College Attrition and the Dynamics of Information Revelation *

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

May 31, 2016

Abstract

This paper investigates the role played by informational frictions in college and the workplace. We estimate a dynamic structural model of schooling and work decisions, where individuals have imperfect information about their schooling ability and labor market productivity. We take into account the heterogeneity in schooling investments by distinguishing between two- and four-year colleges, graduate school, as well as science and non-science majors for four-year colleges. Individuals may also choose whether to work full-time, part-time, or not at all. A key feature of our approach is to account for correlated learning through college grades and wages, whereby individuals may leave or re-enter college as a result of the arrival of new information on their ability and productivity. Our findings indicate that the elimination of informational frictions would increase the college graduation rate by 9 percentage points, and would increase the college wage premium by 32.7 percentage points through increased sorting on ability.

JEL Classification: C35; D83; J24

*We thank the attendees of the 2012 Conference in Honor of John Kennan, the 2012 Cowles Foundation Conference on Structural Microeconomics, the 2012 North American Summer Meeting of the Econometric Society (Northwestern), the 2014 SED conference (Toronto), the 2014 SITE Workshop on Empirical Implementation of Theoretical Models of Strategic Interaction and Dynamic Behavior (Stanford), the 2014 WEA Annual Meetings (Denver), the 2014 Swiss Leading House conference on the Economics of Study Choices, the 2015 Barcelona GSE Structural Microeconometrics Workshop, the 2015 Labour Workshop (Warwick), the 11^{th} World Congress of the Econometric Society (Montréal), the 2015 SEA Meetings (New Orleans), the 2016 AEA Meetings (San Francisco), as well as seminar participants at Arizona State, Carnegie Mellon, CREST, IDB, Toulouse School of Economics, University of Chicago, University of Colorado-Boulder, University of North Carolina at Chapel Hill, University of Pennsylvania, and Yale University for helpful comments and suggestions.

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

‡London School of Economics and Political Science and CEP. E-mail address: E.M.Aucejo@lse.ac.uk

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

¶Social Science Research Institute, Duke University. E-mail address:tyler.ransom@duke.edu

1 Introduction

About half of students entering college in the United States do not earn a bachelor's degree within five years, a proportion that has been increasing since the 1970's (Bound et al. 2010). To the extent that there is a large wage premium to receiving a four-year college degree (Heckman et al. 1996, 2006, Goldin & Katz 2008, Bound & Turner 2011), this suggests that imperfect information and learning may be important to the decision to leave college. In this paper, we quantify the importance of imperfect information on academic ability as well as labor market productivity in the context of college enrollment decisions and of the transitions between college and work. We estimate the impact of these informational frictions on college enrollment, attrition and re-entry, and document the extent to which imperfect information acts as a barrier to college completion and sorting on ability in the labor market.

In order to quantify the role played by informational frictions in the decision to enter, leave or return to college, we estimate a dynamic model of schooling and work decisions in the spirit of Keane & 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 or receiving a GED, individuals decide in each period whether to attend college and/or work part-time or 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. Upon college graduation, his schooling options become whether to attend graduate school or not. 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) and 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 area may be informative about abilities in another area.

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 & 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 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 and the observed transitions between college and work, and to evaluate the impact of mperfect 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). The correlation between productivities in the unskilled and skilled sectors is also large, stressing the importance of allowing for correlated learning in this context. On the other hand, schooling abilities are only weakly correlated with productivity in both sectors, thus indicating that grades earned in college actually reveal little information about future labor market performance once we account for background characteristics and college readiness.

We then simulate our model under a counterfactual scenario where all individuals have perfect information on their abilities by the end of high school, and examine how schooling choices and sorting patterns would be affected. We find that the share of four-year college graduates would increase by around 9 percentage points (38%) relative to the baseline, virtually all through a decrease in dropout rates. Simulations further reveal that ability sorting would be much stronger in the perfect information scenario. We provide evidence that imperfect information on ability significantly limits the extent to which individuals can pursue their comparative advantage through schooling choices. We also find that, in the counterfactual scenario, wages in the skilled sector would increase by 16.7% while wages in the unskilled sector would decrease by 16%, resulting in a large (32.7 percentage points) increase in the college wage premium. Taken together, these results provide evidence that informational frictions play a major role in schooling decisions as well as subsequent labor market outcomes.

Our analysis builds on seminal research by Manski & Wise (1983) and Manski (1989), which argues that college entry can be seen as an experiment that may not lead to a college

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

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 ability. 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 & Urzúa 2009). Of particular relevance to us are the articles by Stinebrickner & Stinebrickner (2012, 2014*a*), 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 & Stinebrickner 2012, 2014*a*).[3] In this paper we relax this assumption, which is important to predict the substantial college re-entry rates of 40% which are observed in the data. By quantifying the importance of learning on schooling abilities as well as labor market productivities, and evaluating the joint informational value of schooling and labor market outcomes, our paper brings together the literatures on schooling choices, and occupational choices under imperfect information (see, e.g., Miller 1984, Sanders 2010, James 2011, Antonovics & Golan 2012).

The remainder of the paper is organized as follows. Section 2 presents the data. 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

[2] See also recent work by 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 (2012) 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.

between January 1, 1980 and December 31, 1984. Respondents were first interviewed in 1997 and have continued to be interviewed annually (for a total of 15 Rounds as of 2011, which corresponds to the most recent data used in the paper) on such topics as labor force activities, education, as well as marriage and fertility.

2.1 Data construction

From the NLSY97 data we classify individuals based on their labor force participation 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 Sciences, Technology, Engineering, and Mathematics (STEM). See Table A.1 for details on the exact majors in each category. Hereafter, 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.[4]

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

The two variables which provide information to students regarding their abilities are college grade point averages (GPA) and wages. College GPA, which we hereafter simply refer to as grades, is measured on a four-point scale and calculated as the average GPA across all semesters in the calendar year. 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.[6]

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

[5]Following this criterion, any individual who is unemployed in October is classified in the home production sector.

