College Attrition and the Dynamics of Information Revelation*

Peter Arcidiacono† Esteban Aucejo‡
Arnaud Maurel§ Tyler Ransom¶

April 14, 2023

Abstract

We examine how informational frictions impact schooling and work outcomes. We estimate a dynamic structural model where individuals face uncertainty about their ability and productivity, which respectively determine their schooling utility and work wages. We account 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. We find that removing informational frictions would barely affect overall college graduation rates, but would substantially increase the college and white-collar wage premiums, while reducing the college graduation gap by family income.

JEL Classification: C35; D83; J24

*First version: June 2016. We thank the Editor, Jim Heckman, four anonymous referees, Stephane Bonhomme, Chris Flinn, Mitch Hoffman, 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.

†Duke University, IZA and NBER. E-mail address: psarcidi@econ.duke.edu

‡Arizona State University, W.P. Carey School of Business, CEP, CEPR, CESifo and NBER. E-mail address: Esteban.Aucejo@asu.edu

§Duke University, IZA and NBER. E-mail address: apm16@duke.edu

¶University of Oklahoma, IZA and GLO. E-mail address: ransom@ou.edu

1 Introduction

About 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 own academic ability and labor market productivity in determining students' outcomes and sorting in the labor market.

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. Upon college graduation, which is probabilistic from the individual 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 in our model 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.

We estimate a richer model than previously possible by making use of recent 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), estimation continues to be computationally simple even in the presence of unobserved heterogeneity that is known to the individual. Using 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 that are initially unknown to the individuals. Focusing on the ability components which are 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 large, stressing the importance of allowing for correlated learning in this context. On the other hand, our estimation results indicate that schooling abilities are only weakly correlated with productivity in both sectors, thus indicating that grades earned in college actually reveal little new information about future labor market performance, once we account for background characteristics and college readiness. The lack of learning about work abilities prior to college graduation results in college graduates that are somewhat well sorted on the basis of college abilities but not on work abilities.

We then simulate our model under a counterfactual scenario where all individuals have perfect 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 the share of four-year college graduates is virtually unchanged relative to the baseline, though this masks significant decreases in the share dropping out of college and significant increases in the share who never attend college. It also masks a substantial reduction in the gap in college graduation rates between individuals from high- and low-income families. The mechanism for this convergence is 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 indeed are a good match, increasing their college graduation rates. The reverse holds true for high-income

[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.

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 and having already accumulated some years of college experience.

Simulations further 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 more than doubles, and the wage gap between white-collar and blue-collar workers almost triples.

Our analysis builds on seminal 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 authors, 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]

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 the NLSY97. 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

[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.

[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.

the literatures on schooling choices, and on occupational choices under imperfect information (see, e.g., Miller, 1984; James, 2011; Antonovics and Golan, 2012; Papageorgiou, 2014; Sanders, 2014). Beyond the educational choice context, our paper also fits within the rich empirical literature on dynamic learning models, which, since the seminal work of Erdem and Keane (1996), have often been estimated in marketing. A key additional challenge that arises in the schooling context is that we do not observe individuals making the same choice multiple times, which makes 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 annually on topics such as family background 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. After imposing some additional data restrictions, the final sample includes 22,398 person-year observations for 2,300 men.[4]

2.1 Descriptive Statistics

Table 1 presents background characteristics conditional on the first college option chosen. Individuals who attend college at some point and start at a four-year institution have, on

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

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. Similarly, Table 2 presents summary statistics but focuses instead on occupation (i.e., blue-collar vs. white-collar) conditional on college graduation status. The sample corresponds to individual-year observations. As expected, individuals in the white-collar sector have higher SAT scores, high school GPA, and family income (when a teenager) than those in blue-collar occupations once conditioning on graduation outcomes.[5] Overall, these compositional differences between two- and four-year colleges (and between majors in four-year colleges) and blue-collar and white-collar occupations highlight the importance of distinguishing between educational and occupational sectors when modeling choices.

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.[6] 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

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

[6] 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.

a later period.[7] 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. Therefore, a realistic model should allow for flexible schooling options.

2.2 Descriptive Investigation of Learning

The large dropout and stopout rates suggest that students are likely 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.

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. In this regard, Panels A and B of Table 4 show that the 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.

[7] 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. $24.86\%/(24.86\% + 35.94\%) = 41.67\%$.

Finally, to illustrate learning in the labor market as a reason for stopouts to return to college, we performed 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 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 2.190 while those who leave but do not later return show a mean log wage of 2.362. 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 suggests that individuals likely learn about their own abilities when participating in different sectors. However, in order to fully isolate the empirical relevance of learning, we next present a structural model of college and labor market decisions where learning constitutes a key ingredient in determining college and labor market outcomes.

3 Model

3.1 Overview

After graduating from high school, individuals in each period make a joint schooling and work decision. 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. In addition, individuals can also choose whether to work either full-time or part-time in the blue-collar or white-collar sector, these work options being available to the individuals regardless of the schooling choices.[8] Individuals may always choose blue-collar work but face frictions in obtaining white-collar work. Finally, home production (i.e., neither work nor attending college) is an available option every period.[9]. After graduation, schooling options are no longer part of the choice

[8] Keane and Wolpin (2001) and Joensen (2009) estimate dynamic structural models of schooling and work decisions and also allow for work while in college.

[9] Overall, individuals can choose among 20 different options before graduation

set. Therefore, 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 occupations). 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.

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. Since the different schooling abilities will likely be correlated in addition to being correlated with different labor market abilities, grades in one of the schooling options will provide information regarding the student'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 take into account 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. It is interesting to note that, in this setting, even though working while in college may be detrimental to academic performance (see, e.g., Stinebrickner and Stinebrickner, 2003), it also serves as an additional channel through which individuals can learn both about their productivities and schooling abilities while in school. Our framework incorporates this tradeoff.

We now detail the main elements of the model. We first discuss the components individuals are forming beliefs over, namely, the grade and wage equations, and the probability of graduating. We then describe how individuals update their beliefs. 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 attended, 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,[10] $X_{ij\tau}$, 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).[11] In the following and throughout the paper, we assume 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:

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

The academic index $AI_{ij\tau}$ gives 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 varies over the periods of college enrollment and across majors. As such, while remaining parsimonious, this specification allows for different

[10] 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 be in his $\tau = 3^{\text{rd}}$ period of college enrollment.

[11] The specification for grades in two-year college also includes an indicator for whether an individual has spent more than one year in this type of college.

effects of ability on grades for lower- and upper-classmen. 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.[12] 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 and blue-collar work options. With probability $1 - \tilde{\lambda}_{it}^{(d_{t-1})}$, the worker cannot choose white-collar work. Those who are not college graduates or advanced degree holders may work in the white-collar sector if they receive an offer. Likewise, college graduates or advanced degree holders may *not* receive a white-collar offer, in which case they work in the blue-collar sector if choosing employment. We emphasize that $\tilde{\lambda}_{it}^{(d_{t-1})}$ is a function of individual characteristics such that this probability matches observed white-collar employment rates in the data.

Log wages in sector l are assumed to differ based on whether the individual is currently attending college or not. For those who are not working while in school, 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} (i.e., 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), 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.

[12]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.

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

$$w^s_{ilt} = \delta_t + \lambda_{0l} + \lambda_{1l} PI_{ilt} + \varepsilon^s_{ilt} \tag{6}$$

where the idiosyncratic shocks, ε^s_{ilt}, are assumed to be distributed $\mathcal{N}(0, \sigma^2_{sl})$. Hence, the return to the productivity index varies 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.

We account for aggregate changes in wages over time through calendar year indicators, δ_t. If we did not control for these nonstationarities, we may 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 3.5.2.

3.4 Consumption

We now discuss how consumption enters our model. 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 loans to pay for their education). Denoting consumption by C, labor income by W, parental transfers by PT, educational grants (both need- and merit-based) by G, educational loans by L (the latter two of which are a function of EFC,[13] family income I, SAT test score), tuition and fees by T, and school loan repayments (which are taking place when $t \geq t^*$) by SLR, we obtain the budget constraint denoted in eq.(7).[14] The repayment of loans is prespecified by a formula

[13]EFC stands for Expected Family Contribution, which is a number that colleges use to compute how much need-based financial aid a given student is eligible for.

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

(see Appendix G for more details), which avoids modeling how individuals decide to pay for them.[15] We first define the latent consumption level C^* as:

$$C^* = \begin{cases} W & \text{if working \& not in school \& t} < t^* \\ W + PT + G\left(EFC, I, SAT\right) + L\left(EFC, I, SAT\right) - T & \text{if working while in school \& t} < t^* \\ PT + G\left(EFC, I, SAT\right) + L\left(EFC, I, SAT\right) - T & \text{if in school \& not working \& t} < t^* \\ W - SLR & \text{if working \& not in school \&} \ t \geq t^* \\ \underline{C} & \text{if not in school \& not working} \end{cases} \tag{7}$$

The (realized) consumption level is then given by:

$$C = \max(C^*, \underline{C}) \tag{8}$$

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 $\underline{C} = \$2,800$. Note that this consumption floor also operates on the other cases. We discuss in detail in Appendix E the specification of the consumption process, including how each component of consumption is computed. We provide details on how individuals accumulate debt in Appendix G.

3.5 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 3.6 when we describe preferences. We discuss the other components here.

3.5.1 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

[15] In fact, as long as the loan repayment period starts after the estimation period, then we do not need to take a statement on how repayments should impact estimation.

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:

$$S_{ij\tau} = G_{ij\tau} - \gamma_{0j} - X_{ij\tau}\gamma_{1j} \tag{9}$$

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

$$S_{ij\tau} = \frac{G_{ij\tau} - \lambda_{0j} - \lambda_{1j}(\gamma_{0j} + X_{ij\tau}\gamma_{1j})}{\lambda_{1j}} \tag{10}$$

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:

$$S_{ilt} = w_{ilt} - \delta_t - \gamma_{0l} - X_{ilt}\gamma_{1l} \tag{11}$$

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

$$S_{ilt} = \frac{w_{ilt} - \delta_t - \lambda_{0l} - \lambda_{1l}(\gamma_{0l} + X_{ilt}\gamma_{1l})}{\lambda_{1l}} \tag{12}$$

Finally, individuals may choose to work while in college, in which case they will receive two ability signals $(S_{ij\tau}, S_{ilt})$.[16]

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, skilled or unskilled labor market). The diagonal elements corresponding to the choices made are given by the inverse of the variances of the

[16] Once an individual graduates from college, we assume that learning occurs only through wage signals.

idiosyncratic shocks.[17] 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.[18]

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):

$$E_t(A_i) = (\Lambda_{t-1}^{-1}(A_i) + \Omega_{it})^{-1}(\Lambda_{t-1}^{-1}(A_i)E_{t-1}(A_i) + \Omega_{it}\widetilde{S}_{it}) \tag{13}$$

$$\Lambda_t(A_i) = (\Lambda_{t-1}^{-1}(A_i) + \Omega_{it})^{-1} \tag{14}$$

As in a more standard one-dimensional learning model, prior variances at the beginning of a given period decrease towards zero as individuals receive additional ability signals in the previous periods, thus giving more weight to the prior ability and less to the current-period signal.

3.5.2 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:

$$\delta_t = \phi\delta_{t-1} + \zeta_t \tag{15}$$

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 (Markov process of order 1) is common in the literature

[17] 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.

[18] 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 $\widetilde{S}_{it}$:

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

(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.

3.5.3 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.[19] We then assume that the probability of graduating conditional on $X_{ig\tau}$ takes a logit form:

$$\Pr(\text{grad}_{i\tau} = 1|X_{ig\tau}) = \frac{\exp(X_{ig\tau}\psi)}{1 + \exp(X_{ig\tau}\psi)} \tag{16}$$

Individuals are assumed to know the parameters ψ and form expectations over their probabilities of graduating using (16).

3.5.4 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 with probability, $\tilde{\lambda}_{it}^{d_{t-1}}$, which is specified as follows:

$$\tilde{\lambda}_{it}^{d_{t-1}} = \begin{cases} \frac{\exp(\widetilde{Z}_{it}'\delta_\lambda)}{1+\exp(\widetilde{Z}_{it}'\delta_\lambda)} & \text{if } i \text{ did not work in the WC sector at time } t-1 \\ 1 & \text{if } i \text{ worked in the WC sector at time } t-1 \end{cases} \tag{17}$$

where $\widetilde{Z}_{it}$ denotes a vector of state variables in period t (i.e. age and an indicator for graduation) and the superscript d_{t-1} emphasizes the dependence on the prior decision as outlined in (17). Similar to graduation, individuals are assumed to know the parameters δ_λ.

[19] 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 to allow for correlated Bayesian learning based on a mixed continuous-discrete distribution of signals.

Section 4 discusses the identification of $\tilde{\lambda}_{it}$.

3.6 Flow utilities

We now define the flow payoffs for each of the schooling and work combinations. We denote 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 $Z_{it} = (Z_{0it}, Z_{1it}, Z_{2it})$:

$$\begin{aligned} U_{jkl}(Z_{it}, \varepsilon_{ijklt}) &= \alpha_{jkl} + \alpha_C E(U(C_{ijklt})) + Z_{1it}\alpha_j + Z_{2it}(\alpha_k + \alpha_l) + \varepsilon_{ijklt} && (18)\\ &= u_{jkl}(Z_{it}) + \varepsilon_{ijklt} && (19) \end{aligned}$$

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

We discuss below the different components of the flow utility payoffs, starting with the utility of consumption. We assume that individuals have CRRA preferences over their consumption, with a risk aversion parameter which we set equal to $\theta = 0.4$; see E.5 for details on how we chose the value of θ. For a given sector l and period t, the expected utility of consumption depends on the expected labor income in that sector, along with the beliefs about the aggregate shocks affecting sector l in period t. Consumption while in college also depends on parental transfers, educational grants (both need- and merit-based), and loans, along with tuition fees.

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, controls for the previous choice (to allow for switching costs similar in spirit to Keane and Wolpin, 1997), and individual (latent) specific effects (assumed independent of the other covariates).[20] Our approach is consistent

[20]More specifically, we assume that agents can be characterized by a finite group of types and that these

with students accumulating human capital while in college, with some 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.[21]

- 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), years of college completed, interactions of previous college choices with years of college completed, and work intensity (whose effects are allowed to vary by labor market sector). Expected ability is included in Z_{1it} as an (inverse) proxy of the cost of effort associated with college attendance.[22] Finally, indicators for work intensity by sector capture the costs associated with balancing both activities simultaneously.

  Our model, however, does not incorporate endogenous effort adjustment. While this would in principle be an interesting extension, we conjecture that in the absence of measurement of effort, such a model would not be identified.[23]

- 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, the inclusion of four-year college graduation acts 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.

