng)}\left\{u_{ht+1}^{w,ng} - \ln p_{ht+1}^{w,ng} + \beta V_{t+2}\left(x_t+1, d_{t+1}=h, g_{t+1}=0\right)\right\} + \\ \left(1-\tilde{\lambda}_{t+1}^{(jkl,ng)}\right)\left\{u_{ht+1}^{n,ng} - \ln p_{ht+1}^{n,ng} + \beta V_{t+2}\left(x_t+1, d_{t+1}=h, g_{t+1}=0\right)\right\} \end{array}\right) \\
&= u_{jklt} + \beta\rho_{t+1}\left(\begin{array}{c} \tilde{\lambda}_{t+1}^{(jkl,g)}\left\{-\ln p_{ht+1}^{w,g}\right\} + \\ \left(1-\tilde{\lambda}_{t+1}^{(jkl,g)}\right)\left\{-\ln p_{ht+1}^{n,g}\right\} \end{array}\right) + \beta\left(1-\rho_{t+1}\right)\left(\begin{array}{c} \tilde{\lambda}_{t+1}^{(jkl,ng)}\left\{-\ln p_{ht+1}^{w,ng}\right\} + \\ \left(1-\tilde{\lambda}_{t+1}^{(jkl,ng)}\right)\left\{-\ln p_{ht+1}^{n,ng}\right\} \end{array}\right) + \\
&\qquad \beta^2\rho_{t+1}V_{t+2}\left(x_t+1, d_{t+1}=h, g_{t+1}=1\right) + \\
&\qquad \beta^2\left(1-\rho_{t+1}\right)V_{t+2}\left(x_t+1, d_{t+1}=h, g_{t+1}=0\right)
\end{aligned} \tag{F.7}
$$

where superscript w signifies "received offer," superscript n signifies "no offer received," superscript ng signifies "not graduated," and superscript g signifies "graduated." Because we assume that "received an offer in the previous period" is *not* a state variable, the two sets of V_{t+2} terms can be combined. Continuing on through period $t+3$:

$$
\begin{aligned}
v_{jkl} = u_{jklt} + \beta \rho_{t+1} & \begin{pmatrix} \tilde{\lambda}_{t+1}^{(jkl,g)} \left\{ - \ln p_{ht+1}^{w,g} \right\} + \\ \left(1 - \tilde{\lambda}_{t+1}^{(jkl,g)}\right) \left\{ - \ln p_{ht+1}^{n,g} \right\} \end{pmatrix} + \beta \left(1 - \rho_{t+1}\right) \begin{pmatrix} \tilde{\lambda}_{t+1}^{(jkl,ng)} \left\{ - \ln p_{ht+1}^{w,ng} \right\} + \\ \left(1 - \tilde{\lambda}_{t+1}^{(jkl,ng)}\right) \left\{ - \ln p_{ht+1}^{n,ng} \right\} \end{pmatrix} + \\
& \beta^2 \rho_{t+1} \begin{pmatrix} \tilde{\lambda}_{t+2}^{(h,g)} \left\{ u_{ht+2}^{w,g} - \ln p_{ht+2}^{w,g} + \beta V_{t+3} \left(x_t + 1, d_{t+2} = h, g_{t+1} = 1\right) \right\} + \\ \left(1 - \tilde{\lambda}_{t+2}^{(h,g)}\right) \left\{ u_{ht+2}^{n,g} - \ln p_{ht+2}^{n,g} + \beta V_{t+3} \left(x_t + 1, d_{t+2} = h, g_{t+1} = 1\right) \right\} \end{pmatrix} + \\
& \beta^2 \left(1 - \rho_{t+1}\right) \begin{pmatrix} \tilde{\lambda}_{t+2}^{(h,ng)} \left\{ u_{ht+2}^{w,ng} - \ln p_{ht+2}^{w,ng} + \beta V_{t+3} \left(x_t + 1, d_{t+2} = h, g_{t+1} = 0\right) \right\} + \\ \left(1 - \tilde{\lambda}_{t+2}^{(h,ng)}\right) \left\{ u_{ht+2}^{n,ng} - \ln p_{ht+2}^{n,ng} + \beta V_{t+3} \left(x_t + 1, d_{t+2} = h, g_{t+1} = 0\right) \right\} \end{pmatrix} \\
= u_{jklt} + \beta \rho_{t+1} & \begin{pmatrix} \tilde{\lambda}_{t+1}^{(jkl,g)} \left\{ - \ln p_{ht+1}^{w,g} \right\} + \\ \left(1 - \tilde{\lambda}_{t+1}^{(jkl,g)}\right) \left\{ - \ln p_{ht+1}^{n,g} \right\} \end{pmatrix} + \beta \left(1 - \rho_{t+1}\right) \begin{pmatrix} \tilde{\lambda}_{t+1}^{(jkl,ng)} \left\{ - \ln p_{ht+1}^{w,ng} \right\} + \\ \left(1 - \tilde{\lambda}_{t+1}^{(jkl,ng)}\right) \left\{ - \ln p_{ht+1}^{n,ng} \right\} \end{pmatrix} + \\
& \beta^2 \rho_{t+1} \begin{pmatrix} \tilde{\lambda}_{t+2}^{(h,g)} \left\{ - \ln p_{ht+2}^{w,g} \right\} + \\ \left(1 - \tilde{\lambda}_{t+2}^{(h,g)}\right) \left\{ - \ln p_{ht+2}^{n,g} \right\} \end{pmatrix} + \beta^2 \left(1 - \rho_{t+1}\right) \begin{pmatrix} \tilde{\lambda}_{t+2}^{(h,ng)} \left\{ - \ln p_{ht+2}^{w,ng} \right\} + \\ \left(1 - \tilde{\lambda}_{t+2}^{(h,ng)}\right) \left\{ - \ln p_{ht+2}^{n,ng} \right\} \end{pmatrix} + \\
& \beta^3 \rho_{t+1} V_{t+3} \left(x_t + 1, d_{t+2} = h, g_{t+1} = 1\right) + \\
& \beta^3 \left(1 - \rho_{t+1}\right) V_{t+3} \left(x_t + 1, d_{t+2} = h, g_{t+1} = 0\right)
\end{aligned}
\tag{F.8}
$$