[6]Note that one can think of this variable as a measure of performance on the job. As such, we do not

2. If a person does not report hourly compensation, we impute earnings as annual income divided by annual hours worked. Specifically, 6.7% of observations that are classified as working part- or full-time have earnings that are defined using this alternative measure.

3. Finally, we top- and bottom-code the resulting earnings distribution at the 99.5 percentile and 2.5 percentile, respectively.

It is worth noting that grades and college majors are missing (or reported as "don't know") frequently in our data. Grades are missing for 28% of college students, and majors are missing for 25% of four-year college students. This partly reflects the fact that interviews primarily occur between October and March when academic outcomes have yet to be realized, resulting in a significant fraction of students reporting college attendance but not answering college-specific questions.[7] We address this missing data issue as discussed in Section 5.6.

Finally, we restrict our analysis to males who have (*i*) either graduated high school or have obtained their GED and (*ii*) have valid Armed Services Vocational Aptitude Battery (ASVAB) test scores. As measures of academic preparation, we use the math and verbal components of the SAT score when observed.[8] If an individual did not take the SAT, we predict the individual's SAT score using each of his ASVAB component test scores.[9] This predicted SAT score is used synonymously with actual SAT score throughout. We drop all current and future observations for any respondents missing wage observations 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 21,343 person-year observations for 2,713 males. Table A.2 in Appendix A gives further details about the sample selection.

need to assume that employers have perfect information about workers' abilities, but rather that workers are paid according to their realized productivity.

[7]The modal interview months are November and December. Data are collected for the period of time between interviews, so if, for example, a respondent is interviewed in October 2004 and again in November 2005, college information related to the Fall 2004 and Spring 2005 semesters would both be reported in the November 2005 interview. This lag in reporting is also likely to contribute to the missing data problem.

[8]The distribution of raw SAT scores for each component is standardized to be zero-mean and standard deviation 1 for the NLSY97 population who took the SAT.

[9]Each ASVAB component test score is standardized to be zero-mean and standard deviation 1 for the NLSY97 population. The sub-tests used are Arithmetic Reasoning, Mathematical Knowledge, Numerical Operations, Paragraph Comprehension and Word Knowledge. Estimates of these two regressions are listed in Table A.4

2.2 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 average, higher SAT test scores, with science majors having higher scores than non-science majors, even for SAT verbal. 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 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. Overall, this difference in composition between two- and four-year colleges (and between majors in four-year colleges) stresses the need to distinguish between college and major type when modeling college enrollment decisions.

Table 2 gives the frequencies of continuous enrollment (either in two- or four-year colleges) until graduation from a four-year college, stopping out (i.e. leaving college before graduating from a four-year college and returning to school at some point) and dropping out (i.e. permanently leaving college before four-year graduation). Evident from Table 2 is that dropping out and stopping out are more common in two-year colleges than four-year colleges. Four-year science majors have the lowest proportions of dropping out and stopping out. This again points at the need to distinguish between these two types of colleges and majors in our model. Due to the ongoing nature of the survey and the fact that some respondents are still enrolled in college, the information from Table 2 can be used to calculate lower bounds on the overall stopout rates. The frequencies reported in this table indicate that stopout rates are sizable. For example, of those who had graduated with a four-year college degree by round 15 of the survey, 17% were stopouts. For those beginning college in a four-year college science major, this number is 9.8%, compared with 14.1% for those who began as non-science majors. For those originating in a two-year college, the stopout rate among four-year graduates is significantly higher (39.5%). As another example of stopout prevalence, Table 2 shows that, of those who had left college at any point before graduating, 40.8% re-enrolled at some later point.

2.3 Descriptive investigation of learning

Given the substantial dropout and stopout rates, we next see if there is any descriptive evidence suggesting that learning about one's abilities may play a role in these decisions.

The first panel of Table 3 shows than those who stay in school at period $t+1$ have significantly higher grades at period t that those who were in school at t but not at $t+1$. Further, higher grades are associated with staying in the major if one is in a four-year institution, and with switching to a four-year college if one is at a two-year institution. These differences, however, may not reflect learning as those with lower grades may have come from worse family backgrounds and/or be of lower ability and at greater risk of dropping out.

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, and whether the individual was working part- or full-time), and compute the residuals. We then compare the mean residual at t for different educational choices at $t+1$ in the third and fourth panels of Table 3. Those who have higher grade residuals are more likely to stay in school and less likely to switch majors. These descriptive findings are consistent with two stories, which may not be mutually exclusive: (*i*) individuals decide to leave or switch college/major as they learn about their schooling ability, or (*ii*) those who leave or switch college/major tend to have a lower ability, that they observe perfectly even before starting college, but that is unknown to the econometrician. Telling apart these two explanations is a key objective of our structural estimation, which will be discussed in the rest of the paper.

Finally, to illustrate learning in the labor market as a reason for stopouts to return to college, we perform a similar analysis as in Table 3 but now using information on wages. The first panel of Table 4 presents mean log wages for those who remain working and those who have stopped out, broken out by next-period decision, while the second panel of Table 4 shows the difference between actual and expected log wages.[10] Overall, these tables show that those who decide to return to school earn lower wages in previous year. For example, those who have left college for the labor force and choose to return to school have 5% lower wages on average the year before returning to school, even when controlling for a rich set of individual, family background, and schooling and labor force experience characteristics. This pattern is consistent with learning on labor market productivity contributing to the decision to return to college. However, in order to quantify the impact of learning on college re-entry, one needs to account for the possibility that individuals also differ in their expected (but unobserved to the econometrician) productivity by the end of high school. Our empirical analysis incorporates and distinguishes between these two types of abilities that are unobserved to the econometrician, but initially known or unknown, respectively, to

[10]Notice that these tables are conditioning on the sample of individuals who are working and have attended at least one year of college but did not graduate.

the individuals themselves.