3.7 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 $(d_{it})_{t=1...T}$, a combination of schooling and work decisions, to

types are constant across agents within each group.

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

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

[23]See recent work by De Groote (2022) who estimates a dynamic model of high school effort choice in the absence of learning.

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 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:

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

where γ denotes Euler's constant.

3.8 Finite dependence

Following Arcidiacono and Miller (2011), we re-express the future payoffs so we can avoid solving the full backward recursion problem. Recall that it is the difference in the conditional value functions that are relevant in estimation, not the conditional value functions themselves. For simplicity, denote 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, $v_{jkl}-v_h$, that is not recursive.[24] To accomplish this, we rely on the finite dependence structure of the problem, such that it can be reformulated in terms of two-period ahead flow payoffs and conditional choice probabilities (CCPs are estimated in a first stage).[25] However, given the presence of search frictions, it is also necessary to include (convenient) weights to achieve the cancellation of the recursive term.[26]

To be more specific on how the cancellation in the finite dependence path works, we express the value function associated with $d_{it} = h$ (home production) as follows, denoting by $\tilde{\lambda}^{(h)}_{i,t+1,s}$ the job offer arrival rate in the white-collar sector:

$$\begin{aligned} E_t\left[V_{t+1}\left(Z_{it+1}\right)|d_{it}=h\right] = \quad & \tilde{\lambda}^{(h)}_{i,t+1,s} E_t\left[V_{t+1}\left(Z_{it+1}\right)|d_{it}=h, \text{Offer}_{it}=1\right] \\ & + (1-\tilde{\lambda}^{(h)}_{i,t+1,s}) E_t\left[V_{t+1}\left(Z_{it+1}\right)|d_{it}=h, \text{Offer}_{it}=0\right] \end{aligned} \tag{23}$$

One can distinguish between three possible events in period $t+1$: (i) the agent hasn't received a job offer in the white-collar sector, (ii) the agent has received a job offer in the white-collar sector and has accepted it, and (iii) the agent has received a job offer in the white-collar sector and has rejected it. Then, as we show in more detail in Appendix F, including convenient weights on the cases (ii) and (iii) makes it possible to achieve cancellation of the recursive term when uncovering an expression for $v_{jkl} - v_h$.[27] Suppressing

[24] 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.

[25] More specifically, 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 $t+2$.

[26] Given that white-collar job offers arrive with probability $\tilde{\lambda}^{(h)}_{t+1}$, then we respectively weight the acceptance and rejection probabilities in $t+1$ of the offers (in the event that i receives an offer) by $\frac{1}{\tilde{\lambda}^{(h)}_{t+1}}$ and $1 - \frac{1}{\tilde{\lambda}^{(h)}_{t+1}}$, allowing us to achieve cancellation.

[27] Appendix F also details the case where there is uncertainty about graduation, which we avoid here for expositional reasons.

the i subscript, the differenced conditional value function is then given by:

$$v_{jkl}(Z_t) - v_h(Z_t) = \left(\begin{array}{l} u_{jkl}(Z_t) - \beta E_t \left\{ \tilde{\lambda}_{t+1}^{(jkl)} \ln \left[p_h \left(Z_{t+1} | \text{Offer}_{t+1} = 1 \right) \right] + \right. \\ \left. \left(1 - \tilde{\lambda}_{t+1}^{(jkl)} \right) \ln \left[p_h \left(Z_{t+1} | \text{Offer}_{t+1} = 0 \right) \right] \middle| Z_t, d_t = (j,k,l) \right\} \\ +\beta E_t \left\{ \left(\frac{1}{\tilde{\lambda}_{t+1}^{(h)}} \right) \tilde{\lambda}_{t+1}^{(h)} \ln \left[p_{jkl} \left(Z_{t+1} | \text{Offer}_{t+1} = 1 \right) \right] + \right. \\ \left(1 - \frac{1}{\tilde{\lambda}_{t+1}^{(h)}} \right) \tilde{\lambda}_{t+1}^{(h)} \ln \left[p_h \left(Z_{t+1} | \text{Offer}_{t+1} = 1 \right) \right] + \\ (1 - \tilde{\lambda}_{t+1}^{(h)}) \ln \left[p_h \left(Z_{t+1} | \text{Offer}_{t+1} = 0 \right) \right] - \\ \left. \left(\frac{1}{\tilde{\lambda}_{t+1}^{(h)}} \right) \tilde{\lambda}_{t+1}^{(h)} u_{jkl}(Z_{t+1}) \middle| Z_t, d_t = h \right\} \\ +\beta^2 E_t \left\{ \tilde{\lambda}_{t+2}^{(jkl)} \ln \left[p_h(Z_{t+2} | \text{Offer}_{t+2} = 1) \right] + \right. \\ \left. \left(1 - \tilde{\lambda}_{t+2}^{(jkl)} \right) \ln \left[p_h(Z_{t+2} | \text{Offer}_{t+2} = 0) \right] \middle| Z_{t+1}, d_t = h, d_{t+1} = (j,k,l) \right\} \\ -\beta^2 E_t \left\{ \left(\tilde{\lambda}_{t+2}^{(h)} \ln \left[p_h(Z_{t+2} | \text{Offer}_{t+2} = 1) \right] \right) + \right. \\ \left. \left(1 - \tilde{\lambda}_{t+2}^{(h)} \right) \ln \left[p_h(Z_{t+2} | \text{Offer}_{t+2} = 0) \right] \middle| Z_{t+1}, d_t = (j,k,l), d_{t+1} = h \right\} \end{array} \right) \tag{24}$$

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 ability (to the econometrician, but known by the agents) and preference parameters 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. We allow for three unobserved binary heterogeneity factors, namely one for schooling ability, one for schooling preferences, and one for work productivity and preferences. Therefore, eight types in total capture the unobserved heterogeneity, where individuals are characterized as low (L) or high (H) for each of the three latent dimensions.[28] We rely on an auxiliary measurement system, which is primarily used

[28] The possible combinations for schooling ability, schooling preferences, and work productivity and preferences 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). We also estimated two extended specifications, allowing for 16 and 32 heterogeneity types. For each of these cases, several of the estimated population type probabilities ended up being negligibly small (below 0.5%).

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. 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.

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 for identification the auxiliary measurement system discussed in Subsection 4.1. 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.[29] 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 types (i.e. that 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.

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:

$$\Pr(D|M = m) = \Pr(D|M = m, R = 1)\pi_1^{M=m} + \Pr(D|M = m, R = 2)(1 - \pi_1^{M=m}) \tag{25}$$

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:

$$\begin{pmatrix} \Pr(D|M = m_1) \\ \Pr(D|M = m_2) \end{pmatrix} = \begin{pmatrix} \pi_1^{M=m_1} & 1 - \pi_1^{M=m_1} \\ \pi_1^{M=m_2} & 1 - \pi_1^{M=m_2} \end{pmatrix} \begin{pmatrix} \Pr(D|R = 1) \\ \Pr(D|R = 2) \end{pmatrix} \tag{26}$$

[29] 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.

This identifies the vector of type-specific conditional choice probabilities $\begin{pmatrix}\Pr(D|R=1)\\ \Pr(D|R=2)\end{pmatrix}$ as long as 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.[30]

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 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

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 signal 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 primarily identified from the covariances between the outcomes (grades and log wages) and past ability signals.

[30] 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).

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 plays an important role in 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, $\mathrm{Cov}(Y, S|X, D = 1)$ as well as $\mathrm{Cov}(Y^2, S|X, D = 1)$ are both functions of the same covariance term $\mathrm{Cov}(R, S|X, D = 1) = \mathrm{Cov}(R^2, S|X, D = 1)$. The share of the covariance $\mathrm{Cov}(Y, S|X = x, D = 1)$ that is attributable to the unobserved ability A is then identified from the covariances $\mathrm{Cov}(Y, S|X = x, D = 1)$ and $\mathrm{Cov}(Y^2, S|X = x, D = 1)$, letting x vary on the support of X. This, in turn, identifies the signal-to-noise ratios associated with the different 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 Structural utility function parameters

We 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.[31] In other words, it follows from the specification of our choice model that:

$$\lim_{y \to \infty} \Pr(D_t = W|\text{Offer}_t = 1, Z_t, E_{t-1}(A_{-W}), E_{t-1}(A_W) = y, R) = 1 \tag{27}$$

[31] Note that 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.

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 directly follows that, for this subset of individuals, the offer arrival rates are equal to the probabilities of working in the white-collar sector $\Pr(D_t = W|Z_t, E_{t-1}(A_{-W}), E_{t-1}(A_W), R)$. These type-specific conditional probabilities are identified for this subset of individuals in a similar fashion as on p.23. 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 by $\widetilde{\lambda}_t^R$ to make the dependence on R explicit. Finally, it follows that the probabilities to work in the white-collar sector conditional on receiving an offer, which are given by $\Pr(D_t = W|Z_t, E_{t-1}(A_{-W}), E_{t-1}(A_W), R) = \frac{\Pr(D_t=W|Z_t,E_{t-1}(A_{-W}),E_{t-1}(A_W),R)}{\widetilde{\lambda}_t^R}$, 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 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 B6 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.[32] 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 (2020) (see, in particular, Theorem 3 and discussion in Section 3.3 of that paper). Identification relies on the finite dependence property of our model, discussed in detail in

[32] 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.

Subsection 3.8.

5 Estimation

For expositional reasons, we first present the estimation procedure for the specification without type-specific unobserved heterogeneity (Subsections 5.1-5.4). We then discuss how the procedure can be extended to allow for unobserved heterogeneity types (Subsections 5.5-5.6).[33]

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.[34] 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 $(G_{i\tau})_\tau$ ($\tau \in \{1, \ldots, T_c\}$) the grades, $(w_{iB\tau})_\tau$ ($\tau \in \{1, \ldots, T_B\}$) the log wages in the blue-collar sector, $(w_{iW\tau})_\tau$ ($\tau \in \{1, \ldots, T_W\}$) the log wages in the white-collar sector, and

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

[34] 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 B5 for a full summary of the estimation steps.

$(d_{i\tau})_\tau$ the decisions ($\tau \in \{1, \ldots, T_d\}$), as a five-dimensional integral:

$$\begin{aligned} &L(d_{i1}, \ldots, d_{iT_d}, G_{i1}, \ldots, G_{iT_c}, w_{iB1}, \ldots, w_{iBT_B}, w_{iW1}, \ldots, w_{iWT_W}) \\ &= \int L(d_{i1}, \ldots, d_{iT_d}, G_{i1}, \ldots, G_{iT_c}, w_{iB1}, \ldots, w_{iBT_B}, w_{iW1}, \ldots, w_{iWT_W} | A) \varphi(A) dA \end{aligned} \tag{28}$$

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 log wages):

$$L(d_{i1}, \ldots, d_{iT_d}, G_{i1}, \ldots, G_{iT_c}, w_{iB1}, \ldots, w_{iBT_B}, w_{iW1}, \ldots, w_{iWT_W}) = L_{id} \times L_{iy} \tag{29}$$

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

$$\begin{aligned} L_{id} &= L(d_{i1}) L(d_{i2} | d_{i1}, G_{i1}) \cdots \\ &\quad \times L(d_{iT_d} | d_{i1}, d_{i2}, \ldots, d_{iT_d - 1}, G_{i1}, G_{i2}, \ldots w_{iB1}, w_{iB2}, \ldots, w_{iW1}, w_{iW2}, \ldots) \end{aligned} \tag{30}$$

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:

$$\begin{aligned} L_{iy} = \int & L(G_{i1} | d_{i1}, A) \cdots L(G_{iT_c} | d_{i1}, d_{i2}, \ldots, A) L(w_{iB1} | d_{i1}, A) \cdots L(w_{iBT_B} | d_{i1}, d_{i2}, \ldots, A) \\ & \times L(w_{iW1} | d_{i1}, A) \cdots L(w_{iWT_W} | d_{i1}, d_{i2}, \ldots, A) \varphi(A) dA \end{aligned} \tag{31}$$

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 (29) results in the choice part of the log-likelihood being additively separable from the outcome (grades and log wages) part of the log-likelihood.[35]

[35] 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.

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:[36]

- 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 previous iteration. This follows from the Bayesian updating formulas for the posterior ability mean and covariance given in Section 3.5.1 (eq. 13). The (population) covariance matrix of the ability distribution is then updated as follows for each iteration k of the EM estimation:

$$\Delta^k = \frac{1}{N}\sum_{i=1}^{N}\left(\Lambda_i^k(A) + E_i^k(A)E_i^k(A)'\right) \tag{32}$$

  where N denotes the number of individuals in the sample, $E_i^k(A)$ the posterior ability mean ($E_i^k(A)'$ its transpose) and $\Lambda_i^k(A)$ 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 k of the EM estimation, denoting by $\varphi_i^k(\cdot)$ the pdf of the posterior ability distribution computed at the E-step, we maximize the expected complete log-likelihood El_i^k:

$$\begin{aligned} El_i^k &= \int \ln\left[L(G_{i1}|d_{i1},A)\cdots L(G_{iT_c}|d_{i1},d_{i2},\ldots,A)L(w_{iB1}|d_{i1},A)\cdots\right. \\ &\qquad \left.\times\ L(w_{iBT_B}|d_{i1},d_{i2},\ldots,A)\right]\varphi_i^k(A)dA \\ &= El_{i,2}^k + El_{i,4S}^k + El_{i,4N}^k + El_{i,W}^k + El_{i,B}^k \end{aligned} \tag{33}$$

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

For instance, the parameters of the grade equation for college sector j are updated by maximizing the contribution $El^k_{i,j}$, $j \in \{2, 4S, 4N\}$, which writes, denoting by $\varphi^k_{ij}(\cdot)$ the pdf of the posterior distribution of A_j:

$$El^k_{i,j} = \int \left\{ \ln\left[L(G_{i,j,1}|d_{i1}, A_j)\right] + \cdots + \ln\left[L(G_{i,j,T_j}|d_{i1}, d_{i2}, \ldots, A_j)\right] \right\} \varphi^k_{ij}(A_j) dA_j \tag{34}$$

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:

$$\begin{aligned} &\int \ln(L(G_{ij\tau}|d_{i1}, d_{i2}, \ldots, A_j))\varphi^k_{ij}(A_j)dA_j = \\ -\frac{1}{2}\ln(2\pi\sigma^2_{j\tau}) - &\frac{1}{2\sigma^2_{j\tau}}\left(\lambda^2_{1j}\Lambda^k_{ijj}(A) + (G_{ij\tau} - \lambda_{0j} - \lambda_{1j}AI^k_{ij\tau})^2\right) \end{aligned} \tag{35}$$

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

$$\min \sum_{i,\tau} \left(\ln(\sigma^2_{j\tau}) + \frac{1}{\sigma^2_{j\tau}}\left(\lambda^2_{1j\tau}\Lambda^k_{ijj}(A) + (G_{ij\tau} - \lambda_{0j\tau} - \lambda_{1j\tau}AI^k_{ij\tau})^2\right)\right) \tag{36}$$

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 wage parameters in (6) proceeds in a similar manner, with two main differences: (i) the idiosyncratic wage shock variances σ^2_{sl} 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 (36) by taking the sum across sectors, too.