F.2.2 Home-college-home path with frictions and stochastic graduation

Using the same notation as above, we get

$$
\begin{aligned}
v_h &= u_{ht} + \beta E_t \left[V_{t+1}\left(x_t, d_t = h\right)\right] \\
&= u_{ht} + \beta \begin{pmatrix} \tilde{\lambda}_{t+1}^{(h,ng)} E_t \left[V_{t+1}\left(x_t, d_t = h, \text{offer}_{t+1} = 1\right)\right] + \\ \left(1 - \tilde{\lambda}_{t+1}^{(h,ng)}\right) E_t \left[V_{t+1}\left(x_t, d_t = h, \text{offer}_{t+1} = 0\right)\right] \end{pmatrix} \\
&= u_{ht} + \beta \begin{pmatrix} \left(\frac{1}{\tilde{\lambda}_{t+1}^{(h,ng)}}\right) \tilde{\lambda}_{t+1}^{(h,ng)} E_t \left[V_{t+1}\left(x_t, d_t = h, \text{offer}_{t+1} = 1, \text{accept}_{t+1} = 1\right)\right] + \\ \left(1 - \frac{1}{\tilde{\lambda}_{t+1}^{(h,ng)}}\right) \tilde{\lambda}_{t+1}^{(h,ng)} E_t \left[V_{t+1}\left(x_t, d_t = h, \text{offer}_{t+1} = 1, \text{accept}_{t+1} = 0\right)\right] + \\ \left(1 - \tilde{\lambda}_{t+1}^{(h,ng)}\right) E_t \left[V_{t+1}\left(x_t, d_t = h, \text{offer}_{t+1} = 0\right)\right] \end{pmatrix} \\
&= u_{ht} + \beta \begin{pmatrix} \left(\frac{1}{\tilde{\lambda}_{t+1}^{(h,ng)}}\right) \tilde{\lambda}_{t+1}^{(h,ng)} \left\{u_{jklt+1}^{w} - \ln p_{jklt+1}^{w} + \beta V_{t+2}\left(x_t + 1, d_{t+1} = (j,k,l)\right)\right\} + \\ \left(1 - \frac{1}{\tilde{\lambda}_{t+1}^{(h,ng)}}\right) \tilde{\lambda}_{t+1}^{(h,ng)} \left\{u_{ht+1}^{w} - \ln p_{ht+1}^{w} + \beta V_{t+2}\left(x_t, d_{t+1} = h\right)\right\} + \\ \left(1 - \tilde{\lambda}_{t+1}^{(h,ng)}\right) \left\{u_{ht+1}^{n} - \ln p_{ht+1}^{n} + \beta V_{t+2}\left(x_t, d_{t+1} = h\right)\right\} \end{pmatrix} \\
&= u_{ht} + \beta \begin{pmatrix} \left(\frac{1}{\tilde{\lambda}_{t+1}^{(h,ng)}}\right) \tilde{\lambda}_{t+1}^{(h,ng)} \left\{u_{jklt+1}^{w} - \ln p_{jklt+1}^{w} + \beta V_{t+2}\left(x_t + 1, d_{t+1} = (j,k,l)\right)\right\} + \\ \left(\tilde{\lambda}_{t+1}^{(h,ng)} - 1\right) \left\{- \ln p_{ht+1}^{w} + \beta V_{t+2}\left(x_t, d_{t+1} = h\right)\right\} + \\ \left(1 - \tilde{\lambda}_{t+1}^{(h,ng)}\right) \left\{- \ln p_{ht+1}^{n} + \beta V_{t+2}\left(x_t, d_{t+1} = h\right)\right\} \end{pmatrix} \\
&= u_{ht} + \beta \begin{pmatrix} \left\{u_{jklt+1}^{w} - \ln p_{jklt+1}^{w}\right\} + \\ \left(1 - \tilde{\lambda}_{t+1}^{(h,ng)}\right) \left\{- \ln p_{ht+1}^{n} + \ln p_{ht+1}^{w}\right\} \end{pmatrix} + \\
&\qquad \beta^2 \rho_{t+2} V_{t+2}\left(x_t + 1, d_{t+1} = (j,k,l), g_{t+2} = 1\right) + \\
&\qquad \beta^2 \left(1 - \rho_{t+2}\right) V_{t+2}\left(x_t + 1, d_{t+1} = (j,k,l), g_{t+2} = 0\right)
\end{aligned}
\tag{F.9}
$$