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. After graduating from a four-year college, the schooling option includes whether or not to enroll in graduate school. In addition to choosing among the different schooling options, individuals also choose whether to work full-time, part-time, or not at all. These work options are available regardless of their schooling choice.[11]

Individuals only have imperfect information about their abilities. Importantly, we treat abilities as multidimensional vectors with five components. Namely, individuals have different abilities for each of the schooling options (two-year, four-year science, and four-year non-science) as well as abilities in the labor market without a college degree (labeled unskilled) and with a college degree (labeled skilled). 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 productivities in the labor market. Similarly, unskilled wages may be informative not only with regard to unskilled productivity, but also with regard to the individual's schooling abilities and skilled productivity.

Individuals 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,

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

and one through their wage. It is interesting to note that, in this setting, while working while in college may be detrimental to academic performance (see, e.g., Stinebrickner & Stinebrickner 2003), it is also 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 components 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.

3.2 Grades

In the following, we denote by $j \in \{a, bs, bn\}$ the type of college and major attended, where a (for *Associate*) denotes a two-year college, bs (for *Bachelor*, Science) a four-year college science major, and bn (for *Bachelor*, Non-science) a four-year college non-science major. Individuals are indexed by i.

We assume that grades depend on A_{ij}, where A_{ij} is the unobserved schooling ability about which individuals have some initial beliefs given by the prior distribution $\mathcal{N}(0, \sigma^2_{Aj})$. Grades also depend on a set of covariates for college type j and period of college enrollment τ,[12] $X_{ij\tau}$, that are known to the individual and include observed skill measures (i.e. high school grades, math and verbal SAT), 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).[13] Importantly, $X_{ij\tau}$ also includes characteristics that are observed to the individual but not the econometrician which may reflect unobserved ability. In practice, we account for unobserved (to the econometrician only) heterogeneity by allowing for two different unobserved heterogeneity types, where the type is a permanent characteristic of the individual.[14] In the following and throughout the paper, we assume that unobserved ability and unobserved heterogeneity type are mutually independent, and that they are also independent from the observed covariates at period $t = 1$.

[12] 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.

[13] 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.

[14] In estimation we account for unobserved heterogeneity using a finite mixture approach a la Heckman & Singer (1984).

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

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

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:

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

Hence, the return to the academic index varies over period of college enrollment and across majors. As such, while remaining parsimonious, this specification allows for different effects of ability on grades for lower- and upper-classmen. 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. We assume that there are two sectors in the labor market, which are indexed by l and referred to as *skilled* ($l = s$) and *unskilled* ($l = u$) in the following. All four-year college graduates and individuals with a graduate school degree who participate in the labor market are part of the skilled sector. The unskilled sector includes all other labor market participants, namely high school graduates or GED recipients, college students, college dropouts and stopouts (before graduating from college), as well as two-year college graduates.

Denoting by t calendar time (year), log wages in sector l are assumed to depend linearly on sector-specific productivity A_{il}, a set of observed (to the individual) characteristics X_{ilt} (i.e. years of education, indicators for working part-time or full-time or being enrolled in college, labor marker experiences, age, ability measures, race, parental education, and college major for those in the skilled sector), sector-specific time dummies δ_{lt}, and idiosyncratic shocks ε_{ilt}:

$$w_{ilt} \quad = \quad \delta_{lt} + X_{ilt}\gamma_{1l} + A_{il} + \varepsilon_{ilt} \tag{4}$$

Similar to the model for grades, X_{ilt} may contain information on abilities that are known to the individual but not the econometrician. 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.

We account for nonstationarities in wages by including calendar year dummies, δ_{lt}, thus incorporating business cycle effects. If we did not control for these non-stationarities, 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 3.4.2.

3.4 Beliefs

Individuals are uncertain about (i) their future preference shocks, (ii) their schooling ability and labor market productivity, (iii) the evolution of the market shocks (the δ_{lt}'s), and (iv) (four-year) college graduation. The first component, future preference shocks, will be discussed in Section 3.5 when we describe preferences. We discuss the other components here.

3.4.1 Beliefs over schooling ability and labor market productivity

We denote A_i as the five-dimensional ability vector, $A_i \equiv (A_{ia}, A_{ibs}, A_{ibn}, A_{is}, A_{iu})'$ (simply referred to as *ability* in the following). We assume that individuals are rational and update their beliefs in a Bayesian fashion. Their initial ability beliefs are given by the population distribution of A, which is supposed to be multivariate normal with mean zero and covariance matrix Δ. Importantly, we do not restrict Δ to be diagonal, thus allowing for correlated learning across the five different ability components.

At each period τ of college attendance, individuals use their realizations of grades and wages (if they work while in college) to update their beliefs about their schooling abilities in all college options $(A_{ia}, A_{ibs}, A_{ibn})$, as well as their labor market productivity in both sectors (A_{is}, A_{iu}). 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:

$$S_{ij\tau} = G_{ij\tau} - \gamma_{0j} - X_{ij\tau}\gamma_{1j} \quad (5)$$

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

$$S_{ij\tau} = \frac{G_{ij\tau} - \lambda_{0j} - \lambda_{1j}(\gamma_{0j} + X_{ij\tau}\gamma_{1j})}{\lambda_{1j}} \quad (6)$$

Similarly, individuals who participate in the labor market update their ability beliefs after receiving their wages. The signal for sector l and period t is given by:

$$S_{ilt} = w_{ilt} - \delta_{lt} - X_{ilt}\gamma_{1l} \quad (7)$$

Finally, individuals may choose to work while in college, in which case they will receive two ability signals $(S_{ij\tau}, S_{iut})$.[15]

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 idiosyncratic shocks.[16] 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.