5.3 Estimation of the graduation and search friction parameters and labor market 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 16, Section 3.5.3).

To estimate 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. Appendix H discusses the implementation in detail.

We estimate the parameters ϕ and σ_ζ^2 in (15) by maximum likelihood using the estimated values of δ_t as data.

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 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 3.8).

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

1. Estimate the CCPs (p) 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. By using the finite dependence property of our model, we only need to integrate out over the future shocks for two periods.[37]

[37] 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 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 type is orthogonal to the covariates at $t = 1$. To this end, we estimate (in the first stage) a measurement system to identify three unobserved discrete heterogeneity factors: one for schooling ability, one for schooling preferences, and one for work productivity and preferences.

Accounting for type-specific unobserved heterogeneity breaks down the non-separability between the choice and outcome components of the likelihood described above as the full log-likelihood function can be rewritten as:

$$\sum_i \ln\left[\sum_{r=1}^{R} \pi_r L_{imr} L_{idr} L_{ibr} L_{iyr}\right] \tag{37}$$

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.

Estimation with unobserved heterogeneity proceeds as follows. We estimate in a first step the conditional probabilities of being of each type (q_{ir}) along with the parameters of the measurement system using the EM algorithm (see Appendix C) .[38] This approach follows in spirit Arcidiacono and Miller (2011), where the EM algorithm allows us to restore the additive separability of the likelihood function despite the presence of unobserved heterogeneity.[39]

After recovering the conditional probabilities of each type (q_{ir}), we can 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.[40]

[38]In particular, the unobserved type is treated as if it were observed. We then update the posterior probability of i belonging to the r-th type using the joint likelihood of the measurements according to (given Bayes' rule):

$$q_{ir} = \frac{\pi_r \mathcal{L}_{imr}}{\sum_{r'=1}^{R} \pi_{r'} \mathcal{L}_{imr'}} \tag{38}$$

[39]Rather than updating the structural parameters of the decision process at each step, we rely on a two-stage estimation strategy.

[40]Note that this is identical to the case without unobserved heterogeneity, except that now the q_{ir}'s are used

Finally, standard errors are estimated using a parametric bootstrap procedure with 150 replications.[41] We discuss this procedure in detail in Appendix I.

5.6 Missing college majors and grades

In our data, college grades and four-year college majors are each missing at a non-trivial rate. This is especially true for those who drop out of college by the end of the first period. These individuals likely received negative grade realizations, which could bias our results if ignored. We consider this issue within our estimation procedure by treating the first instance of missing grades or major as another unobserved discrete latent variable. In the case of missing grades, we approximate their distribution with a finite support discrete distribution.[42]

Bear in mind that the estimation approach discussed above can be easily adjusted to allow for these additional latent variables. Specifically, along with the type-specific unobserved heterogeneity distribution, the distribution of unobserved grades and majors (conditional on each heterogeneity type) is estimated in the first stage of our estimation procedure.[43] Therefore, the maximization of the full log-likelihood conditions on both the unobserved heterogeneity type as well as the major or grade quartile.[44] We summarize our complete estimation procedure in Table B5.

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. Our model entails estimating

as weights. The CCPs are identified from the data and could in principle be estimated nonparametrically. However, we decided to estimate them using a parametric specification to avoid the curse of dimensionality. Nevertheless, we implemented highly flexible specifications that include a large set of covariates in addition to accounting for the unobserved types. Appendix Table B15 presents the specifications and estimates corresponding to the CCPs, while Appendix H provides more details on how the flexible conditional choice probabilities are estimated.

[41] For finite mixture models like ours, 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.

[42] More specifically, we discretized the distribution of grades in quartiles. The corresponding cut points of the grades distribution occur at $\{0, 2.5, 3.0, 3.6, 4.0\}$.

[43] The implementation is fully described in Appendix D.

[44] Accounting for missing grades and majors in this fashion results in a finite mixture model with $R \times 2 \times 4 = 8R$ points of support.

over 2,000 parameters (see Table B6). 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 which determines the students' probabilities of being each unobserved type are given in Tables B9–B11 of Appendix B. 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 B13 and B14.

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, though the effects is smaller in 2-year colleges. Working full-time hinders performance, and this is especially true for the four-year science option. Interestingly and consistent with the findings of Hansen, Heckman, and Mullen (2004) regarding the effects of latent ability on achievement test scores, returns to the ability index are found to be smaller after sophomore year for both groups of majors in four-year colleges.

Turning to the type-specific unobserved ability (known to the agent), those with high school-specific ability see higher grades in the four-year options, especially in science. There is also evidence that high schooling preferences are associated with higher grades in science though the effects are small. Unobserved preferences for work are not associated with higher grades, and none of the other unobserved components are associated with higher grades in two-year college.

6.2 Wage parameters

Estimates of the wage equations are given in Table 7.[45] Each year of work experience in the blue-collar sector corresponds to roughly a four-percentage-point 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 1.7 percentage points higher in the white-collar sector.

Returns to schooling are higher in the white-collar sector than in the blue-collar sector. For each year of schooling (up to a maximum of four), workers see 2.7 (4.9) percentage points

[45]Estimates of the AR(1) processes governing the aggregate shocks are reported in Table B12.

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 5.2 (7.4) percentage points in the blue (white) collar sector. Graduating in the sciences is even more lucrative with an additional increase of 9.9 (14.7) percentage points in the blue (white) collar sector. All else equal, the total premium for a non-science graduate relative to someone with no college experience is then respectively 16 and 27 percentage points in the blue- and white-collar sectors; the similar numbers for science graduates are 26 and 42 percentage points. 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 a blue-collar occupation, 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 (15 vs. 10 percentage points).

The type-specific unobserved heterogeneity parameters suggest that the low-work-preference type faces an earnings penalty and that this is especially true if they are also a low-schooling-ability type. With the exception of the coefficient on Black (negative and significant in both sectors), the other background measures (HS grades, parent graduated from college, and Hispanic) are small and insignificant in both sectors.

Finally, the returns we have described apply when the individual is not in school. Returns are dampened to all characteristics 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 about 0.67 in both sectors, implying that in-school work is associated with about a 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. A 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 0.78. Positive, though weaker, correlations are also seen with 2-year college ability, ranging from 0.28 with 4-year science ability to 0.44 with 4-year non-science ability. Work abilities are also strongly correlated (estimated correlation coefficient of 0.67).

The correlations between schooling abilities and labor market productivity are generally positive but markedly lower than the correlations across college types and majors. None of the correlations are significantly different from zero. Taken together, these patterns provide clear indication that grades earned in college, regardless of schooling type, reveal little new information about future labor market performance.

The variances of each of the unobserved ability measures and the outcomes are given in the bottom two rows of Table 8. These estimates provide clear evidence that individuals have a substantial amount of uncertainty about their own abilities by the end of high school. A sizable share of the dispersion in college grades and wages is attributable to the ability components that are gradually revealed to the individuals. Specifically, those ability components account for between 24% of the variance of grades in two-year college, to as much as 46% of the variance of log wages in the white-collar sector. Even in the case with the smallest variance—ability in the blue-collar sector—a one-standard-deviation increase in ability would translate into a 29% increase in wages.

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. In the first year of college where the signal-to-noise ratios are 0.45, 0.41, and 0.25 for four-year science, four-year non-science, and two-year, respectively.[46] The signal-to-noise ratios in the blue-collar sector are similar regardless of whether the individual is in school at the time of employment, 0.41 and 0.37 for in-school and out-of-school, respectively. However, working while out of school is more informative in the white-collar sector; the signal-to-noise ratios here are 0.45 and 0.55 for in-school and out-of-school, respectively.

[46]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).

6.4 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 E.5). The marginal utility of consumption given this risk aversion parameter is large and positive. 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 high schooling ability and preferences are associated with higher utility in the schooling options, though the estimates are noisy.

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 presented 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 $\hat{\Delta}$).[47] 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 B8). 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

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

utility estimates and the reduced-form CCPs to represent the future value term.[48] 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.[49] We perform this forward simulation 10 times for each individual in the estimation sample.

Tables B16 and B17 show how the model matches the choice probabilities in the data for non-graduates and 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.

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 forward simulating 10 times, for each individual in the sample, the outcomes and sequences of choices.

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, science graduates have similar non-science posteriors to those of non-science graduates.

The second key pattern from Table 11 is how little sorting there is across educational paths on the basis of work abilities.[50] Some exceptions are that those who drop out of

[48]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.

[49]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.

[50]The substantial unresolved uncertainty for college graduates can be seen in Appendix Table B19 which

college late have received signals that they are stronger in the blue-collar sector than in the white-collar sector, and those who have worked in both sectors while in college sort into the sciences if the signals were positive and in the non-sciences otherwise. But overall, those who obtain college degrees have beliefs about their work abilities that are not especially different from those who do not obtain a college degree.

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 three 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 search frictions, meaning that all individuals have the option of working in the white-collar sector in every period. The final counterfactual adds to the first by relaxing credit constraints. In particular, we set each person's in-college non-wage consumption to the 75th percentile for all individuals and remove all loans.

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 and third counterfactual, individuals remain uncertain regarding having the option to work in the white-collar sector. In addition, there are three other sources of uncertainty: the probability of graduating from a four-year college conditional on attendance, aggregate labor market shocks common to both sectors, and individual preference shocks.[51] We then solve the model backwards to get the counterfactual choice probabilities, and then forward simulate to obtain the distribution of choices and the average abilities across different choice paths.

shows the equivalent of Table 11 except the entries are the posterior ability variances.

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

7.1 Information and educational choices

Table 12 reports the college completion status frequencies in the baseline and in the three counterfactual scenarios. Comparing column 2 (full information) to column 1 (baseline) shows only a slight increase in the college graduation rate from providing information. However, there are efficiency gains as there is a 3.4 percentage point increase (13%) in the number of individuals who never went to college. Dropout rates also decline, as well as the share of individuals who switch majors and the share who stop out and then dropout. Time to degree also falls by 0.3 years.

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 the number of individuals graduating slightly increasing but the number of individuals never attending college increasing.

The other significant difference between the full information counterfactual and the baseline is the shift in majors from non-science to science. This is the result of the 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 there.

Comparing the full-information counterfactual to the one where search frictions are also removed (column 3) or when credit constraints are relaxed (column 4), we see increases in college graduation rates. Considering the removal of search frictions (column 3), increases are the product of two countervailing forces. First, when search frictions are present, college graduates see higher probabilities of having a white-collar option. 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. This latter effect dominates, as evidence by the net increase in college graduation rates. Considering the removal of credit constraints (column 4), the rise in graduation rates is the result of higher in-school consumption coupled with no loan repayments in the future.

The patterns in Table 12 mask significant heterogeneity in the effects of information. Table 13 shows counterfactual results by whether the individual's family is above or below the median income in the data.[52] Comparing full information (column 2) to the baseline (column 1) shows that information increases the graduation rates of low-income individuals by almost 5 percentage points (a 31% increase). This is counterbalanced by a 4.6 percentage point drop in graduation rates of high-income individuals. All told, full information cuts the gap in graduation rates between low and high-income individuals by more than half.

The mechanism for this convergence is 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 actually are 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 revealing of information in the baseline leads to higher graduation rates for high-income individuals because 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 remaining to do.

Counterfactual 3 (column 4) shows that, conditional on full information, relaxing credit constraints does not close the gap in graduation rates between high and low-income individuals. While removing credit constraints increases college graduation rates for low-income individuals, it does so even more for high-income individuals as they are more on the margin of graduating from college.

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 counterfactual simulations (counterfactual 1).[53] As in the baseline, Table 14 shows substantial sorting on the basis of college ability. However, the correlations are now much stronger.

But in contrast to Table 11, there is now 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. Those who never attend college have above-average blue-collar abilities, but below-average white-collar abilities.

[52] We refer to those with families above and below median income as high- and low-income, respectively.

[53] Results for counterfactuals 2 and 3 are in Tables B20 and B21, respectively.

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 would 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 magnifies sorting. Whereas white-collar science graduates in the baseline earned 42% more than blue-collar non-graduates, the gap increases to 93% in the counterfactual. The source of 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. Information provision, therefore, results in stronger matching of abilities and occupations in the labor market. Indeed, with full information, more individuals 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 77% in the counterfactual.[54]

Relative to the full information counterfactual, additionally removing 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 roughly doubles for both science and non-science graduates relative to the

[54]These figures are calculated as 100 minus the entries in the “Remainder” row of each column, since each of the other rows corresponds to full-time work.

baseline. But the overall premium more than doubles because of the shift from non-science to science majors among college graduates. Since science graduates make more than their non-science counterparts, the overall college premium increases by more than the individual parts. The premium for working in the white-collar sector also increases substantially. This is in part because white-collar ability has a higher variance than blue-collar ability, so its sorting effects get magnified. But it is also due to a greater share of those working in the white collar sector having a college degree in the counterfactual relative to the baseline.

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 get the change in the white-collar premium due to ability, we first take 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. Changes in work abilities account for between 71% and 93% of each of the premia; the remainder is due to changes in the composition of who enters each sector-education combination, changes in sector-specific work experience, and, in the case of non-graduates, changes in years of college.

To conclude, our results show that informational frictions play an important role in shaping labor market outcomes. Providing full information to students about their own abilities by the end of high school substantially changes the income composition 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-collar and blue-collar workers.

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. At the end of each year, individuals update their ability and productivity beliefs through college grades and wages. 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 the 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 the skilled and unskilled labor market. In contrast, grades earned in college turn out to reveal little information about future labor market performance. To the extent that part of the mission of higher education is to help prepare students for the labor market, this finding suggests that there is room for improvement in the screening mechanisms in place in college.

Finally, simulations conducted under a counterfactual full information scenario indicate little change in graduation rates in the aggregate. However, this mask two key features. First, it significantly closes the college graduation rate gap between children from low and high income households. Low income households have expectations that college is likely to be a poor match. Revealing information about abilities up front results in some low income individuals being induced into college. Second, information amplifies sorting into college based on (previously unknown) academic abilities but also sorting on white collar abilities. 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.

References

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

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 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.

———. 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.

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.