As before, a clever choice of weighting the offer acceptance and offer rejection future value terms in (F.9) gives us the cancellation that we are looking for. Continuing on through period $t + 3$:

$$
\begin{aligned}
v_h = u_{ht} + \beta & \left(\begin{array}{c} \left\{ u^{w}_{jklt+1} - \ln p^{w}_{jklt+1} \right\} + \\ \left(1 - \tilde{\lambda}^{(h,ng)}_{t+1}\right) \left\{ -\ln p^{n}_{ht+1} + \ln p^{w}_{ht+1} \right\} \end{array} \right) + \\
& \beta^2 \rho_{t+2} \left(\begin{array}{c} \tilde{\lambda}^{(jkl,g)}_{t+2} \left\{ u^{w,g}_{ht+2} - \ln p^{w,g}_{ht+2} + \beta V_{t+3}\left(x_t+1, d_{t+2}=h, g_{t+2}=1\right) \right\} + \\ \left(1 - \tilde{\lambda}^{(jkl,g)}_{t+2}\right) \left\{ u^{n,g}_{ht+2} - \ln p^{n,g}_{ht+2} + \beta V_{t+3}\left(x_t+1, d_{t+2}=h, g_{t+2}=1\right) \right\} \end{array} \right) + \\
& \beta^2 \left(1-\rho_{t+2}\right) \left(\begin{array}{c} \tilde{\lambda}^{(jkl,ng)}_{t+2} \left\{ u^{w,ng}_{ht+2} - \ln p^{w,ng}_{ht+2} + \beta V_{t+3}\left(x_t+1, d_{t+2}=h, g_{t+2}=0\right) \right\} + \\ \left(1 - \tilde{\lambda}^{(jkl,ng)}_{t+2}\right) \left\{ u^{n,ng}_{ht+2} - \ln p^{n,ng}_{ht+2} + \beta V_{t+3}\left(x_t+1, d_{t+2}=h, g_{t+2}=0\right) \right\} \end{array} \right) \\
= u_{ht} + \beta & \left(\begin{array}{c} \left\{ u^{w}_{jklt+1} - \ln p^{w}_{jklt+1} \right\} + \\ \left(1 - \tilde{\lambda}^{(h,ng)}_{t+1}\right) \left\{ -\ln p^{n}_{ht+1} + \ln p^{w}_{ht+1} \right\} \end{array} \right) + \\
& \beta^2 \rho_{t+2} \left(\begin{array}{c} \tilde{\lambda}^{(jkl,g)}_{t+2} \left\{ -\ln p^{w,g}_{ht+2} \right\} + \\ \left(1 - \tilde{\lambda}^{(jkl,g)}_{t+2}\right) \left\{ -\ln p^{n,g}_{ht+2} \right\} \end{array} \right) + \beta^2 \left(1-\rho_{t+2}\right) \left(\begin{array}{c} \tilde{\lambda}^{(jkl,ng)}_{t+2} \left\{ -\ln p^{w,ng}_{ht+2} \right\} + \\ \left(1 - \tilde{\lambda}^{(jkl,ng)}_{t+2}\right) \left\{ -\ln p^{n,ng}_{ht+2} \right\} \end{array} \right) + \\
& \beta^3 \rho_{t+2} V_{t+3}\left(x_t+1, d_{t+2}=h, g_{t+2}=1\right) + \\
& \beta^3 \left(1-\rho_{t+2}\right) V_{t+3}\left(x_t+1, d_{t+2}=h, g_{t+2}=0\right)
\end{aligned}
\tag{F.10}
$$