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 $\Sigma_t(A_i)$ the posterior ability mean and covariance at the end of period t, we have (see DeGroot 1970):

$$E_t(A_i) = (\Sigma_{t-1}^{-1}(A_i) + \Omega_{it})^{-1}(\Sigma_{t-1}^{-1}(A_i)E_{t-1}(A_i) + \Omega_{it}\widetilde{S}_{it}) \quad (8)$$

$$\Sigma_t(A_i) = (\Sigma_{t-1}^{-1}(A_i) + \Omega_{it})^{-1} \quad (9)$$

As in a more standard one-dimensional learning model, prior variances decrease towards zero as individuals receive additional ability signals, thus giving more weight to the prior ability and less to the signal.

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

[16] Note that, for the college options, the idiosyncratic variances will depend on the year of enrollment.

3.4.2 *Beliefs over market shocks*

We now specify how individuals form their beliefs about the labor market. Individuals observe the current values of δ_{st} and δ_{ut}. We assume that the process governing the aggregate shocks is an AR(1) with a sector-specific autocorrelation parameter:

$$\delta_{lt} = \phi_{1l}\delta_{lt-1} + \zeta_{lt} \tag{10}$$

where the ζ_{lt} are assumed to be i.i.d. over time, but allowed to be correlated across sectors, following a $\mathcal{N}(0, \sigma_\zeta^2)$. The assumption that the aggregate shocks follow an AR(1) process, or a discretized version of it (Markov process of order 1) is common in the literature (see, e.g., Adda et al. 2010, Robin 2011). Given the realizations of the market shocks δ_{lt-1}'s, individuals then integrate over possible realizations of the ζ_{lt}'s when forming their expectations over the future.[17]

3.4.3 *Beliefs over graduation*

We treat graduation as probabilistic. Individuals are only at risk of graduating if they have completed at least three years of college and if they are currently attending a four-year institution. Individuals in this risk set face a probability of graduation in 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 SAT scores, high school grades, and demographics, as well as an unobserved (to the econometrician) heterogeneity type dummy. 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.[18] We then assume that the probability of graduating conditional on $X_{ig\tau}$ takes a logit form:

$$\Pr(\text{grad}_{i\tau} = 1 | X_{ig\tau}) = \frac{\exp(X_{ig\tau}\psi)}{1 + \exp(X_{ig\tau}\psi)} \tag{11}$$

Individuals are assumed to know the parameters ψ and form expectations over their probabilities of graduating using (11).

[17] In practice, we estimate separately for each sector the autocorrelation parameter and the variance of the shocks from the (estimated) calendar time dummies $(\widehat{\delta}_{st})_t$ and $(\widehat{\delta}_{ut})_t$.

[18] While we do allow the prior ability beliefs 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 which would substantially complicate our model.

3.5 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 $j \in \{a, bs, bn, 0\}$ for those who have not graduated from a four-year institution and where, indexing graduate school by gs, $j \in \{gs, 0\}$ for those who have graduated from a four-year college. We denote the work options by $k \in \{p, f, 0\}$ where p and f respectively refer to part-time and full-time work. The choice $d_{it} = (0, 0)$ refers to the home production option: no work ($k = 0$) and no school ($j = 0$).

Up to an intercept term and an idiosyncratic preference shock, we assume that the utility of the choice (j, k) is additively separable in its schooling and work components. Let Z_{1it} denote the variables that affect the utility of school and Z_{2it} denote the variables that affect the utility of work. These variables need not be mutually exclusive; both include many of the same variables that affect grades and wages. The flow payoff for choice $d_{it} = (j, k)$ is then given by (letting $Z_{it} = (Z_{1it}, Z_{2it})'$):

$$\begin{aligned} U_{ijk}(Z_{it}, \varepsilon_{ijkt}) &= \alpha_{jk} + Z_{1it}\alpha_j + Z_{2it}\alpha_k + \varepsilon_{ijkt} && (12) \\ &= u_{jk}(Z_{it}) + \varepsilon_{ijkt} && (13) \end{aligned}$$

where the idiosyncratic preference shocks ε_{ijkt} are assumed to be i.i.d. following a Type 1 extreme value distribution. 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.[19] Specific to Z_{2it} are the expected log wages in sector k, which depend on the expected productivity in sector k as well as on the beliefs about the aggregate shocks affecting sector k in period t.[20] Z_{1it} and Z_{2it} also include demographics, ability measures (SAT math and verbal, high school GPA), as well as controls for the previous choice to allow for switching costs, similar in spirit to Keane & Wolpin (1997). Finally, both sets of variables also include controls for unobserved type-specific heterogeneity, allowing for unobserved preferences for the different schooling and work alternatives.

Finally, the home production sector ($d_{it} = (0, 0)$) 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.

[19] Expected ability is included in Z_{1it} as an (inverse) proxy of the cost of effort associated with college attendance.

[20] See also Belzil & Hansen (2002), Arcidiacono (2004, 2005), Pavan (2011), Beffy et al. (2012) and Wiswall & Zafar (2015, 2016) who estimate models of human capital accumulation where the utility of work is assumed to be logarithmic in earnings. In practice, we impose that the coefficient on log wages is the same in both sectors. Note that differences in hours between part- and full-time work will come out in the intercept term.