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

Clark, Brian, Clément Joubert, and Arnaud Maurel. 2017. "The career prospects of overeducated Americans." *IZA Journal of Labor Economics* 6 (3):1–29.

De Groote, Olivier. 2022. "A Dynamic Model of Effort Choice in High School." Working paper.

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.

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.

Hansen, Karsten T., James J. Heckman, and Kathleen J. Mullen. 2004. "The Effect of Schooling and Ability on Achievement Test Scores." *Journal of Econometrics* 121:39–98.

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.

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. 2003. “Working during School and Academic Performance.” *Journal of Labor Economics* 21:473–491.

———. 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 college occupation and college completion status

	Non-graduates		Graduates	
	Blue Collar	White Collar	Blue Collar	White Collar
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)
HS GPA	-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	$\lvert t\text{-stat}\rvert$
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	$\lvert t\text{-stat}\rvert$
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	$\lvert t\text{-stat}\rvert$
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	$\lvert t\text{-stat}\rvert$
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	1350	3.567
Return to school	2.190	0.432	113	
Total	2.350	0.498	1463	

(b) Log wage *residuals*

	Mean residual	Std Dev	N	$\|t\text{-stat}\|$
Stay in work	0.068	0.477	1350	2.001
Return to school	-0.025	0.413	113	
Total	0.060	0.474	1457	

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	-0.120	(0.089)	-0.325	(0.065)	-0.218	(0.062)
Hispanic	-0.044	(0.095)	-0.067	(0.075)	-0.074	(0.068)
Parent graduated college	0.090	(0.063)	0.111	(0.050)	0.015	(0.062)
HS Grades (z-score)	0.202	(0.038)	0.200	(0.027)	0.159	(0.026)
Work full-time	-0.285	(0.068)	-0.146	(0.041)	-0.040	(0.057)
Work part-time	-0.046	(0.053)	0.032	(0.036)	0.056	(0.053)
Year 2 or higher in college					0.182	(0.041)
Unobserved type						
(H, H, H)	0.199	(0.106)	0.078	(0.075)	-0.058	(0.085)
(H, H, L)	0.194	(0.097)	0.080	(0.067)	0.014	(0.071)
(H, L, H)	0.131	(0.114)	0.038	(0.080)	-0.048	(0.086)
(H, L, L)	0.162	(0.096)	0.058	(0.068)	-0.014	(0.075)
(L, H, H)	0.041	(0.115)	0.050	(0.077)	-0.028	(0.082)
(L, H, L)	0.079	(0.083)	0.003	(0.066)	-0.047	(0.079)
(L, L, H)	-0.018	(0.103)	0.021	(0.079)	-0.043	(0.079)
λ_0 (ability index intercept)	0.394	(0.161)	0.563	(0.087)	0.000	(—)
λ_1 (ability index loading)	0.890	(0.062)	0.821	(0.036)	1.000	(—)
Mean of dependent variable	2.612		2.620		2.351	
Person-year obs.	881		1,639		1,272	

Notes: Bootstrap standard errors in parentheses. Reference categories for multinomial variables are as follows: "White" for race/ethnicity, "Not working while in school" for work intensity, and "(L, L, L)" for unobserved type. Coefficients for birth year and age dummies are omitted.

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

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

	White Collar		Blue Collar	
	Coeff.	Std. Error	Coeff.	Std. Error
Black	-0.098	(0.024)	-0.110	(0.018)
Hispanic	0.025	(0.027)	0.004	(0.018)
Parent graduated college	0.030	(0.021)	0.017	(0.015)
HS Grades (z-score)	-0.034	(0.014)	-0.016	(0.008)
Work experience (any sector)	0.039	(0.003)	0.041	(0.001)
Work experience (white collar sector)	0.017	(0.003)	0.010	(0.003)
Years of college completed	0.049	(0.010)	0.027	(0.004)
College graduate (any major)	0.074	(0.027)	0.052	(0.017)
College graduate (science major)	0.147	(0.041)	0.099	(0.027)
Work part-time	-0.032	(0.017)	-0.064	(0.008)
Unobserved type				
(H, H, H)	0.109	(0.021)	0.039	(0.020)
(H, H, L)	0.081	(0.019)	0.032	(0.018)
(H, L, H)	0.193	(0.023)	0.066	(0.020)
(H, L, L)	0.133	(0.020)	0.046	(0.018)
(L, H, H)	0.103	(0.022)	0.066	(0.018)
(L, H, L)	0.075	(0.022)	-0.012	(0.019)
(L, L, H)	0.119	(0.023)	0.075	(0.019)
λ_0 (in-school work index intercept)	0.750	(0.096)	0.659	(0.131)
λ_1 (in-school work index loading)	0.626	(0.037)	0.662	(0.055)
Mean of dependent variable	2.669		2.332	
Person-year obs.	2,373		12,755	

Notes: Bootstrap standard errors in parentheses. Reference categories for multinomial variables are as follows: "White" for race/ethnicity, "Work full-time" for work intensity, and "(L, L, L)" for unobserved type. Coefficients for birth year, age, and year dummies are omitted.

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

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	0.673	1.000			
	(0.023)	(—)			
Science	0.054	0.086	1.000		
	(0.056)	(0.053)	(—)		
Non-Science	0.088	-0.059	0.778	1.000	
	(0.046)	(0.042)	(0.058)	(—)	
2-year	0.063	0.095	0.280	0.439	1.000
	(0.073)	(0.056)	(0.132)	(0.097)	(—)
Variance of Unobserved Abilities	0.167	0.084	0.413	0.402	0.260
	(0.008)	(0.003)	(0.045)	(0.029)	(0.030)
Variance of Raw Outcomes	0.360	0.272	0.961	0.902	1.064

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 individuals and at most 1,810 individuals.

Table 9: Idiosyncratic Variances

Employment			Schooling			
Work Type	White Collar	Blue Collar	Schooling Period	Science	Non-Science	2-year
In-school	0.206	0.120	1	0.500	0.573	0.772
	(0.011)	(0.004)		(0.047)	(0.041)	(0.055)
Out-of-school	0.136	0.143	2	0.332	0.364	0.626
	(0.002)	(0.001)		(0.041)	(0.028)	(0.054)
			3	0.489	0.480	0.740
				(0.046)	(0.029)	(0.042)
			4	0.631	0.367	
				(0.057)	(0.025)	
			5+	0.572	0.535	
				(0.045)	(0.028)	

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

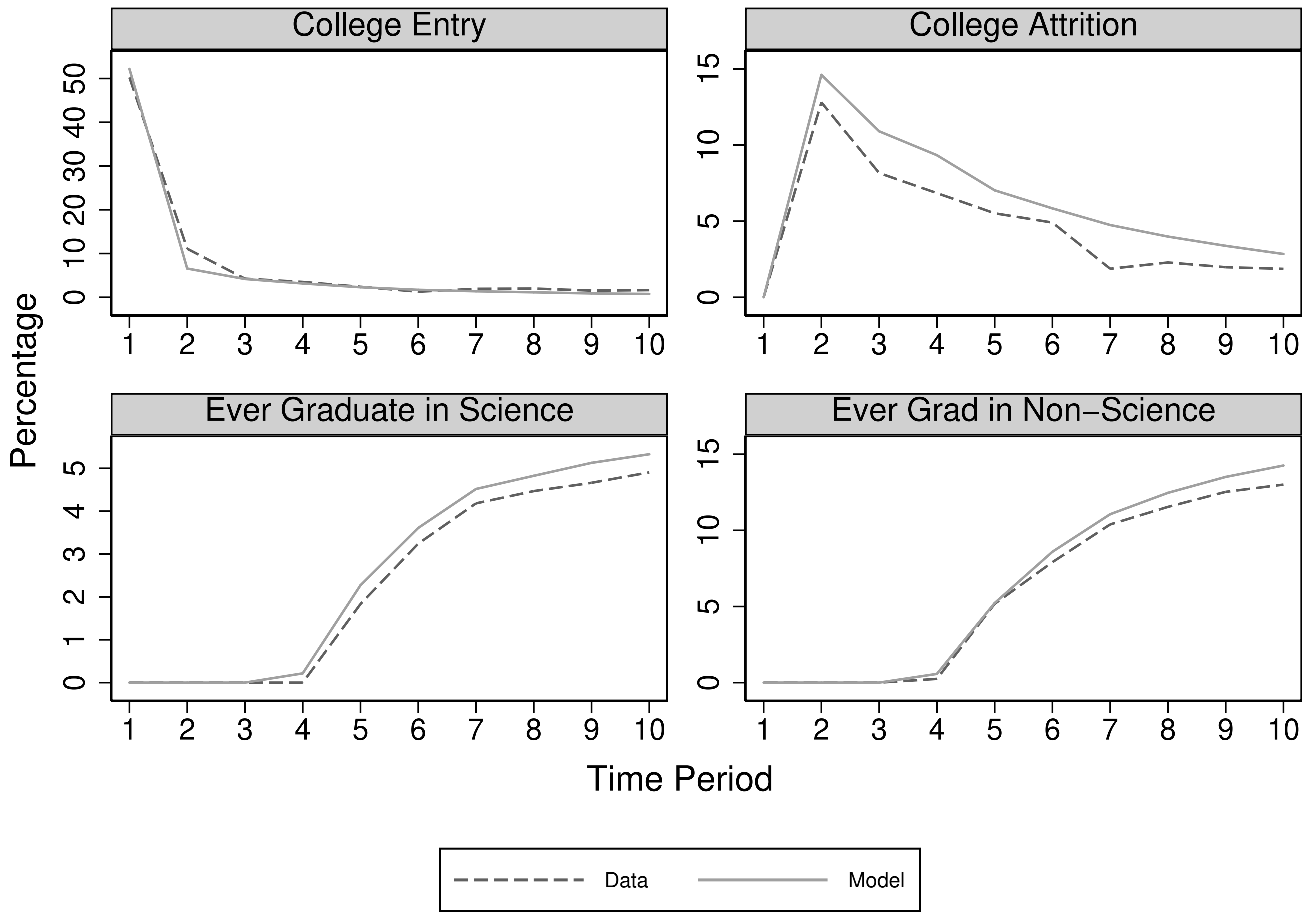
Variable	2-year	4-year Sci	4-year Non-Sci	Work PT	Work FT	White Collar
Black	-0.085	-0.124	-0.115	0.044	0.037	0.028
	(0.057)	(0.091)	(0.100)	(0.039)	(0.032)	(0.037)
Hispanic	0.044	-0.093	-0.041	-0.071	-0.018	0.030
	(0.050)	(0.108)	(0.114)	(0.041)	(0.028)	(0.040)
HS Grades (z-score)	-0.017	0.261	0.221	0.000	0.020	0.050
	(0.025)	(0.045)	(0.051)	(0.018)	(0.014)	(0.017)
Parent graduated college	0.036	0.051	0.063	0.029	-0.029	0.184
	(0.054)	(0.088)	(0.097)	(0.034)	(0.030)	(0.032)
Family Income ($10,000)	0.004	0.045	0.058	-0.014	-0.007	-0.006
	(0.007)	(0.009)	(0.009)	(0.004)	(0.003)	(0.003)
Prior academic ability	0.306	1.481	1.625			
	(0.126)	(0.221)	(0.163)			
Previous high school	0.985	2.650	1.730	1.185	0.907	-0.599
	(0.099)	(0.168)	(0.110)	(0.082)	(0.081)	(0.178)
Previous 2-year college	2.413	1.046	0.721	0.043	0.301	0.110
	(0.084)	(0.164)	(0.111)	(0.095)	(0.090)	(0.134)
Previous 4-year science	0.826	4.555	1.971	0.665	0.392	-0.165
	(0.178)	(0.138)	(0.131)	(0.104)	(0.100)	(0.123)
Previous 4-year non-science	0.293	1.971	3.581	0.628	0.630	0.080
	(0.154)	(0.163)	(0.097)	(0.089)	(0.080)	(0.092)
Previous work part-time	0.001	0.471	0.415	2.222	1.388	-0.981
	(0.095)	(0.121)	(0.099)	(0.049)	(0.054)	(0.089)
Previous work full-time	0.069	0.151	0.564	0.986	2.316	-0.925
	(0.102)	(0.144)	(0.104)	(0.062)	(0.043)	(0.087)
Previous work white-collar	-0.035	-0.362	-0.112	-1.325	-1.491	2.721
	(0.143)	(0.170)	(0.134)	(0.091)	(0.084)	(0.151)
College graduate				-0.232	-0.116	0.531
				(0.103)	(0.115)	(0.088)
Currently work white-collar				-0.060		
				(0.110)		
Currently work part-time	0.682	-0.618	-0.508			
	(0.124)	(0.159)	(0.130)			
Currently work full-time	-0.698	-1.408	-1.837			
	(0.142)	(0.190)	(0.162)			
Unobserved type						
(H, H, H)	0.146	0.309	0.107	-0.104	-0.082	0.130
	(0.096)	(0.215)	(0.134)	(0.088)	(0.112)	(0.119)
(H, H, L)	0.112	0.293	0.134	-0.016	-0.016	0.034
	(0.088)	(0.193)	(0.127)	(0.073)	(0.098)	(0.115)
(H, L, H)	0.196	0.383	0.131	-0.210	-0.051	0.090
	(0.100)	(0.221)	(0.141)	(0.093)	(0.118)	(0.135)
(H, L, L)	0.107	0.310	0.154	-0.050	-0.070	0.084
	(0.090)	(0.196)	(0.126)	(0.078)	(0.107)	(0.120)
(L, H, H)	0.078	0.235	0.086	-0.114	-0.058	0.110
	(0.094)	(0.202)	(0.131)	(0.075)	(0.090)	(0.123)
(L, H, L)	0.065	0.334	0.030	-0.057	0.008	0.155
	(0.095)	(0.204)	(0.134)	(0.073)	(0.097)	(0.123)
(L, L, H)	0.069	0.214	0.058	-0.154	-0.051	0.006
	(0.095)	(0.195)	(0.132)	(0.074)	(0.099)	(0.127)
$\mathbb{E}[u(\text{consumption})] \div 1,000$	3.117	(0.459)				
Constant Relative Risk Aversion parameter (θ)	0.4					
Log likelihood	-26,351					
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 0.37. Missing GPAs are estimated to be $\leq$ 2.5 w.p. 0.66, 2.5–3.0 w.p. 0.12, 3.0–3.6 w.p. 0.13, and 3.6–4.0 w.p. 0.09.

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, L, L)" for unobserved type. We omit the following coefficients: 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).