F.2.3 Putting it together when there is stochastic graduation in the model

Equations (F.8) and (F.10) can be combined under the following assumptions:

1. $\rho_{t+1} = \rho_{t+2}$; that is, there is no age- or calendar-time component to the college graduation probability

2. $V_{t+3}\left(\cdot,\cdot,g_{t+2}=g\right) \equiv V_{t+3}\left(\cdot,\cdot,g_{t+1}=g\right)$ for $g \in \{0,1\}$; i.e., utility does not depend on *how long* one has been a college graduate.

Both of these assumptions hold according to our model, wherein the college graduation process does not include age or calendar time, and wherein the flow utility does not depend on duration of life as a college graduate—it depends only on current college graduation status.

Combining the formulas in (F.8) and (F.10), setting $\rho_{t+2} = \rho_{t+1}$, and simplifying gives us:

$$
\begin{aligned}
v_{jkl} - v_h = u_{jklt} &+ \beta\rho_{t+1} \begin{pmatrix} \tilde{\lambda}_{t+1}^{(jkl,g)} \left\{ -\ln p_{ht+1}^{w,g} \right\} + \\ \left(1 - \tilde{\lambda}_{t+1}^{(jkl,g)}\right) \left\{ -\ln p_{ht+1}^{n,g} \right\} \end{pmatrix} + \beta \left(1 - \rho_{t+1}\right) \begin{pmatrix} \tilde{\lambda}_{t+1}^{(jkl,ng)} \left\{ -\ln p_{ht+1}^{w,ng} \right\} + \\ \left(1 - \tilde{\lambda}_{t+1}^{(jkl,ng)}\right) \left\{ -\ln p_{ht+1}^{n,ng} \right\} \end{pmatrix} + \\
&\beta^2\rho_{t+1} \begin{pmatrix} \tilde{\lambda}_{t+2}^{(h,g)} \left\{ -\ln p_{ht+2}^{w,g} \right\} + \\ \left(1 - \tilde{\lambda}_{t+2}^{(h,g)}\right) \left\{ -\ln p_{ht+2}^{n,g} \right\} \end{pmatrix} + \beta^2 \left(1 - \rho_{t+1}\right) \begin{pmatrix} \tilde{\lambda}_{t+2}^{(h,ng)} \left\{ -\ln p_{ht+2}^{w,ng} \right\} + \\ \left(1 - \tilde{\lambda}_{t+2}^{(h,ng)}\right) \left\{ -\ln p_{ht+2}^{n,ng} \right\} \end{pmatrix} - \\
&\beta \begin{pmatrix} \left(\frac{1}{\tilde{\lambda}_{t+1}^{(h,ng)}} \right) \tilde{\lambda}_{t+1}^{(h,ng)} \left\{ u_{jklt+1}^{w} - \ln p_{jklt+1}^{w} \right\} + \\ \left\{ - \left(1 - \tilde{\lambda}_{t+1}^{(h,ng)}\right) \ln p_{ht+1}^{n} - \left(1 - \frac{1}{\tilde{\lambda}_{t+1}^{(h,ng)}}\right) \tilde{\lambda}_{t+1}^{(h,ng)} \ln p_{ht+1}^{w} \right\} \end{pmatrix} - \\
&\beta^2\rho_{t+1} \begin{pmatrix} \tilde{\lambda}_{t+2}^{(jkl,g)} \left\{ -\ln p_{ht+2}^{w,g} \right\} + \\ \left(1 - \tilde{\lambda}_{t+2}^{(jkl,g)}\right) \left\{ -\ln p_{ht+2}^{n,g} \right\} \end{pmatrix} - \beta^2 \left(1 - \rho_{t+1}\right) \begin{pmatrix} \tilde{\lambda}_{t+2}^{(jkl,ng)} \left\{ -\ln p_{ht+2}^{w,ng} \right\} + \\ \left(1 - \tilde{\lambda}_{t+2}^{(jkl,ng)}\right) \left\{ -\ln p_{ht+2}^{n,ng} \right\} \end{pmatrix}
\end{aligned}
\tag{F.11}
$$