3.6 The optimization problem

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

$$E\left[\sum_{t=1}^{T}\beta^{t-1}\sum_{j}\sum_{k}(u_{jk}(Z_{it})+\varepsilon_{ijkt})1\{d_{it}=(j,k)\}\right] \tag{14}$$

where $\beta \in (0,1)$ is the discount factor. The expectation is taken with respect to the distribution of the future idiosyncratic shocks, the signals associated with the different choice paths, the beliefs over the aggregate shocks to wages, 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_{jkt} is given by:

$$v_{jkt}(Z_{it}) = u_{jk}(Z_{it}) + \beta E_t(V_{t+1}(Z_{it+1})|Z_{it}, d_{it} = (j,k))$$

where the term $E_t(.|)$ 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 $t-1$. Finally, the assumption that the ε's are i.i.d. Type 1 extreme value yields the following familiar log-sum formula:

$$v_{jkt}(Z_{it}) = u_{jk}(Z_{it}) + \beta E_t\left[\ln\left(\sum_{j}\sum_{k}\exp(v_{jkt+1}(Z_{it+1}))\right)\middle| Z_{it}, d_{it}=(j,k)\right] + \beta\gamma \tag{15}$$

where γ denotes Euler's constant.

4 Identification

Before turning to the estimation procedure, we first informally discuss the identification of the model. For the sake of exposition, we start by considering the case of the model without type-specific unobserved heterogeneity, before briefly discussing the identification of the unobserved heterogeneity parameters. As is common for these types of dynamic discrete choice models (see, e.g., Rust 1994, Magnac & Thesmar 2002, Arcidiacono & Miller 2015*a*), identification of the flow utility parameters relies on the distributional assumptions imposed on the idiosyncratic shocks, the normalization of the home production utility and the discount factor β, which is set equal to 0.9 throughout the paper.

First consider the identification of the outcome equations (grades and log wages). The grades $G_{ij\tau}$ are only observed for the individuals who are enrolled in a type-j (college, major) in their τ-th period of college enrollment. To the extent that college enrollment decisions depend on the ability (A_i), this raises a selection issue. We show the identification of the grade equation parameters by using, for each period τ, the prior ability at the beginning of the period ($E_{\tau-1}(A_{ij})$) as a control function in the grade equation. Specifically, we consider the following augmented regression, for $j \in \{bs, bn\}$ and $\tau > 2$:

$$G_{ij\tau} = \lambda_{0j} + \lambda_{1j}(\gamma_{0j} + X_{ij\tau}\gamma_{1j}) + \lambda_{1j}E_{\tau-1}(A_{ij}) + \nu_{ij\tau} \tag{16}$$

where it follows from the Bayesian updating rule (see Equation (8) on p. 13) that $E_{\tau-1}(A_{ij})$ can be expressed as a weighted sum of all the past ability signals. Under the key assumption, consistent with the specification of the flow utilities discussed in Subsection 3.5, that college decisions only depend on ability through the ability beliefs, application of ordinary least squares to Equation (16) identifies the location and scale parameters $(\lambda_{0j}, \lambda_{1j})$, with the ability index coefficients $(\gamma_{0j}, \gamma_{1j})$ being identified from the first and second period grades.

Grades in the first two years of four-year college, as well as in two-year colleges, can indeed be expressed as follows:

$$G_{ij\tau} = \gamma_{0j} + X_{ij\tau}\gamma_{1j} + E_{\tau-1}(A_{ij}) + \nu_{ij\tau} \tag{17}$$

Application of ordinary least squares therefore directly identifies $(\gamma_{0j}, \gamma_{1j})$. Identification of the ability index coefficients also follows from the assumption that enrollment decisions only depend on ability through the past ability signals. Similar arguments can be used for the identification of the parameters of the log wage equations in each sector.

Turning to the distribution of the unobserved ability vector A, the signal-to-noise ratios as well as the ability covariance matrix Δ are identified from the observed history of grade and wage signals, and in particular from the correlations between current grades and wages, and past ability signals. Of particular interest in our context are the correlations between the different ability components, which are primarily identified from individuals switching across types of colleges and majors, as well as between unskilled and skilled labor market sectors. College employment, however, provides a more direct cross-sectional source of identification for the correlations between productivity in the unskilled sector and all three schooling abilities.[21]

[21] See Table A.3 which lists the number of individuals in our sample that ever participate in any given combination of sectors and schooling options.

Finally, in our specification with latent heterogeneity types, one also needs to tell apart the type-specific unobserved (to the econometrician only) heterogeneity components, from the ability A unobserved to the econometrician as well as, at least initially, the individual himself. Initial choices made before the individuals get a chance to learn about their ability A offer a natural solution to this deconvolution problem, and as such are a key source of identification. For instance, low-SAT individuals who choose to enroll in a four-year college right after high school graduation would be predicted to have high type-specific unobserved preferences for one of the four-year college options. Alternatively, low-SAT individuals who are enrolled in college may decide to leave college after receiving high grades. Individuals exhibiting this type of behavior would be predicted to have a high type-specific schooling ability.

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).[22]

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. Specifically, estimation in this case proceeds in two key stages. In the first stage, one can estimate the parameters from the grade and wage equations and the college graduation process, as well as the choice probabilities associated with all schooling and work alternatives.The second stage is devoted to the estimation of the flow utility parameters, taking as given the first-stage estimates. The validity of this sequential approach rests on the likelihood being separable in the contributions of the choices and outcomes.[23]

Namely, consider the case of an individual i attending college during T_c periods, who

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

[23]Since college graduation does not depend on learned ability except through the choices, the additive separability of this portion is clear so we omit it from this discussion.