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

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 100 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						
$x = 0$	0.03	0.05	0.50	0.40	0.15	1.33
$x > 0$, white collar only	-0.03	0.00	0.48	0.36	0.12	0.47
$x > 0$, blue collar only	0.05	0.08	0.49	0.39	0.17	3.55
$x > 0$, mixture	0.10	0.10	0.47	0.35	0.12	1.02
Continuous enrollment, graduate in non-science with x years of in-school work experience						
$x = 0$	0.02	-0.03	0.28	0.37	0.17	2.99
$x > 0$, white collar only	0.05	-0.03	0.31	0.41	0.18	0.98
$x > 0$, blue collar only	-0.01	-0.08	0.30	0.39	0.18	8.99
$x > 0$, mixture	-0.05	-0.12	0.25	0.35	0.19	1.87
Stop out (SO)						
SO, graduate in science	-0.02	0.03	0.21	0.13	0.02	0.95
SO, graduate in non-science	-0.06	-0.14	0.22	0.33	0.19	3.13
SO then DO, start in 2yr	-0.03	-0.03	-0.11	-0.11	-0.01	5.12
SO then DO, start in science	-0.00	0.02	-0.42	-0.40	-0.21	1.67
SO then DO, start in non-science	-0.01	0.05	-0.29	-0.37	-0.18	2.67
Truncated	-0.06	-0.09	0.01	0.04	0.08	5.42
Drop out (DO) after x years of school						
$x = 1$	-0.00	0.01	-0.14	-0.16	-0.13	16.57
$x = 2$	-0.02	-0.01	-0.20	-0.23	-0.16	8.08
$x = 3$	-0.00	0.03	-0.24	-0.27	-0.12	4.37
$x = 4$	0.00	0.07	-0.29	-0.33	-0.12	2.23
$x \geq 5$	0.03	0.10	-0.28	-0.29	0.01	2.25
Never attended college						
Never attend college	0.01	0.02	0.00	-0.00	0.00	26.35

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	Full info. & no search frictions	Full info. & reduced credit constraints
Continuous completion (CC), Science	6.37	9.05	9.89	10.67
Continuous completion (CC), Non-Science	14.83	11.86	11.99	13.33
Stop out (SO) but graduated Science	0.95	1.62	1.74	1.97
Stop out (SO) but graduated Non-Science	3.13	2.92	3.60	4.10
Stop out (SO) then drop out	9.45	8.60	8.57	10.05
Truncated	5.42	4.25	4.46	6.40
Drop out (DO)	33.50	31.91	30.70	29.68
Never went to college	26.35	29.79	29.06	23.79
Graduate from 4-year college	25.27	25.45	27.22	30.08
Ever switch major	23.95	20.01	21.59	21.77
Time to degree (years)	5.12	4.82	4.88	4.93

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. "Full info. & reduced credit constraints" maintains full information, removes college loans, and sets in-college non-wage consumption to its 75th percentile for all individuals.

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	Full info. & no search frictions	Full info. & reduced credit constraints
Panel A: Above-median family income in high school				
Dropout	30.57	30.50	29.16	28.22
Never went to college	19.72	27.24	25.83	20.19
Graduate from 4-year college	34.50	29.87	32.03	35.98
Ever Switch Major	25.96	20.96	22.70	22.33
Panel B: Below-median family income in high school				
Dropout	36.43	33.32	32.24	31.15
Never went to college	32.98	32.33	32.29	27.38
Graduate from 4-year college	16.05	21.03	22.41	24.17
Ever Switch Major	20.97	18.87	20.24	21.08

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. "Full info. & reduced credit constraints" maintains full information, removes college loans, and sets in-college non-wage consumption to its 75th percentile for all individuals.

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						
$x = 0$	0.16	0.02	1.27	0.84	0.20	1.16
$x > 0$, white collar only	1.32	0.68	1.00	0.77	0.20	0.60
$x > 0$, blue collar only	0.44	0.34	1.25	0.79	0.23	5.92
$x > 0$, mixture	1.36	0.73	0.95	0.65	0.14	1.38
Continuous enrollment, graduate in non-science with x years of in-school work experience						
$x = 0$	-0.16	-0.59	0.68	1.18	0.54	1.94
$x > 0$, white collar only	0.99	0.01	0.37	0.93	0.44	0.53
$x > 0$, blue collar only	0.25	-0.13	0.63	1.10	0.44	8.25
$x > 0$, mixture	0.90	0.17	0.37	0.92	0.42	1.14
Stop out (SO)						
SO, graduate in science	0.36	0.15	1.03	0.76	0.20	1.62
SO, graduate in non-science	0.01	-0.30	0.63	1.01	0.48	2.92
SO then DO, start in 2yr	-0.20	-0.12	-0.37	-0.35	-0.01	3.27
SO then DO, start in science	-0.31	-0.17	0.35	0.10	-0.07	1.80
SO then DO, start in non-science	-0.24	-0.22	0.12	0.35	0.26	3.54
Truncated	-0.14	-0.18	0.06	0.15	0.10	4.25
Drop out (DO) after x years of school						
$x = 1$	-0.12	0.01	-0.40	-0.43	-0.13	15.77
$x = 2$	-0.17	-0.07	-0.16	-0.15	-0.02	7.61
$x = 3$	-0.25	-0.16	0.20	0.24	0.13	5.12
$x = 4$	-0.24	-0.14	0.31	0.31	0.19	2.47
$x \geq 5$	-0.15	-0.22	0.40	0.37	0.14	0.95
Never attended college						
Never attend college	-0.07	0.08	-0.56	-0.64	-0.30	29.79

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

	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	0.42	0.95	0.90	3.53	5.53	6.43
White collar, Non-Science graduate	0.25	0.68	0.61	7.97	5.76	6.54
White collar, Non-graduate	0.14	0.48	0.29	5.68	1.41	5.69
Blue collar, Science graduate	0.17	0.26	0.29	2.33	4.15	4.05
Blue collar, Non-Science graduate	0.05	0.02	0.03	6.35	6.27	6.10
Blue collar, Non-graduate	0.00	0.02	0.03	41.91	54.43	48.56
Remainder	—	—	—	32.23	22.45	22.62

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	Baseline	Counterfactual	No Frictions Cfl
College wage premium	0.19	0.44	0.43	—	0.20	0.17
Science college premium	0.30	0.62	0.61	—	0.27	0.24
Non-science college premium	0.15	0.30	0.27	—	0.14	0.11
White-collar wage premium	0.23	0.74	0.57	—	0.38	0.29

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

Appendices

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, occupation, 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 blue- or white-collar occupations. Using data from the March Current Population Survey (CPS) 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 occupation 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 Section 5.6).

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 5.6), 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 5.6.

B Supporting Tables

Table B5: 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 E for complete details
1	Measurement system	- measurements, - demographics	- Unobserved type probabilities	- Find global optimum - See Appendix C for complete details
2	Missing data	- unobserved type probabilities, - demographics, - choices	- Missing outcome (major or GPA) type probabilities	- See Appendix D for complete details
3	Learning parameters	- outcomes (wages, GPA), - unobserved type probabilities, - missing outcome type probabilities, - demographics, - other state variables (experiences, educational degrees)	- learning parameter estimates	
4	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 E and H for complete details
5	Compute future value (FV) terms	All data, all parameter estimates	Future value terms	- FV formulas according to (24) - See Appendix F for complete details
6	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 5

Table B6: Variables Included in Each Component of the Model, by Estimation Stage

Variable	Stage 1	Stage 2	Stage 3		Stage 4						Stage 6
	Measure-ment System	Reduced-form Choice	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)						✓	✓				
EFC (continuous)		✓									
Aggregate labor market											
Year dummies		✓	✓								
Individual characteristics											
Age dummies (≤18, 19, 20, ≥21)			✓	✓							
Age (linear)		✓			✓				✓		
Average cumulative wage & GPA		✓									
Expected prior work ability			✓								
Expected prior acad. ability								✓			✓
Education											
Bachelor's degree		✓	✓					✓	✓		✓
Science major		✓									
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	✓	✓	✓	✓				✓	✓	✓	✓
Number of parameters	221	967	170		22	114	114	240	10	33	164

Notes: Grand total of all parameters is 2,055. In the Stage 4 columns for transfers, grants and loans, we combine the parameter counts in both 2-year and 4-year models. We also combine the parameter counts in the logit models of having positive amounts and the linear or log-linear models of total amounts conditional on amounts being positive. In the Stage 4 column for the static choice model, we include the two parameters of the AR(1) model for wages.

Table B7: Unobserved Type Coefficient Signs and Significance Across All Equations of the Model

	Unobserved Type Identity							
	High Schooling Ability				Low Schooling Ability			
	High Schooling Ability & Prefs.		Low Schooling Ability & Prefs.		High Schooling Ability & Prefs.		Low Schooling Ability & Prefs.	
	High Work Ab & Pr.	Low Work Ab & Pr.	High Work Ab & Pr.	Low Work Ab & Pr.	High Work Ab & Pr.	Low Work Ab & Pr.	High Work Ab & Pr.	Low Work Ab & Pr.
Model Equation	(H, H, H)	(H, H, L)	(H, L, H)	(H, L, L)	(L, H, H)	(L, H, L)	(L, L, H)	(L, L, L)
Panel A: Measurement System								
ASVAB Arithmetic Reasoning	(Ref.)	(Ref.)	(Ref.)	(Ref.)	-*	-*	-*	-*
ASVAB Coding Speed	(Ref.)	(Ref.)	(Ref.)	(Ref.)	-*	-*	-*	-*
ASVAB Mathematical Knowledge	(Ref.)	(Ref.)	(Ref.)	(Ref.)	-*	-*	-*	-*
ASVAB Numerical Operations	(Ref.)	(Ref.)	(Ref.)	(Ref.)	-*	-*	-*	-*
ASVAB Paragraph Comprehension	(Ref.)	(Ref.)	(Ref.)	(Ref.)	-*	-*	-*	-*
ASVAB Word Knowledge	(Ref.)	(Ref.)	(Ref.)	(Ref.)	-*	-*	-*	-*
SAT Math	(Ref.)	(Ref.)	(Ref.)	(Ref.)	-*	-*	-*	-*
SAT Verbal	(Ref.)	(Ref.)	(Ref.)	(Ref.)	-*	-*	-*	-*
No. AP Exams	(Ref.)	(Ref.)	-*	-*	(Ref.)	(Ref.)	-*	-*
Late for Classes	(Ref.)	(Ref.)	+*	+*	(Ref.)	(Ref.)	+*	+*
Regularly Break Rules	(Ref.)	(Ref.)	+*	+*	(Ref.)	(Ref.)	+*	+*
Took Extra Classes/Lessons	(Ref.)	(Ref.)	-	-	(Ref.)	(Ref.)	-	-
Ever Took Classes During School Break	(Ref.)	(Ref.)	+*	+*	(Ref.)	(Ref.)	+*	+*
Reason Took Classes During Break	(Ref.)	(Ref.)	-*	-*	(Ref.)	(Ref.)	-*	-*
Have High Standards at Work	+*	(Ref.)	+*	(Ref.)	+*	(Ref.)	+*	(Ref.)
Make Every Effort to Do What is Expected	+*	(Ref.)	+*	(Ref.)	+*	(Ref.)	+*	(Ref.)
Percent Chance Work at Age 30	+*	(Ref.)	+*	(Ref.)	+*	(Ref.)	+*	(Ref.)
Parental Assessment of Age-30 Work Pr.	+	(Ref.)	+	(Ref.)	+	(Ref.)	+	(Ref.)
Panel B: Learning Outcomes								
White Collar Log Wages	+*	+*	+*	+*	+*	+*	+*	(Ref.)
Blue Collar Log Wages	+	+	+*	+*	+*	-	+*	(Ref.)
4-year Science Grades	+	+*	+	+	+	+	-	(Ref.)
4-year Non-Science Grades	+	+	+	+	+	+	+	(Ref.)
2-year Grades	-	+	-	-	-	-	-	(Ref.)
Panel C: White Collar Offer Arrival								
Receive White Collar Offer	+*	+*	+*	+*	-	+	+	(Ref.)
Panel D: Logit for Graduation in $t+1$								
Graduate in $t+1$	+*	+*	+	+	+	+	+	(Ref.)
Panel E: Flow Utilities								
2-year College	+	+	+*	+	+	+	+	(Ref.)
4-year Science	+	+	+	+	+	+	+	(Ref.)
4-year Non-Science	+	+	+	+	+	+	+	(Ref.)
Work Part-Time	-	-	-*	-	-	-	-*	(Ref.)
Work Full-Time	-	-	-	-	-	+	-	(Ref.)
Work White Collar	+	+	+	+	+	+	+	(Ref.)
Home Production	(Ref.)	(Ref.)	(Ref.)	(Ref.)	(Ref.)	(Ref.)	(Ref.)	(Ref.)

Notes: * indicates statistical significance at the 5% level. "(Ref.)" indicates that the type served as the reference category for that outcome. See Tables 6 (grades), 7 (log wages), 10 (flow utilities), B9–B11 (measurement system), B13 (graduation), and B14 (offer arrival), for exact parameter estimates.

Table B8: Estimates of Probability Mass for Each Unobserved Type

Type Identity	Mass Probability	Std. Error
(H, H, H)	0.154	(0.006)
(H, H, L)	0.210	(0.006)
(H, L, H)	0.051	(0.004)
(H, L, L)	0.164	(0.005)
(L, H, H)	0.097	(0.004)
(L, H, L)	0.041	(0.004)
(L, L, H)	0.109	(0.005)
(L, L, L)	0.174	(0.006)

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 ability and preferences }. e.g. "Unobserved type (H, L, H)" corresponds to a worker with high schooling ability, low schooling preferences, and high work ability and preferences. Labels are identified through the measurement system detailed in Appendix C.