G Details on debt accumulation

This appendix section details our treatment of accumulated debt in the model. As mentioned in the body of the paper, we allow consumption to depend on loans borrowed during college. As individuals will eventually need to repay these loans, accumulated debt is an important state variable in our analysis.

G.1 Accumulated debt

We compute accumulated debt by relying on loan data from the NPSAS survey. As discussed in Appendix D, we use the subsample of 2008 NPSAS respondents who are 18 years old to compute expected loans at age 18 as a function of individual characteristics (EFC and family income for 2-year colleges; EFC, family income, SAT math and SAT verbal for 4-year colleges).

We then assume that debt accumulates according to the number of periods of school enrollment, where each period adds the individual-specific deterministic expected loan amount, compounded by an interest rate ι which we calibrate. For reasons we discuss in the next subsection, we assume that individuals take out larger loans the older they are. Specifically, we assume that the expected loan amount for a person who has been out of high school for t years is equal to $(1+\iota)^t$ times the deterministic loan amount calibrated from 18-year-olds in the NPSAS, where $t = 0$ corresponds to the first year after high school graduation.

In our logit model that we use to compute the CCPs, we allow accumulated debt to be an individual-alternative-specific covariate. Mathematically, it is defined as follows:

$$debt_{ijt} = \begin{cases} \left(exper_{i,2}E(loan_{i,2}) + exper_{i,4}E(loan_{i,4})\right)(1+\iota)^t + E(loan_{i,2})(1+\iota)^t & \text{if } j \in 2yr \\ \left(exper_{i,2}E(loan_{i,2}) + exper_{i,4}E(loan_{i,4})\right)(1+\iota)^t + E(loan_{i,4})(1+\iota)^t & \text{if } j \in 4yr \\ \left(exper_{i,2}E(loan_{i,2}) + exper_{i,4}E(loan_{i,4})\right)(1+\iota)^t & \text{if } j \notin college \end{cases} \tag{G.1}$$

where $E(loan_{i,c})$ indicates the expected loan amount of individual i when choosing college sector $c \in \{2, 4\}$. $exper_{i,c}$ indicates the number of periods in the past that the individual enrolled in college sector c, and t (minus 1) is the number of years since high school graduation; i.e. $t = 0$ for someone in their first year after high school graduation.

Table G1 is an illustrative example of how loans and debt evolve over the lifetime of a fictitious individual in our sample.

Table G1: Example of loans and debt for a fictitious individual

t	age_t	d_t	$E(loan_{i,t})$	$debt_{it}$
0	18	4yr	$E(loan_{i,4})$	$E(loan_{i,4})$
1	19	2yr	$E(loan_{i,2})(1+\iota)$	$E(loan_{i,4})(1+\iota) + E(loan_{i,2})(1+\iota)$
2	20	Home	0	$E(loan_{i,4})(1+\iota)^2 + E(loan_{i,2})(1+\iota)^2$
3	21	2yr	$E(loan_{i,2})(1+\iota)^3$	$E(loan_{i,4})(1+\iota)^3 + 2E(loan_{i,2})(1+\iota)^3$
4	22	4yr	$E(loan_{i,4})(1+\iota)^4$	$2E(loan_{i,4})(1+\iota)^4 + 2E(loan_{i,2})(1+\iota)^4$

G.2 Finite dependence with debt

With debt as a state variable, satisfying the finite dependence assumption requires some more assumptions. As mentioned in the previous subsection, the main assumption we make is that the expected loan amount in $t+1$ for a person who was at home in t is higher by $(1+\iota)$ than the expected loan amount in t for a person who chose to attend college.

Mathematically, we have the following two finite dependence paths, assuming some amount of accumulated debt in period $t-1$, $debt_{it-1}$. Note that this quantity is individual-specific, not individual-alternative-specific. That is, $debt_{it-1} = debt_{ijt-1}, j \notin college$.