participates in the unskilled (resp. skilled) labor market during T_u (resp. T_s) 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 $A = (A_a, A_{bs}, A_{bn}, A_s, A_u)'$. This breaks down the dependence across the grades, log wages, choices and between all of these variables. The contribution to the likelihood then writes, denoting by $(G_{i\tau})_\tau$ ($\tau \in \{1, \ldots, T_c\}$) the grades, $(w_{iu\tau})_\tau$ ($\tau \in \{1, \ldots, T_u\}$) the log wages in the unskilled sector, $(w_{is\tau})_\tau$ ($\tau \in \{1, \ldots, T_s\}$) the log wages in the skilled sector, and $(d_{i\tau})_\tau$ the decisions ($\tau \in \{1, \ldots, T_d\}$), as a five-dimensional integral:

$$\begin{aligned} & L(d_{i1}, \ldots, d_{iT_d}, G_{i1}, \ldots, G_{iT_c}, w_{iu1}, \ldots, w_{iuT_u}, w_{is1}, \ldots, w_{isT_s}) \\ = & \int L(d_{i1}, \ldots, d_{iT_d}, G_{i1}, \ldots, G_{iT_c}, w_{iu1}, \ldots, w_{iuT_u}, w_{is1}, \ldots, w_{isT_s} | A) \varphi(A) dA \end{aligned}$$

where $\varphi(.)$ 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):

$$L(d_{i1}, \ldots, d_{iT_d}, G_{i1}, \ldots, G_{iT_c}, w_{iu1}, \ldots, w_{iuT_u}, w_{is1}, \ldots, w_{isT_s}) = L_{id} \times L_{iy} \tag{18}$$

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

$$L_{id} = L(d_{i1}) L(d_{i2} | d_{i1}, G_{i1}) \ldots L(d_{iT_d} | d_{i1}, d_{i2}, \ldots, d_{iT_d - 1}, G_{i1}, G_{i2}, \ldots w_{iu1}, w_{iu2}, \ldots, w_{is1}, w_{is2}, \ldots)$$

This simply 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, unskilled and skilled log wages is given by:

$$\begin{aligned} L_{iy} = & \int L(G_{i1} | d_{i1}, A) \ldots L(G_{iT_c} | d_{i1}, d_{i2}, \ldots, A) L(w_{iu1} | d_{i1}, A) \ldots L(w_{iuT_u} | d_{i1}, d_{i2}, \ldots, A) \\ & \times L(w_{is1} | d_{i1}, A) \ldots L(w_{isT_s} | d_{i1}, d_{i2}, \ldots, A) \varphi(A) dA \end{aligned} \tag{19}$$

Where $(L(w_{iu\tau} | d_{i1}, \ldots, A))_\tau$, $(L(w_{is\tau} | d_{i1}, \ldots, A))_\tau$, and $(L(G_{i\tau} | d_{i1}, \ldots, A))_\tau$ denote the normal pdf's of the unskilled and skilled log wage as well as college grade distributions, respectively, all conditional on the ability A and the sequence of choices. Taking logs of (18) results in the choice part of the log likelihood being additively separable from the outcome (grades and log wages) part of the log likelihood.[24]

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

5.2 Estimation of grade and wage parameters

Estimation of the parameters of the outcome 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 et al. 1977). The estimation procedure iterates over the following two steps, until convergence:[25]

- 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 (8), for the posterior ability mean and covariance, given in Section 3.4.1. The (population) covariance matrix of the ability distribution is then updated as follows, for each iteration k of the EM estimation:

$$\Delta^k = \frac{1}{N}\sum_{i=1}^{N}\left(\Sigma_i^k(A) + E_i^k(A)E_i^k(A)'\right) \tag{20}$$

  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 $\Sigma_i^k(A)$ 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, skilled or unskilled labor).

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

$$\begin{aligned} El_i^k &= \int \ln(L(G_{i1}|d_{i1},A)\dots L(G_{iT_c}|d_{i1},d_{i2},\dots,A)L(w_{iu1}|d_{i1},A)\dots L(w_{uiT_u}|d_{i1},d_{i2},\dots,A))\varphi_i^k(A)dA \\ &= El_{i,a}^k + El_{i,bs}^k + El_{i,bn}^k + El_{i,s}^k + El_{i,u}^k \end{aligned} \tag{21}$$

For instance, the parameters of the unskilled wage equation are updated by maximizing the contribution $El_{i,u}^k$, which writes, denoting by $\varphi_{iu}^k(.)$ the pdf of the posterior distribution

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

of A_u:

$$El_{i,u}^k = \int (\ln(L(w_{iu1}|d_{i1}, A_u)) + \ldots + \ln(L(w_{iuT_u}|d_{i1}, d_{i2}, \ldots, A_u)))\varphi_{iu}^k(A_u)dA_u \quad (22)$$

Note that this term is additively separable over time. For any given period τ of unskilled labor market participation, it follows from the normality assumptions on the idiosyncratic productivity shocks and the unobserved ability that:

$$\int \ln(L(w_{iu\tau}|d_{i1}, d_{i2}, \ldots, A_u))\varphi_{iu}^k(A_u)dA_u =$$
$$-\frac{1}{2}\ln(2\pi\sigma_u^2) - \frac{1}{2\sigma_u^2}\left(\Sigma_{iuu}^k(A) + (w_{iu\tau} - X_{iut}\gamma_{1u} - \delta_{ut} - E_{iu}^k(A))^2\right) \quad (23)$$

where t refers to calendar time, while $E_{iu}^k(A)$ and $\Sigma_{iuu}^k(A)$ denote respectively the posterior mean and variance of the ability in the unskilled sector (computed at the E-step).[26] This equality implies that the wage equation parameters $(\gamma_{1u}, \delta_{ut})$ can be simply updated by regressing via OLS the log wages in the unskilled sector net of the posterior unskilled ability mean (which plays the role of a selection correction term) on the set of observed characteristics and calendar time dummies. The idiosyncratic shock variance (σ_u^2) is then updated as follows:

$$\sigma_u^{2,k+1} = \frac{\sum_{i,\tau}\left(\Sigma_{iuu}^k(A) + (w_{iu\tau} - X_{iut}\gamma_{1u} - \delta_{ut} - E_{iu}^k(A))^2\right)}{N_u^{\text{obs}}} \quad (24)$$

where N_u^{obs} is the total number of wage observations in the unskilled sector. Skilled wage equation parameters are updated similarly.