Table B9: 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	0.890	0.343	0.709	0.543	0.647	0.623	0.464	0.254
	(0.036)	(0.050)	(0.037)	(0.050)	(0.038)	(0.042)	(0.048)	(0.059)
Black	-0.663	-0.326	-0.509	-0.210	-0.507	-0.547	-0.668	-0.243
	(0.031)	(0.051)	(0.031)	(0.044)	(0.032)	(0.036)	(0.042)	(0.050)
Hispanic	-0.354	-0.122	-0.232	-0.157	-0.221	-0.359	-0.237	0.032
	(0.037)	(0.053)	(0.034)	(0.045)	(0.038)	(0.038)	(0.050)	(0.053)
Born in 1980	-0.090	-0.033	-0.023	-0.031	-0.087	-0.014	-0.255	-0.222
	(0.042)	(0.056)	(0.037)	(0.055)	(0.044)	(0.046)	(0.054)	(0.064)
Born in 1981	-0.018	0.019	0.015	-0.002	0.015	0.010	-0.153	-0.211
	(0.036)	(0.053)	(0.039)	(0.051)	(0.037)	(0.042)	(0.055)	(0.062)
Born in 1982	-0.056	-0.103	-0.048	-0.059	-0.109	-0.021	-0.129	-0.294
	(0.039)	(0.056)	(0.037)	(0.051)	(0.042)	(0.040)	(0.055)	(0.067)
Born in 1983	-0.048	-0.000	-0.010	-0.042	-0.027	0.038	-0.067	-0.196
	(0.039)	(0.052)	(0.039)	(0.051)	(0.042)	(0.045)	(0.051)	(0.063)
Parent graduated college	0.335	0.269	0.365	0.188	0.334	0.273	0.299	0.180
	(0.030)	(0.041)	(0.026)	(0.037)	(0.030)	(0.030)	(0.038)	(0.048)
Family Income ($10,000)	0.018	0.009	0.017	0.011	0.017	0.023	0.020	0.024
	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.005)
Unobserved type (L, L, H)	-1.356	-0.857	-1.320	-1.076	-1.258	-0.980	-0.948	-0.757
	(0.026)	(0.033)	(0.024)	(0.033)	(0.027)	(0.028)	(0.033)	(0.037)
Unobserved type (L, L, L)	-1.356	-0.857	-1.320	-1.076	-1.258	-0.980	-0.948	-0.757
	(0.026)	(0.033)	(0.024)	(0.033)	(0.027)	(0.028)	(0.033)	(0.037)
Unobserved type (L, H, H)	-1.356	-0.857	-1.320	-1.076	-1.258	-0.980	-0.948	-0.757
	(0.026)	(0.033)	(0.024)	(0.033)	(0.027)	(0.028)	(0.033)	(0.037)
Unobserved type (L, H, L)	-1.356	-0.857	-1.320	-1.076	-1.258	-0.980	-0.948	-0.757
	(0.026)	(0.033)	(0.024)	(0.033)	(0.027)	(0.028)	(0.033)	(0.037)
Unobserved type (H, L, H)	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
	(—)	(—)	(—)	(—)	(—)	(—)	(—)	(—)
Unobserved type (H, L, L)	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
	(—)	(—)	(—)	(—)	(—)	(—)	(—)	(—)
Unobserved type (H, H, H)	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
	(—)	(—)	(—)	(—)	(—)	(—)	(—)	(—)
Unobserved type (H, H, L)	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
	(—)	(—)	(—)	(—)	(—)	(—)	(—)	(—)
Std. Dev. of noise	0.614	0.871	0.591	0.776	0.647	0.655	0.802	0.946
	(0.009)	(0.013)	(0.009)	(0.011)	(0.009)	(0.010)	(0.012)	(0.014)
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.

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

Table B10: Measurement System Estimates for Schooling Ability & Preferences Measurements

Variable	No. of AP classes	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				-6.304	-3.364	2.469
				(0.538)	(0.206)	(0.115)
Black	-0.472	0.509	-0.444	1.512	0.622	-0.164
	(0.248)	(0.109)	(0.100)	(0.509)	(0.171)	(0.130)
Hispanic	0.031	0.632	-0.185	0.369	0.743	0.119
	(0.234)	(0.114)	(0.109)	(0.590)	(0.201)	(0.137)
Born in 1980	-0.542	1.477	0.220		1.201	-0.782
	(0.275)	(0.154)	(0.122)		(0.185)	(0.124)
Born in 1981	-0.229	1.025	0.121		0.738	-0.405
	(0.258)	(0.139)	(0.112)		(0.201)	(0.107)
Born in 1982	-0.035	0.606	0.385	-1.470	0.602	
	(0.226)	(0.157)	(0.108)	(0.554)	(0.204)	
Born in 1983	0.070	0.226	-0.038	-2.263		
	(0.226)	(0.168)	(0.115)	(0.562)		
Parent graduated college	0.837	0.049	-0.098	2.056	0.340	-0.133
	(0.166)	(0.107)	(0.076)	(0.526)	(0.144)	(0.112)
Family Income ($10,000)	0.022	0.010	-0.002	0.244	-0.012	0.024
	(0.015)	(0.011)	(0.009)	(0.049)	(0.018)	(0.012)
Unobserved type (L, L, H)	-1.187	1.524	1.443	-0.287	0.538	-2.742
	(0.110)	(0.062)	(0.048)	(0.273)	(0.093)	(0.047)
Unobserved type (L, L, L)	-1.187	1.524	1.443	-0.287	0.538	-2.742
	(0.110)	(0.062)	(0.048)	(0.273)	(0.093)	(0.047)
Unobserved type (L, H, H)	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
	(—)	(—)	(—)	(—)	(—)	(—)
Unobserved type (L, H, L)	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
	(—)	(—)	(—)	(—)	(—)	(—)
Unobserved type (H, L, H)	-1.187	1.524	1.443	-0.287	0.538	-2.742
	(0.110)	(0.062)	(0.048)	(0.273)	(0.093)	(0.047)
Unobserved type (H, L, L)	-1.187	1.524	1.443	-0.287	0.538	-2.742
	(0.110)	(0.062)	(0.048)	(0.273)	(0.093)	(0.047)
Unobserved type (H, H, H)	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
	(—)	(—)	(—)	(—)	(—)	(—)
Unobserved type (H, H, L)	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
	(—)	(—)	(—)	(—)	(—)	(—)
Cut point 1	2.239	2.439	-0.605			
	(0.218)	(0.163)	(0.118)			
Cut point 2	2.837	2.981	0.445			
	(0.229)	(0.165)	(0.115)			
Cut point 3		3.448	1.041			
		(0.167)	(0.118)			
Cut point 4		3.876	1.485			
		(0.167)	(0.117)			
Cut point 5		4.144	2.268			
		(0.169)	(0.124)			
Cut point 6		4.510	3.088			
		(0.177)	(0.132)			
Cut point 7		5.448				
		(0.196)				
Std. Dev. of noise				7.884		
				(0.253)		
Observations	2,310	2,303	2,088	1,386	1,141	151

Notes: Bootstrap standard errors are listed below each coefficient in parentheses. The first three 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."

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

Table B11: Measurement System Estimates for Working Ability & Preferences 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 working at age 30
Black	0.675	0.047	-0.236	-0.217
	(0.128)	(0.092)	(0.139)	(0.181)
Hispanic	0.001	0.229	-0.334	0.104
	(0.123)	(0.098)	(0.130)	(0.222)
Born in 1980	-0.086	-0.078	-0.070	0.040
	(0.132)	(0.114)	(0.135)	(0.196)
Born in 1981	0.177	0.370		
	(0.127)	(0.107)		
Born in 1982	-0.042	0.064		
	(0.137)	(0.115)		
Born in 1983	-0.124	0.111		
	(0.132)	(0.105)		
Parent graduated college	-0.003	-0.307	-0.085	-0.158
	(0.089)	(0.078)	(0.122)	(0.166)
Family Income ($10,000)	0.039	0.007	0.011	0.074
	(0.010)	(0.008)	(0.013)	(0.022)
Unobserved type (L, L, H)	3.201	5.499	0.474	0.202
	(0.088)	(0.066)	(0.099)	(0.136)
Unobserved type (L, L, L)	Ref.	Ref.	Ref.	Ref.
	(—)	(—)	(—)	(—)
Unobserved type (L, H, H)	3.201	5.499	0.474	0.202
	(0.088)	(0.066)	(0.099)	(0.136)
Unobserved type (L, H, L)	Ref.	Ref.	Ref.	Ref.
	(—)	(—)	(—)	(—)
Unobserved type (H, L, H)	3.201	5.499	0.474	0.202
	(0.088)	(0.066)	(0.099)	(0.136)
Unobserved type (H, L, L)	Ref.	Ref.	Ref.	Ref.
	(—)	(—)	(—)	(—)
Unobserved type (H, H, H)	3.201	5.499	0.474	0.202
	(0.088)	(0.066)	(0.099)	(0.136)
Unobserved type (H, H, L)	Ref.	Ref.	Ref.	Ref.
	(—)	(—)	(—)	(—)
Cut point 1	-4.767	-4.480	-2.272	-2.984
	(0.415)	(0.288)	(0.128)	(0.197)
Cut point 2	-4.006	-3.187	-1.178	-1.964
	(0.278)	(0.165)	(0.112)	(0.170)
Cut point 3	-2.860	-2.247		
	(0.178)	(0.134)		
Cut point 4	-1.966	-1.022		
	(0.150)	(0.103)		
Cut point 5	-0.406	0.692		
	(0.131)	(0.100)		
Cut point 6	1.841	4.696		
	(0.129)	(0.111)		
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%+

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

Table B12: Labor market shock forecasting estimates

Parameter	Estimate	Std. Error
Autocorrelation	0.494	(0.099)
Std. Dev. of shock	0.019	(0.002)
Observations	16	

Notes: Estimates of Equation (15). Bootstrap standard errors in parentheses. We estimate a single AR1 process for both labor market sectors.

Table B13: Estimates of Probability of Graduation

Variable	Coeff.	Std. Error
Constant	-3.269	(0.286)
Black	-0.656	(0.135)
Hispanic	-0.421	(0.158)
HS Grades (z-score)	0.229	(0.055)
Parent graduated college	-0.060	(0.112)
Family Income ($10,000)	0.017	(0.009)
College experience completion profiles:		
0 years of 2yr	0.487	(0.215)
2+ years of 2yr	0.639	(0.176)
2 years of 4yr	1.235	(0.217)
3 years of 4yr	1.627	(0.218)
4 years of 4yr	2.645	(0.271)
5 years of 4yr	0.570	(0.592)
6+ years of 4yr	0.597	(0.393)
2 years of 4yr and 0 years of 2yr	-2.816	(0.311)
4 years of 4yr and 0 years of 2yr	-0.608	(0.310)
5 years of 4yr and 0 years of 2yr	2.244	(0.677)
6+ years of 4yr and 0 years of 2yr	1.531	(0.433)
Science major	-0.495	(0.119)
Prior ability science × Science major	1.469	(0.223)
Prior ability non-sci. × Non-Sci. major	1.332	(0.145)
Work part-time	-0.117	(0.125)
Work full-time	0.271	(0.137)
Unobserved type (H, H, H)	0.393	(0.155)
Unobserved type (H, H, L)	0.399	(0.138)
Unobserved type (H, L, H)	0.175	(0.164)
Unobserved type (H, L, L)	0.272	(0.148)
Unobserved type (L, H, H)	0.083	(0.173)
Unobserved type (L, H, L)	0.015	(0.136)
Unobserved type (L, L, H)	0.020	(0.168)
Person-year observations	1,115	

Notes: Parameter estimates of Equation (16) which is 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, L, L)" for unobserved type.

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

Table B14: Estimates of White Collar Offer Arrival Parameters

Variable	Coeff.	Std. Error
Constant	-0.675	(0.245)
Age	-0.164	(0.015)
College graduate	1.772	(0.119)
Unobserved type (H, H, H)	0.596	(0.139)
Unobserved type (H, H, L)	0.562	(0.138)
Unobserved type (H, L, H)	0.377	(0.146)
Unobserved type (H, L, L)	0.492	(0.141)
Unobserved type (L, H, H)	-0.072	(0.128)
Unobserved type (L, H, L)	0.014	(0.120)
Unobserved type (L, L, H)	0.206	(0.131)
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, L, L)" for the unobserved type. We restrict the offer arrival probability to equal 1 for those who worked in the white-collar sector in the previous period.

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

Table B15: Parameter Estimates of Static Choice Model

Variable	2-year	4-year Sci	4-year Non-Sci	Work PT	Work FT	White Collar
Constant	-2.664	-4.704	-3.526	-2.158	-1.442	-2.106
	(0.126)	(0.185)	(0.130)	(0.108)	(0.105)	(0.182)
Black	-0.090	0.100	0.240	-0.049	-0.090	0.042
	(0.078)	(0.107)	(0.081)	(0.056)	(0.046)	(0.086)
Hispanic	0.101	-0.037	-0.064	-0.087	0.014	0.116
	(0.077)	(0.112)	(0.085)	(0.054)	(0.047)	(0.089)
HS Grades (z-score)	0.199	0.953	0.656	-0.034	-0.053	0.131
	(0.032)	(0.049)	(0.038)	(0.024)	(0.020)	(0.034)
Parent graduated college	0.432	0.850	0.789	-0.003	-0.273	0.508
	(0.063)	(0.085)	(0.064)	(0.049)	(0.043)	(0.078)
Born in 1980	0.082	0.177	0.196	0.177	0.210	0.409
	(0.090)	(0.140)	(0.102)	(0.075)	(0.062)	(0.094)
Born in 1981	-0.070	-0.052	0.080	0.172	0.146	0.334
	(0.094)	(0.138)	(0.097)	(0.077)	(0.057)	(0.098)
Born in 1982	-0.069	0.287	-0.045	0.163	0.077	0.329
	(0.093)	(0.129)	(0.100)	(0.067)	(0.056)	(0.109)
Born in 1983	-0.029	0.081	0.059	0.112	0.005	0.168
	(0.087)	(0.117)	(0.097)	(0.067)	(0.057)	(0.095)
Family Income ($10,000)	0.017	0.053	0.063	-0.028	-0.014	0.007
	(0.008)	(0.008)	(0.007)	(0.006)	(0.005)	(0.009)

continued

Variable	2-year	4-year Sci	4-year Non-Sci	Work PT	Work FT	White Collar
Age	-0.187	-0.266	-0.282	-0.134	-0.146	0.174
	(0.016)	(0.025)	(0.019)	(0.012)	(0.011)	(0.020)
Age squared	0.006	0.007	0.005	0.001	-0.004	-0.004
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Experience	0.022	-0.050	-0.005	0.302	0.388	-0.043
	(0.034)	(0.053)	(0.043)	(0.022)	(0.016)	(0.027)
Experience squared	-0.006	0.010	-0.001	-0.019	-0.012	0.004
	(0.003)	(0.004)	(0.003)	(0.001)	(0.001)	(0.002)
Years of college	-0.031	0.668	0.488	0.275	0.160	0.359
	(0.055)	(0.069)	(0.045)	(0.034)	(0.028)	(0.037)
Years of college squared	0.018	-0.024	-0.011	-0.019	0.003	-0.036
	(0.007)	(0.007)	(0.005)	(0.003)	(0.003)	(0.004)
Prior academic ability	1.439	2.201	2.029			
	(0.203)	(0.186)	(0.126)			
Accumulated debt ($1,000)	-0.005	-0.017	0.002	-0.005	0.007	
	(0.004)	(0.005)	(0.004)	(0.003)	(0.002)	
Accumulated debt squared $\div$100	-0.009	0.001	-0.004	0.004	-0.006	
	(0.001)	(0.003)	(0.003)	(0.002)	(0.001)	
Non-grad $\times \mathbb{E}[u(\text{consumption})] \div 1,000$	1.642	1.642	1.642	1.642	1.642	
	(0.165)	(0.165)	(0.165)	(0.165)	(0.165)	
Previous high school	0.970	2.692	1.772	1.164	0.912	-0.530
	(0.102)	(0.163)	(0.112)	(0.085)	(0.083)	(0.185)
Previous 2-year college	2.422	1.032	0.722	0.033	0.318	0.088