G.2.1 Home-School-Home Path

- In period t the individual chooses Home, so $debt_{it} = debt_{it-1}(1+\iota)$, where we have imposed that $j = h$ in t according to the finite dependence path.
- In period $t+1$ the individual attends college and, according to our assumption, borrows a loan amount equal to $E(loan_{i,c})(1+\iota)^{t+1}$, so

  $debt_{it+1} = debt_{it}(1+\iota) + E(loan_{i,c})(1+\iota)^{t+1}$.
- In period $t+2$ the individual chooses home and has $debt_{it+2} = debt_{it+1}(1+\iota)$.

G.2.2 School-Home-Home Path

- In period t the individual chooses school, so $debt_{it} = debt_{it-1}(1+\iota) + E(loan_{i,c})(1+\iota)^t$, where we have imposed that $j = c$ in t according to the finite dependence path.
- In period $t+1$ the individual chooses home and has $debt_{it+1} = debt_{it}(1+\iota)$.

- In period $t+2$ the individual chooses home and has $debt_{it+2} = debt_{it+1}(1+\iota)$.

G.2.3 Cancellation

It is easily verifiable that $debt_{it+2}$ is the same along both paths by recursively applying the formulas:

- Home-School-Home path:

$$\begin{aligned} debt_{it+2} &= debt_{it+1}(1+\iota) \\ &= \left\{debt_{it}(1+\iota) + E(loan_{i,c})(1+\iota)^{t+1}\right\}(1+\iota) \\ &= \left\{[debt_{it-1}(1+\iota)](1+\iota) + E(loan_{i,c})(1+\iota)^{t+1}\right\}(1+\iota) \\ &= debt_{it-1}(1+\iota)^3 + E(loan_{i,c})(1+\iota)^{t+2} \end{aligned}$$

- School-Home-Home path:

$$\begin{aligned} debt_{it+2} &= debt_{it+1}(1+\iota) \\ &= \{debt_{it}(1+\iota)\}(1+\iota) \\ &= \left\{\left[debt_{it-1}(1+\iota) + E(loan_{i,c})(1+\iota)^{t}\right](1+\iota)\right\}(1+\iota) \\ &= debt_{it-1}(1+\iota)^3 + E(loan_{i,c})(1+\iota)^{t+2} \end{aligned}$$

H Estimation of CCP and offer arrival parameters with search frictions

We make use of a flexible multinomial choice model in order to compute the log CCP terms that enter the formulas for the conditional value functions (see Appendix F). The likelihood function for this model can be written as

$$L = \prod_i \sum_r \pi_r \prod_t \prod_{\{j,k,l\}} \left(P_{ijkltr}\right)^{d_{ijklt}} \tag{H.1}$$

When there are search frictions, however, certain choice alternatives are no longer available under certain cases. Denoting by J^o the entire choice set, and by J^n the limited choice set, the likelihood is modified like so:

$$L = \prod_i \sum_r \pi_r \prod_t \left(\lambda_{itr} \prod_{\{j,k,l\} \in J^o} \left(P^o_{ijkltr}\right)^{d_{ijklt}} + (1-\lambda_{itr}) \prod_{\{j,k,l\} \in J^n} \left(P^n_{ijkltr}\right)^{d_{ijklt}} \right) \tag{H.2}$$

Conditional on the heterogeneity type $R = r$, the log-likelihood of occupational choice of individual i in period t is given by:

$$\ell_{itr} = \begin{cases} \ln\left(\lambda_{itr} \prod_{\{j,k,l\} \in J^o} \left(P^o_{ijkltr}\right)^{d_{ijklt}} + (1-\lambda_{itr}) \prod_{\{j,k,l\} \in J^n} \left(P^n_{ijkltr}\right)^{d_{ijklt}} \right) & \text{if } d_{it} \notin \text{white collar} \\ \ln\left(\lambda_{itr} \prod_{\{j,k,l\} \in J^o} \left(P^o_{ijkltr}\right)^{d_{ijklt}} \right) & \text{if } d_{it} \in \text{white collar} \end{cases} \tag{H.3}$$

where

$$\begin{aligned} \lambda_{itr} &= \frac{\exp\left(Z_{itr}\delta_\lambda\right)}{1+\exp\left(Z_{itr}\delta_\lambda\right)} \\ P^o_{ijkltr} &= \frac{\exp\left(X_{itr}\tilde{\beta}_{jkl}\right)}{\sum_{\{m,n,o\} \in J^o} \exp\left(X_{itr}\tilde{\beta}_{mno}\right)} \\ P^n_{ijkltr} &= \frac{\exp\left(X_{itr}\tilde{\beta}_{jkl}\right)}{\sum_{\{m,n,o\} \in J^n} \exp\left(X_{itr}\tilde{\beta}_{mno}\right)} \end{aligned} \tag{H.4}$$

and where P^{o}_{ijkltr} denotes the (type-specific) choice probability when an offer was received (i.e. the full choice set was available), while P^{n}_{ijkltr} denotes the (type-specific) choice probability when an offer was not received (i.e. limited choice set).