The updating rule above needs to be adjusted to account for the ability index specification of the grade equations along with the time-varying variances of the idiosyncratic shocks. For instance, for four-year colleges (period of enrollment $\tau > 2$), the contribution to the log likelihood writes:

$$\int \ln(L(G_{ij\tau}|d_{i1}, d_{i2}, \ldots, A_j))\varphi_{ij}^k(A_j)dA_j =$$
$$-\frac{1}{2}\ln(2\pi\sigma_{j\tau}^2) - \frac{1}{2\sigma_{j\tau}^2}\left(\lambda_{1j}^2\Sigma_{ijj}^k(A) + (G_{ij\tau} - \lambda_{0j} - \lambda_{1j}AI_{ij\tau}^k)^2\right) \quad (25)$$

where $j \in \{bs, bn\}$, $\Sigma_{ijj}^k(A)$ denotes the posterior variance of the college-j ability (computed at the E-step), $AI_{ij\tau}^k = \gamma_{0j} + X_{ij\tau}\gamma_{1j} + E_{ij}^k(A)$ is the posterior mean of the ability index

[26] t is the calendar time that corresponds to the τ-th period of unskilled labor market participation of individual i. As such, while time is simply indexed by t for notational simplicity, it should be understood as individual-specific here.

in college and major j, and $\varphi_{ij}^{k}(.)$ 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:

$$\min \sum_{i,\tau} \left(\ln(\sigma^2_{j\tau}) + \frac{1}{\sigma^2_{j\tau}} \left(\lambda^2_{1j\tau} \Sigma^k_{ijj}(A) + (G_{ij\tau} - \lambda_{0j\tau} - \lambda_{1j\tau} AI^k_{ij\tau})^2 \right) \right) \tag{26}$$

where $(\lambda_{0j\tau}, \lambda_{1j\tau}) = (0, 1)$ for $\tau \leq 2$, and $(\lambda_{0j\tau}, \lambda_{1j\tau}) = (\lambda_{0j}, \lambda_{1j})$ otherwise.

5.3 Estimation of the graduation parameters

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 11, Section 3.4.3). In the absence of type-specific unobserved heterogeneity, these parameters can be consistently estimated separately from all of the other parameters of the model. We discuss in Subsection 5.5 below how the estimation procedure needs to be adjusted to accommodate type-specific unobserved heterogeneity.

5.4 Estimation of the flow payoffs

With the estimates of the grade, wage and graduation parameters taken as given, we estimate the flow payoffs in a second stage. Following Arcidiacono & Miller (2011), we express the future payoffs in such a way that avoids solving the full backward recursion problem. Namely, the expected value function at time $t + 1$ can be expressed relative to the conditional value function for one of the choices, plus a (known) function of the choice probabilities. With the assumption that the preference shocks are distributed Type 1 extreme value, the expected value function can be expressed as:

$$E_t \left[V_{t+1} \left(Z_{it+1} \right) | d_{it} = (j, k) \right] = E_t \left[v_{j'k't+1} \left(Z_{it+1} \right) - \ln \left(p_{j'k't+1} \left(Z_{it+1} \right) \right) | d_{it} = (j, k) \right] \tag{27}$$

for any choice (j', k'), where $p_{j'k't+1}(Z_{it+1})$ is the conditional choice probability (CCP) of choosing $d_{it+1} = (j', k')$ conditional on the state variables Z_{it+1}.

Recall that in estimation it is differences in the conditional value functions that are relevant, not the conditional value functions themselves. Consider any choice $d_{it} = (j', k')$ as well as the choice $d_{it} = (0, 0)$ (home production). Given these initial choices and the structure imposed by the model, it is easy to show that there exists two sequences of choices such that the probabilities associated with each of the three-period-ahead states are the same after making the choices in the sequences.[27] It follows that:

$$\begin{aligned} &E_t\left[V_{t+3}(Z_{it+3})|d_{it}=(0,0), d_{it+1}=(j',k'), d_{it+2}=(0,0)\right] = \\ &E_t\left[V_{t+3}(Z_{it+3})|d_{it}=(j',k'), d_{it+1}=(0,0), d_{it+2}=(0,0)\right] \end{aligned} \tag{28}$$

The finite dependence property holds here because, conditional on the same lagged decision, the sequence of the previous choices does not affect the current state.

We can then reformulate the problem in terms of two-period ahead flow payoffs and conditional choice probabilities, and then estimate the CCPs in a first stage. The differenced conditional value function for choices prior to college graduation is:[28]

$$v_{jkt}(Z_{it}) - v_{00t}(Z_{it}) = \begin{pmatrix} u_{jk}(Z_{it}) - \beta E_t\left(\ln\left[p_{00t+1}(Z_{t+1})\right]|Z_{it}, d_{it}=(j,k)\right) \\ +\beta E_t\left(\ln\left[p_{jkt+1}(Z_{t+1})\right] - u_{jk}(Z_{t+1})|Z_{it}, d_{it}=(0,0)\right) \\ +\beta^2 E_t\left(\ln\left[p_{00t+2}(Z_{t+2})\right]|Z_{it}, d_{it}=(0,0), d_{it+1}=(j,k)\right) \\ -\beta^2 E_t\left(\ln\left[p_{00t+2}(Z_{t+2})\right]|Z_{it}, d_{it}=(j,k), d_{it+1}=(0,0)\right) \end{pmatrix} \tag{29}$$

Estimation of the flow utility parameters then involves the following steps:

1. Estimate the CCPs via a flexible multinomial logit model.

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, 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.[29]

[27] For a full characterization of models that have the finite dependence property see Arcidiacono & Miller (2015*b*).

[28] The differenced conditional value function is normalized with respect to $v_{0pt}(Z_{it})$, i.e. no school and part-time work, for all choices after college graduation. We do so since home production is rarely chosen among college graduates. Finite dependence still holds in this case, but the expression for $v_{jkt}(Z_{it}) - v_{0pt}(Z_{it})$ (available from the authors on request) is more complex because $u_{0p}(\cdot)$ is not equal to zero, whereas $u_{00}(\cdot)$ is normalized to zero.