continued

Variable	2-year	4-year Sci	4-year Non-Sci	Work PT	Work FT	White Collar
	(0.088)	(0.165)	(0.114)	(0.096)	(0.091)	(0.136)
Previous 4-year science	0.890	4.504	1.912	0.643	0.407	-0.188
	(0.180)	(0.138)	(0.131)	(0.106)	(0.102)	(0.125)
Previous 4-year non-science	0.319	1.940	3.524	0.604	0.634	0.073
	(0.158)	(0.162)	(0.098)	(0.090)	(0.081)	(0.094)
Previous work part-time	0.005	0.471	0.416	2.210	1.377	-0.987
	(0.096)	(0.122)	(0.101)	(0.049)	(0.054)	(0.090)
Previous work full-time	0.090	0.163	0.570	0.978	2.293	-0.953
	(0.104)	(0.145)	(0.106)	(0.064)	(0.043)	(0.088)
Previous work white-collar	-0.011	-0.435	-0.168	-1.346	-1.513	2.730
	(0.145)	(0.170)	(0.134)	(0.091)	(0.085)	(0.152)
Currently work white-collar				-0.137		
				(0.078)		
Currently work part-time	0.877	-0.187	-0.113			
	(0.076)	(0.082)	(0.070)			
Currently work full-time	-0.463	-1.794	-1.926			
	(0.083)	(0.096)	(0.081)			
Currently work part-time in white collar	-0.320	0.315	0.086			
	(0.190)	(0.141)	(0.124)			
Currently work full-time in white collar	-0.098	0.912	0.184			
	(0.140)	(0.142)	(0.122)			
Unobserved type (H, H, H)	0.269	0.262	0.298	-0.131	-0.089	-0.174
	(0.087)	(0.118)	(0.090)	(0.061)	(0.053)	(0.089)

continued

Variable	2-year	4-year Sci	4-year Non-Sci	Work PT	Work FT	White Collar
Unobserved type (H, H, L)	0.271	0.304	0.397	-0.000	-0.058	-0.257
	(0.080)	(0.111)	(0.089)	(0.057)	(0.046)	(0.083)
Unobserved type (H, L, H)	0.224	0.074	0.292	-0.228	-0.069	-0.114
	(0.090)	(0.128)	(0.101)	(0.071)	(0.053)	(0.092)
Unobserved type (H, L, L)	0.202	0.107	0.323	-0.002	-0.052	-0.265
	(0.078)	(0.112)	(0.091)	(0.060)	(0.049)	(0.087)
Unobserved type (L, H, H)	0.023	-0.039	0.213	-0.125	-0.064	-0.143
	(0.090)	(0.120)	(0.092)	(0.070)	(0.055)	(0.101)
Unobserved type (L, H, L)	0.006	-0.105	0.054	-0.008	0.029	-0.096
	(0.075)	(0.092)	(0.075)	(0.050)	(0.041)	(0.086)
Unobserved type (L, L, H)	-0.032	-0.100	0.126	-0.192	-0.091	-0.348
	(0.086)	(0.119)	(0.086)	(0.060)	(0.050)	(0.097)
College graduate				0.204	0.585	0.585
				(0.176)	(0.207)	(0.156)
Black × col. grad.				0.291	-0.065	0.637
				(0.182)	(0.152)	(0.129)
Hispanic × col. grad.				-0.029	-0.115	-0.097
				(0.193)	(0.145)	(0.137)
HS Grades (z-score) × col. grad				-0.151	-0.283	0.318
				(0.070)	(0.058)	(0.058)
Parent grad. col. × col. grad.				0.386	0.404	-0.596
				(0.132)	(0.098)	(0.107)
Born in 1980 × col. grad.				0.197	0.504	-1.110

continued

Variable	2-year	4-year Sci	4-year Non-Sci	Work PT	Work FT	White Collar
				(0.174)	(0.145)	(0.155)
Born in 1981 $\times$ col. grad.				-0.064	0.229	-0.728
				(0.174)	(0.138)	(0.140)
Born in 1982 $\times$ col. grad.				-0.772	-0.117	-0.807
				(0.164)	(0.130)	(0.143)
Born in 1983 $\times$ col. grad.				-0.665	-0.159	-0.517
				(0.159)	(0.134)	(0.131)
Family Income ($10,000) $\times$ col. grad.				-0.024	-0.033	0.034
				(0.011)	(0.009)	(0.012)
Col. grad $\times \mathbb{E}[u(\text{consumption})] \div 1,000$				0.648	0.648	
				(0.218)	(0.218)	
Constant Relative Risk Aversion parameter (θ)	0.4					
Log likelihood	-26,351					
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 0.37. Missing GPAs are estimated to be ≤ 2.5 w.p. 0.66, 2.5–3.0 w.p. 0.12, 3.0–3.6 w.p. 0.13, and 3.6–4.0 w.p. 0.09.

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, L, L)” for unobserved type.

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

Table B16: Model fit: Overall choice frequencies (Dynamic)

Choice alternative	Data Frequency (%)	Model Frequency (%)
2-year & work FT blue collar	1.70	1.83
2-year & work FT white collar	0.15	0.16
2-year & work PT blue collar	1.82	1.97
2-year & work PT white collar	0.10	0.13
2-year only	1.92	2.03
4-year Science & work FT blue collar	0.43	0.50
4-year Science & work FT white collar	0.15	0.18
4-year Science & work PT blue collar	0.94	1.10
4-year Science & work PT white collar	0.14	0.17
4-year Science only	2.27	2.46
4-year Non-Science & work FT blue collar	0.90	1.10
4-year Non-Science & work FT white collar	0.20	0.20
4-year Non-Science & work PT blue collar	1.88	2.28
4-year Non-Science & work PT white collar	0.29	0.32
4-year Non-Science only	4.06	4.67
Work PT blue collar	6.57	6.50
Work PT white collar	0.96	0.97
Work FT blue collar	42.71	41.79
Work FT white collar	8.61	9.01
Home production	24.22	22.64

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 0.3167 and in estimation is 0.3077.

Table B17: Model fit: Graduate choice frequencies

Choice alternative	Data Frequency (%)	Model Frequency (%)
Work PT blue collar	4.09	4.12
Work PT white collar	4.38	3.80
Work FT blue collar	34.54	36.39
Work FT white collar	43.66	42.00
Home production	13.34	13.69

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 0.6847 and in estimation is 0.7044.

Table B18: Average abilities by employment sector and education level at age 28 in baseline and counterfactual models

Sector and education level			White Collar			Blue Collar			4-year Science			4-year Non-Science			2-year		
			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	0.02	0.48	0.42	0.03	0.16	0.14	0.29	0.68	0.66	0.21	0.49	0.47	0.07	0.11	0.09
		Non-Science	0.01	0.39	0.32	-0.01	0.02	0.00	0.19	0.27	0.29	0.24	0.64	0.63	0.10	0.22	0.21
		Non-graduate	0.04	0.38	0.25	0.02	0.10	0.06	-0.14	-0.31	-0.24	-0.15	-0.23	-0.20	-0.05	-0.12	-0.06
	Blue Collar	Science	-0.01	-0.00	-0.00	-0.00	0.08	0.09	0.31	0.82	0.81	0.25	0.48	0.46	0.07	0.09	0.11
		Non-Science	-0.01	-0.07	-0.06	-0.03	-0.05	-0.05	0.19	0.48	0.45	0.24	0.70	0.67	0.10	0.26	0.27
		Non-graduate	0.01	-0.02	-0.02	0.02	0.04	0.04	-0.06	-0.18	-0.19	-0.07	-0.22	-0.23	-0.03	-0.06	-0.06
Part-time	White Collar	Science	-0.06	0.15	0.19	-0.01	0.01	0.02	0.27	0.83	0.78	0.17	0.54	0.46	0.03	0.09	0.09
		Non-Science	0.01	0.14	0.10	-0.03	-0.11	-0.11	0.07	0.48	0.22	0.14	0.79	0.54	0.06	0.26	0.16
		Non-graduate	0.01	0.15	-0.04	-0.02	-0.05	-0.09	-0.08	-0.40	-0.21	-0.05	-0.22	-0.17	-0.00	0.06	-0.10
	Blue Collar	Science	0.02	-0.19	-0.09	0.01	-0.07	-0.02	0.21	0.83	0.82	0.10	0.55	0.49	0.05	0.08	0.17
		Non-Science	-0.07	-0.07	-0.18	-0.04	-0.12	-0.15	0.15	0.49	0.42	0.21	0.73	0.61	0.16	0.25	0.22
		Non-graduate	-0.03	-0.13	-0.14	-0.02	-0.07	-0.08	-0.05	-0.17	-0.18	-0.05	-0.15	-0.19	-0.01	-0.03	-0.08
Home		Science	0.00	-0.22	-0.19	0.01	-0.16	-0.11	0.31	0.82	0.77	0.23	0.54	0.46	0.08	0.14	0.09
		Non-Science	-0.03	-0.23	-0.22	-0.05	-0.28	-0.24	0.13	0.42	0.45	0.23	0.71	0.72	0.06	0.19	0.22
		Non-graduate	-0.04	-0.24	-0.24	-0.04	-0.20	-0.20	-0.06	-0.21	-0.22	-0.06	-0.16	-0.18	-0.04	-0.09	-0.09

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

Table B19: 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						
$x = 0$	0.17	0.08	0.10	0.18	0.23	1.33
$x > 0$, white collar only	0.06	0.06	0.10	0.18	0.23	0.47
$x > 0$, blue collar only	0.12	0.03	0.10	0.18	0.22	3.55
$x > 0$, mixture	0.06	0.03	0.10	0.17	0.22	1.02
Continuous enrollment, graduate in non-science with x years of in-school work experience						
$x = 0$	0.16	0.08	0.19	0.09	0.22	2.99
$x > 0$, white collar only	0.07	0.06	0.19	0.09	0.21	0.98
$x > 0$, blue collar only	0.12	0.03	0.19	0.09	0.21	8.99
$x > 0$, mixture	0.06	0.03	0.18	0.09	0.20	1.87
Stop out (SO)						
SO, graduate in science	0.10	0.03	0.13	0.15	0.19	0.95
SO, graduate in non-science	0.10	0.03	0.19	0.11	0.18	3.13
SO then DO, start in 2yr	0.11	0.03	0.32	0.28	0.14	5.12
SO then DO, start in science	0.11	0.03	0.16	0.20	0.18	1.67
SO then DO, start in non-science	0.11	0.03	0.25	0.16	0.18	2.67
Truncated	0.10	0.02	0.24	0.20	0.16	5.42
Drop out (DO) after x years of school						
$x = 1$	0.14	0.05	0.36	0.33	0.21	16.57
$x = 2$	0.13	0.05	0.31	0.27	0.19	8.08
$x = 3$	0.13	0.05	0.26	0.22	0.18	4.37
$x = 4$	0.12	0.04	0.23	0.19	0.18	2.23
$x \geq 5$	0.10	0.03	0.21	0.16	0.16	2.25
Never attended college						
Never attend college	0.10	0.02	0.41	0.40	0.26	26.35
Time 0 population variance	0.17	0.08	0.41	0.40	0.26	

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 B20: 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						
$x = 0$	-0.06	-0.17	1.29	0.79	0.15	0.93
$x > 0$, white collar only	1.12	0.50	1.03	0.76	0.11	1.16
$x > 0$, blue collar only	0.29	0.27	1.17	0.73	0.21	5.19
$x > 0$, mixture	1.06	0.66	1.05	0.68	0.20	2.61
Continuous enrollment, graduate in non-science with x years of in-school work experience						
$x = 0$	-0.23	-0.61	0.66	1.11	0.44	1.79
$x > 0$, white collar only	0.66	-0.05	0.44	0.97	0.35	1.24
$x > 0$, blue collar only	0.12	-0.15	0.66	1.09	0.50	6.50
$x > 0$, mixture	0.63	0.02	0.52	1.00	0.51	2.46
Stop out (SO)						
SO, graduate in science	0.47	0.29	1.07	0.79	0.20	1.74
SO, graduate in non-science	0.08	-0.27	0.51	0.93	0.41	3.60
SO then DO, start in 2yr	-0.17	-0.11	-0.30	-0.31	0.05	3.32
SO then DO, start in science	-0.27	-0.14	0.28	0.06	0.01	1.88
SO then DO, start in non-science	-0.18	-0.22	0.10	0.30	0.19	3.36
Truncated	-0.17	-0.20	0.06	0.14	0.06	4.46
Drop out (DO) after x years of school						
$x = 1$	-0.11	-0.00	-0.41	-0.46	-0.17	14.79
$x = 2$	-0.14	-0.04	-0.20	-0.20	-0.02	7.54
$x = 3$	-0.24	-0.13	0.13	0.18	0.11	5.05
$x = 4$	-0.23	-0.18	0.22	0.21	0.12	2.20
$x \geq 5$	-0.20	-0.22	0.24	0.28	0.23	1.11
Never attended college						
Never attend college	-0.05	0.08	-0.57	-0.65	-0.32	29.06

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 B21: Average abilities for different choice paths in full-information reduced-credit-constraints 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						
$x = 0$	0.13	-0.03	1.19	0.76	0.14	1.49
$x > 0$, white collar only	1.19	0.65	0.99	0.70	0.18	0.68
$x > 0$, blue collar only	0.40	0.33	1.14	0.69	0.17	7.02
$x > 0$, mixture	1.30	0.78	0.91	0.61	0.14	1.48
Continuous enrollment, graduate in non-science with x years of in-school work experience						
$x = 0$	-0.20	-0.60	0.63	1.11	0.45	2.27
$x > 0$, white collar only	0.92	0.03	0.40	0.94	0.47	0.65
$x > 0$, blue collar only	0.21	-0.12	0.58	1.02	0.41	9.32
$x > 0$, mixture	0.87	0.23	0.34	0.87	0.36	1.10
Stop out (SO)						
SO, graduate in science	0.23	0.08	1.01	0.75	0.24	1.97
SO, graduate in non-science	-0.00	-0.26	0.52	0.88	0.42	4.10
SO then DO, start in 2yr	-0.20	-0.14	-0.41	-0.40	0.01	4.05
SO then DO, start in science	-0.37	-0.14	0.25	-0.00	-0.06	2.05
SO then DO, start in non-science	-0.31	-0.24	0.02	0.24	0.19	3.95
Truncated	-0.18	-0.16	0.02	0.08	0.10	6.40
Drop out (DO) after x years of school						
$x = 1$	-0.09	0.05	-0.49	-0.53	-0.19	13.71
$x = 2$	-0.19	-0.06	-0.26	-0.28	-0.10	6.81
$x = 3$	-0.25	-0.14	0.01	0.06	0.08	5.06
$x = 4$	-0.29	-0.16	0.20	0.17	0.11	2.77
$x \geq 5$	-0.20	-0.26	0.19	0.21	0.06	1.34
Never attended college						
Never attend college	-0.03	0.16	-0.62	-0.71	-0.33	23.79

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 B22: College completion status frequencies: data, baseline model, static model

Status	Data	Baseline model	Static model
Continuous completion (CC), Science	2.88	3.14	3.12
Continuous completion (CC), Non-Science	7.21	7.84	7.29
Stop out (SO) but graduated Science	0.24	0.45	0.45
Stop out (SO) but graduated Non-Science	1.23	1.58	1.52
Stop out (SO) then drop out	2.70	6.02	5.81
Drop out (DO)	20.83	26.08	25.93
Never went to college	28.61	32.12	33.88
Truncated	36.30	22.77	22.00

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. 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.