Having estimated the distribution of heterogeneity types in a first step from the measurement equations, we estimate the unknown parameters $\left(\delta_\lambda, \tilde{\beta}\right)$ by maximizing the following weighted log-likelihood, where the weights are given by the posterior type probabilities q_{ir}:

$$\widetilde{\ell} = \sum_i \sum_r q_{ir} \sum_t \ell_{itr} \tag{H.5}$$

where ℓ_{itr} is given in (H.3) above and q_{ir} is as defined in Subsection 5.5.

I Parametric bootstrap procedure

This appendix section details our parametric bootstrap procedure used to obtain standard errors for the model estimates. We compute the standard errors based on $B = 150$ bootstrap replications. We take the following steps to create each bootstrap replication. These steps are similar to the steps we take to simulate the data when computing the fit of our model in Subsection 6.5. Namely:

1. Sample with replacement N individuals from the the data used in the structural estimation procedure.

2. Generate initial conditions for each sampled individual. This includes the unobservable ability vector, the unobserved type, as well as the initial calendar year and the personal and family background characteristics observed in the data. Draw the ability vector from the estimated population distribution $\mathcal{N}\left(0, \hat{\Delta}\right)$. Similarly, draw the unobserved type $r \in \{1, \ldots, R\}$ from the estimated population distribution of types, which is a categorical distribution with R-length parameter vector $\hat{\pi}$ (i.e. $\hat{\pi}_r$ is the estimated probability of being unobserved type r).

3. For each time period until period $T = 19$—the longest panel length in the sample—repeat the following steps on the cross-section of individuals:

   (a) Generate a white collar job offer according to the estimated $\hat{\delta}_\lambda$ in (H.4).

   (b) Generate choices based on the job offer outcome as well as the estimated flow utility parameters of the structural model described in Subsection 3.5.

   (c) Draw the outcomes (wage and/or grade) corresponding to the choice that was just drawn, using the parametric specifications of the grade and wage processes (see Subsections 3.2-3.3).

   (d) If at risk of graduating, draw the graduation status using the predicted graduation probability described in Subsection 3.4.3.

   (e) Compute the implied posterior ability beliefs given the outcomes and choices generated previously as discussed in Subsection 3.4.1, and update the values of all other deterministic state variables.

4. Finally, for each bootstrap sample generated from the previous steps, estimate the model as discussed in Subsection 5.5.

5. Repeat steps 1–4 $B = 150$ times.

Once we have obtained the vector of parameter estimates for all bootstrap replications $\widehat{\Theta}_b$, $b = 1, \ldots, B$, we estimate the variance of $\widehat{\Theta}$ as follows:

$$\widehat{Var}\left(\widehat{\Theta}\right) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\widehat{\Theta}_b - \overline{\widehat{\Theta}}\right)\left(\widehat{\Theta}_b - \overline{\widehat{\Theta}}\right)' \tag{I.1}$$

where $\overline{\widehat{\Theta}} = \frac{1}{B}\sum_{b=1}^{B}\widehat{\Theta}_b$.

Note that, because we simulate the model to form the parametric bootstrap replicates, we have no missing data and hence there is no need to integrate out over the missing majors and grades and detailed in Appendix C.4. For each bootstrap replicate, we simply weight each individual's choice, graduation, and outcome likelihoods by q_{ir}, which is the probability that i is of unobserved type r in the bootstrap sample.

J Details on Counterfactual Simulations

This appendix section details the steps we take for the counterfactual simulations of our model. In our counterfactual simulations, we assume that agents have full information about their abilities. We then solve each individual's dynamic programming problem using backwards recursion.