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.

C Estimation of measurement system

This appendix section provides an overview of how we specify our measurement system.

C.1 Measurement system

C.1.1 Dimensions of unobserved types

We assume that there are three different dimensions of discrete unobserved 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 ability and preferences

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

C.1.2 Measurements of unobserved types

Here we outline the measurements used for each of the dimensions of unobserved types.

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 ability and preferences To identify schooling preferences, we make use of measures of schooling ability and preferences, from which the marginal distribution (of the joint system) identifies the schooling preferences. The measurements are:

- Number of Advanced Placement (AP) tests taken
- 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 a censored variable.

Work ability and preferences 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. As in the case of schooling ability and preferences, the measurements of work ability and preferences 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 perceived likelihood of working part- or full-time at age 30 (reported as a percent chance on a scale from 0-100)
- The parent's perceived likelihood of the individual working part- or full-time at age 30 (reported on the same scale as above)

C.2 Estimation

To estimate the measurement system, we use a variety of regression 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 unobserved type. 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). For individual i and measurement b, we have the following equation:

$$y_{ibr} = X_i \tilde{\beta}_b + \xi_{ir} \omega_b + \varepsilon_{ib} \tag{C.1}$$

where X_i consists of the demographic and family background variables, ξ_{ir} is a set of dummies indicating which of r types the individual belongs to, and ε_{ib} is measurement error.

Note that, in the measurements for schooling ability and preferences, there will be two different sets of τ's included (i.e. the type dummies associated with schooling ability, and the type dummies associated with schooling preferences).

C.2.1 Component likelihood functions

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:

$$\mathcal{L}_{ibr}^{co}\left(\Theta; y_{ibr}, X_i, \xi_{ir}\right) = \frac{1}{\sigma_b} \varphi\left(\frac{y_{ibr} - X_i \tilde{\beta}_b - \xi_{ir} \omega_b}{\sigma_b}\right) \tag{C.2}$$

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

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:

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

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

Discrete ordered measurements For measurements that are discrete but have an inherent ordering (e.g. number of AP tests taken, 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:

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

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

$$
\begin{aligned}
P_{o,ijbr} &= \Pr\left(y_{ibr} = j\right) \\
&= \Pr\left(\kappa_{j-1,b} < X_i\tilde{\beta}_b + \xi_{ir}\omega_b \leq \kappa_{jb}\right) \\
&= \frac{1}{1 + \exp\left(X_i\tilde{\beta}_b + \xi_{ir}\omega_b - \kappa_{jb}\right)} - \frac{1}{1 + \exp\left(X_i\tilde{\beta}_b + \xi_{ir}\omega_b - \kappa_{j-1,b}\right)}
\end{aligned}
\tag{C.5}
$$

when ε_{ijbr} is distributed Type 1 Extreme Value, and where $\kappa_{0b} = -\infty$ and $\kappa_{J^b} = \infty$.

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:

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

where

$$P_{u,ijbr} = \frac{\exp\left(X_i\tilde{\beta}_{jb} + \xi_{ir}\gamma_{jb}\right)}{\sum_k \exp\left(X_i\beta_{kb} + \xi_{ir}\gamma_{kb}\right)} \tag{C.7}$$

when ε_{ijbr} is distributed Type 1 Extreme Value

C.2.2 Joint likelihood function

The joint likelihood function for the entire model (conditional on unobserved type) is then

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

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

C.2.3 The EM algorithm

We use the EM algorithm to iteratively estimate the measurement models (using the component likelihoods in Section C.2.1), treating unobserved type as if it were observed. We then update the probability of being a particular type using the joint likelihood of the measurements (as specified in Section C.2.2) according to

$$q_{ir} = \frac{\pi_r \mathcal{L}_{ir}}{\sum_{r'=1}^{R} \pi_{r'} \mathcal{L}_{ir'}} \tag{C.9}$$

C.3 Summary of 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 AP tests	school abil. & pref.	ordered categorical	ordered logit	0, 1, or 2+ tests
Number of times late for school	school abil. & pref.	ordered categorical	ordered logit	0, 1, 2, 3, 4, 5, 6-10, or 11+ times
Broke rules regularly	school abil. & pref.	ordered categorical	ordered logit	1-7 Likert scale
Hours per week took extra classes	school abil. & pref.	censored ($\underline{y} = 0$)	normal MLE	—
Took class during break?	school abil. & pref.	binary	binary logit	yes or no
Reason took class during break	school abil. & pref.	binary	binary logit	1) for enrichment; 2) to catch up
Have high standards at work	work abil. & pref.	ordered categorical	ordered logit	1-7 Likert scale
Make every effort to do what is expected	work abil. & pref.	ordered categorical	ordered logit	1-7 Likert scale
Percent chance work at age 30	work abil. & pref.	ordered categorical	ordered logit	0-75%, 76-90%, 91-100%
Parent: percent chance i works at age 30	work abil. & pref.	ordered categorical	ordered logit	0-75%, 76-90%, 91-100%

D Integration of missing outcomes

This appendix details our treatment of missing majors and GPA observations. In the estimation, 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.

The notation used throughout this section mirrors that which is used in that subsection. 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.

D.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. To the extent that the learning likelihood is non-separable across time (because of person-specific ability that is being learned about), we treat the integration accordingly. Below, we list the joint probabilities of i being of a particular unobserved type and unobserved major or GPA quartile (if either or both of these outcomes are missing).

$$\Pr(r|i) = q_{ir} = \frac{\pi_r \mathcal{L}_{ir}}{\sum_{r'=1}^{R} \pi_{r'} \mathcal{L}_{ir'}} \tag{D.1}$$

$$\Pr(r, m|i) = \Pr(r|i)\Pr(m|r, i) = q_{irm} = q_{ir} \frac{\mathcal{L}^*_{idr}\left(G_m, w_u, w_s\right)}{\sum_{r'=1}^{R} \mathcal{L}^*_{idr'}\left(G_m, w_u, w_s\right)} \tag{D.2}$$

$$\Pr(r, g|i) = \Pr(r|i)\Pr(g|r, i) = q_{irg} = q_{ir} \frac{\int_{g' \in g} \mathcal{L}^*_{idr}\left(G_g, w_u, w_s\right) \mathrm{d}g'}{\int_{g' \in g} \mathcal{L}^*_{idr'}\left(G_g, w_u, w_s\right) \mathrm{d}g'} \tag{D.3}$$

$$\Pr(r, m, g|i) = \Pr(r|i)\Pr(m, g|r, i) = q_{irmg} = q_{ir} \frac{\int_{g' \in g} \mathcal{L}^*_{idr}\left(G_{mg}, w_u, w_s\right) \mathrm{d}g'}{\int_{g' \in g} \mathcal{L}^*_{idr'}\left(G_{mg}, w_u, w_s\right) \mathrm{d}g'} \tag{D.4}$$

where (D.1) holds for those who have no missing data and is the same as (38) and (C.9), (D.2) holds for those who only have a missing major, (D.3) holds for those who only have a missing GPA, and (D.4) holds for those who have both outcomes missing. Note that, for (D.3) and (D.4), the integral in the numerator is over quartile g of the GPA distribution, while in the denominator we integrate over the full support of the GPA distribution.

D.2 M-step

At the M-step of our algorithm, we 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

For the choice M-step, we construct $\mathcal{L}^*_{idr}$ to estimate a weighted multinomial logit, using q_{irmg}, q_{irm}, q_{irg}, or q_{ir} as the weights—depending on what is observed and what is unobserved. This multinomial logit is estimated on the entire population. At this point we take as given the parameter values from the measurement system likelihood function, denoted in (38) and (C.8)–(C.9) as $\mathcal{L}_{ir}$.

D.3 Subsequent stage estimation

The estimation steps described above pertain only to the first two stages of our model, in which we estimate the measurement system and recover the distribution of missing majors and GPAs.

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, 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 measurement system likelihood value.

E 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.

E.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)):

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

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.

E.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).

E.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):

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

We include the following in X:

- age
- log family income
- race/ethnicity dummies

E.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

E.2.3 Expected parental transfers

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

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

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

E.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.

E.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.[A5] This value is \$6,394. For two-year colleges, the value is \$1,380.

[A5] 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.

E.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.[A6]

We then obtain

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

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.

E.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:

[A6] 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.

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

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.

E.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.[A7] 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.

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

where A denotes assets, I denotes income, and P and C superscripts denote parent and child, respectively. X represents characteristics about the parent 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 $\widetilde{\ }$ the

[A7]The following commonly held assets are not considered in the EFC: equity of primary residence, retirement funds, and life insurance.

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

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

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

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

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 (E.7) and (E.8), we have

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

E.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.[A8]
- 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.

E.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.

E.4 Flow utility of consumption

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

[A8] 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.

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

where the consumption level C is given by Equations 7 and 8. We distinguish three cases:

1. Not in school and not working:

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

2. Working and not in school:

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

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 $a = \underline{C}^{1-\theta}$):

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

where $\Phi(\cdot)$ denotes the cdf. of a standard normal distribution.[A9]

[A9] 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:

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

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

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

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

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

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:

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

3. In school:

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

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):

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

where $f_{PT}(\cdot)$ and $f_G(\cdot)$ 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 (7). 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:

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

where $f_{PT}(\cdot)$, $f_W(\cdot)$, and $f_G(\cdot)$ 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 (7)).

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.

E.4.1 Example: 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 (E.3) and (E.19))

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

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 (E.19) in the case where the student receives grants would then be

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

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.

E.4.2 Example: 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

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

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: $\sigma_w=\sqrt{\sigma_\varepsilon^2+\lambda_1^2V(A_w|\mathcal{I}_t)}$ where λ_1 is the productivity index loading on in-school wages ($\lambda_1=1$ for out-of-school wages).

E.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 B5), 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.

F More details on finite dependence

This appendix section details the mathematical derivation of our difference in value functions listed in (24). The goal is to find an expression for $v_{jkl} - v_h$ that is not recursive. We accomplish this with finite dependence. 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 $\tilde{\lambda}_{t+1}^{(h)}$, we respectively weight the acceptance and rejection probabilities in $t+1$ (in the event that i receives an offer) by $\frac{1}{\tilde{\lambda}_{t+1}^{(h)}}$ and $1 - \frac{1}{\tilde{\lambda}_{t+1}^{(h)}}$. 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):

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

for all i, (j,k,l) and t, where c 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} \\
&= 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}
\tag{F.4}
$$

A clever 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 $t+3$:

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

F.1.3 Putting it together

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

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

Equation (F.6) precisely matches equation (24), up to small differences in notation.

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 \left(\begin{array}{c} \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{array} \right) \\
&= u_{ht} + \beta \left(\begin{array}{c} \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{array} \right) \\
&= u_{ht} + \beta \left(\begin{array}{c} \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{array} \right) \\
&= u_{ht} + \beta \left(\begin{array}{c} \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{array} \right) \\
&= u_{ht} + \beta \left(\begin{array}{c} \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{array} \right) + \\
&\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. $V_{t+3}(\cdot, \cdot, g_{t+2} = g) \equiv V_{t+3}(\cdot, \cdot, g_{t+1} = g)$ 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 E, 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:

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

- School-Home-Home path:

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

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 (24). The likelihood function for this model can be written as

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

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:

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

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} comes from (38).

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 $r \in \{1, \dots, R\}$ 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 3.6.

   (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 3.5.3.

   (e) Compute the implied posterior ability beliefs given the outcomes and choices generated previously as discussed in Subsection 3.5.1, 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 employ the algorithm discussed in Subsection 5.6 and detailed in Appendix D. 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
- 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)
- Loans have a 30-year repayment horizon[A10]
- 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

We also discretize the AR(1) process governing the aggregate labor market shocks in (15) 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.

[A10]In one of our counterfactuals, we set all loans to zero. See the first paragraph of Section 7 for more details.

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	(18)
β	Discount factor	(20)
$\tilde{\beta}$	Nuisance parameters in measurement system, inputs to consumption, and CCPs	(C.1), (E.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), (15)
δ_λ	White collar offer arrival parameters	(17), (H.4)
ε	Idiosyncratic shocks to grades, wages, preferences, and unobserved type measurements	(1), (4), (18), (20), (C.1), (C.7)
ζ_t	Innovations to aggregate labor market shock	(15)
θ	CRRA parameter on exp. util. of consumption	(18), (E.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	(17), (22)
ξ	Unobserved type in measurement system	(C.1)
π_r	Population probability mass of unobserved type r	(37), (38), (C.9)
ρ_t	Probability of having graduated before time t	(F.8)
σ	Standard deviation of various idiosyncratic shocks	(1), (4), (15), (36), (C.2)
$\tilde{\sigma}$	Nuisance parameter in inputs to consumption	(E.3)
τ	Index of enrollment or employment time period, distinct from calendar time t	(9)
ϕ	AR(1) coefficient on aggregate labor market shock	(15)
φ	Normal distribution pdf	(28), (31), (33), (C.2)
ψ	Graduation logit parameters	(16)
ω	Parameter on unobserved type 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	(28), (32)
Z_{it}	State variables for individual i in period t	(18), (24)
$\tilde{Z}_{it}$	Covariates in job offer arrival logit	(17)
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	(13)
Φ	Standard normal cdf	(C.3), (E.13)
X	Covariate matrices in various equations	(1), (4), (16)
Ω	Inverse of variance of idiosyncratic grade and wage shocks	(13)