J.1 Assumptions to simplify the problem

Due to the high dimensionality of the state space, we make the following assumptions to ensure tractability of our simulations:

- Retirement age is 65, meaning $T = 65 - 17 = 48$
- Terminal value is set equal to zero for all individuals and choice paths.
- Time is discrete at annual frequency
- Agents are able to choose college for only the first ten periods (i.e. until age 28)
- Individuals can choose college only in the first 10 periods, meaning $T^* = 10$.[J1]
- Loans have a 20-year repayment horizon that begins in period $T^* + 10$
- Experience variables are capped as follows. In most cases, these caps correspond to the 99th percentile of what we observe in the data:
  - white-collar work experience for non-graduates is capped at 10 years and for graduates is capped at 15 years
  - blue-collar work experience is capped at 15 years
  - total work experience is capped at 15 years
  - 2-year college experience is capped at 4 years
  - 4-year college experience is capped at 6 years
  - total college experience is capped at 7 years

[J1] In estimation, we do not impose this assumption, meaning that individuals can choose college options throughout our entire sample period. In practice this is not binding as annual college enrollment rates are below 2% when $t > T^*$, compared with above 22% when $t \leq T^*$.

We also discretize the AR(1) process governing the aggregate labor market shocks in (13) using Tauchen's (1986) method. We separate the continuous labor market shocks into quartiles and approximate the transitions using a four-by-four Markov transition matrix.

J.2 Outline of steps to compute simulations

We take the following steps to produce simulated data consistent with our model parameters and counterfactual scenario. For each individual in our cross-sectional sample, we repeat this process 10 times:

1. Generate initial conditions (vector of abilities, unobserved type, and personal and family background characteristics). Draw the ability vector from the estimated population distribution $\mathcal{N}\left(0, \hat{\Delta}\right)$. Similarly, draw the unobserved type $r \in \{1, \ldots, R\}$ from the estimated population distribution of types.

2. Solve the model backwards from retirement age using the simplifications detailed in J.1. The resulting policy functions are the individual's probabilities of making each choice at any given set of states.

3. Generate a sequence of observed states by simulating forward from the initial conditions. This entails drawing a job offer outcome, drawing a choice, and then updating the state space corresponding to the sequence of choices.

The process yields a panel data set of simulated choices that has a structure identical to the data we use in estimation.

K Mathematical symbol glossary

This section contains a glossary with descriptions of each mathematical symbol used in the paper or appendices. See Table K1.

Table K1: Mathematical symbol glossary

Symbol	Description	Main equations of reference
α	Flow utility parameters	(16)
β	Discount factor	(20)
$\tilde{\beta}$	Nuisance parameters in measurement system, inputs to consumption, and CCPs	(C.1), (D.3), (H.4)
γ	Grade and wage parameters	(1), (2), (4), (5)
d_{it}	Individual i's choice in period t	(20), (21)
δ_t	Aggregate labor market shock to wages	(4), (6), (13)
δ_λ	White collar offer arrival parameters	(15), (H.4)
ε	Idiosyncratic shocks to grades, wages, preferences, and unobserved type measurements	(1), (4), (16), (20), (C.1), (C.7)
ζ_t	Innovations to aggregate labor market shock	(13)
θ	CRRA parameter on exp. util. of consumption	(16), (D.11)
ι	Interest rate for loan repayment	(G.1)
κ	Cut points for measurement system ordered logit	(C.5)
λ	Grade or wage return to ability or productivity index	(3), (6)
$\tilde{\lambda}_{it}^{d_{t-1}}$	White collar offer arrival probability	(15), (22)
ℓ	Log-likelihood	(32), (C.16), (H.5)
π_r	Population probability mass of unobserved type r	(36)
ρ_t	Probability of having graduated before time t	(F.8)
σ	Standard deviation of various idiosyncratic shocks	(1), (4), (13), (35), (C.2)
$\tilde{\sigma}$	Nuisance parameter in inputs to consumption	(D.3)
τ	Index of enrollment or employment time period, distinct from calendar time t	(7)
ϕ	AR(1) coefficient on aggregate labor market shock	(13)
φ	Normal distribution pdf	(27), (30), (32), (C.2)
ψ	Graduation logit parameters	(14)
ω	Unobserved type intercept in measurement system	(C.1)
A_i	5-dimensional ability vector that is gradually revealed to the individual	(1), (2), (4), (5)
Δ	5×5 population covariance matrix of abilities	(27), (31)
Z_{it}	State variables for individual i in period t	(16)
$\tilde{X}_{it}$	Covariates in job offer arrival logit	(15)
Θ	Collection of all measurement system parameters, or collection of all model parameters	(C.2), (I.1)
Λ_t	Posterior variance of ability at time t	(11)
Φ	Standard normal cdf	(C.3), (D.13)
X	Covariate matrices in various equations	(1), (4), (14)
Ω	Inverse of variance of idiosyncratic grade and wage shocks	(11)