[29] Another advantage of applying this approach is that we do not have to make assumptions about beliefs

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 $t = 1$. Accounting for type-specific unobserved heterogeneity breaks down the separability between the choice and outcome components of the likelihood described above as our full log likelihood function becomes:

$$\sum_i \ln\left[\sum_{r=1}^{R} \pi_r L_{idr} L_{ibr} L_{iyr}\right] \tag{30}$$

where π_r denotes the population probability of being of type r, and L_{idr}, L_{ibr}, and L_{iyr} respectively denote individual i's contribution to the likelihood of (i) the choices, (ii) four-year college graduation outcomes, and (iii) grade and wage outcomes $y = (G, w_u, w_s)'$, all conditional on the unobserved heterogeneity type r.

Following Arcidiacono & Miller (2011), we use an adaptation of the EM algorithm that restores the additive separability of the likelihood function. Rather than updating the structural parameters of the decision process at each step, we use their two-stage approach and approximate the decision process with a reduced form. Specifically, let L^*_{idr} give the reduced-form choice likelihood conditional on being of type r. The posterior probability of i belonging to the r-th type follows from Bayes' rule:

$$q_{ir} = \frac{\pi_r L^*_{idr} L_{ibr} L_{iyr}}{\sum_{r'=1}^{R} \pi_{r'} L^*_{idr'} L_{ibr'} L_{iyr'}} \tag{31}$$

In the first stage we recover the parameters of the grade and wage processes, the graduation probabilities, the (type-specific) CCPs, and the conditional probabilities of being of each type. The second stage boils down to a weighted multinomial logit with an offset term. Note that this is identical to the case without unobserved heterogeneity except that now the q_{ir}'s are used as weights.[30]

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.

[30]The CCPs are identified from the data and could in principle be estimated nonparametrically. However, we choose to estimate them using a parametric specification to avoid the curse of dimensionality. We use different specifications depending on the graduation status. In the first estimation stage and for those who have not graduated from college, we use a quadratic in age and each type of work and schooling experience. We also include year dummies, a running mean for each of the outcome variables, previous decision dummies, and heterogeneity type interacted with work and schooling experiences. We impose more structure for college

Finally, standard errors are estimated using a parametric bootstrap procedure with 150 replications.[31] We discuss this procedure in detail in Appendix C.

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 after one period. These individuals likely received negative grade realizations which, if ignored, could bias our results. We take this issue into account within our estimation procedure by treating the first instance of unobserved grades or major as another unobserved discrete latent variable. In practice we discretize the observed grades distribution into quartiles, and approximate the distribution of unobserved grades with a discrete distribution with finite support given by these quartiles.[32] The estimation procedure 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 within the first stage of our estimation procedure. The log likelihood which is maximized at the M-step is now conditional on both the unobserved heterogeneity type as well as the major or grade quartile.[33] The implementation of this procedure is more fully described in Appendix B.

6 Results

In this section, we present and discuss the estimation results. All of the results discussed below are obtained assuming the existence of $R = 2$ unobserved heterogeneity types. Type 1

graduates to avoid overfitting. In particular, we impose additive separability between the graduate school and work options, and do not include interaction terms between heterogeneity types and experiences. Finally, we impose additional linearity restrictions on the experience variables in the CCPs that are used in the second estimation stage.

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

[32] The corresponding cut points of the grades distribution occur at $\{0, 2.5, 3.0, 3.6, 4.0\}$.

[33] 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.

(respectively type 2) individuals account for 50.3% (resp. 49.7%) of the overall population.[34]

6.1 Grade parameters

The parameter estimates for the grade equations are presented in Table 5. Since high school grades and SAT scores have been standardized, the coefficients associated with these characteristics represent the effects of one standard deviation increases. For all the schooling options, high school grades are associated with larger effects on college grades than either of the test score components. SAT verbal matters more for college grades than SAT math, with the exception of four-year science where the effects are similar.

Working while in college is associated with lower grades for both types of colleges and majors, although the effects are modest and generally not statistically significant at standard levels. Interestingly and consistent with the findings of Hansen et al. (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. Finally, turning to the type-specific unobserved ability (known to the agent), students in the type 1 group have lower GPA in two-year colleges, and higher GPA in both science and non-science majors in four-year colleges.

6.2 Wage parameters

Estimates of the wage equations are given in Table 6.[35] For both sectors we find significant and sizable returns to a one-standard deviation increase in SAT math scores (6% and 11% in the unskilled and skilled sectors, respectively). On the other hand, the returns to SAT verbal are negative for both sectors. The latter finding echoes several other papers documenting a negative effect of SAT verbal scores after controlling for SAT math scores (see, e.g., Arcidiacono 2004, Kinsler & Pavan 2015, Sanders 2015). High school grades are significant in the skilled sector only, both statistically and economically with a one-standard deviation increase in high school GPA being associated with about 8% higher wages in that sector. Overall, the differences in returns in the labor market for the different ability measures versus the returns to these measures in terms of of college grades (see Section 6.1) provide evidence that the skills necessary to succeed in college do not fully line up with those that are rewarded in the

[34]Results for the specification without type-specific unobserved heterogeneity are available from the authors upon request.

[35]Estimates of the AR(1) processes governing the aggregate shocks are reported in Table A.5.

labor market.

Returns to experience are slightly larger in the skilled sector (8%) than in the unskilled sector (6%), although both returns are of similar order of magnitude. Experience in the unskilled sector does not translate into higher earnings in the skilled sector.[36] This highlights the importance of accounting for sector-specific labor market experience in this context, as both types of experience are rewarded very differently in the labor market.

Returns to schooling in the unskilled sector are positive and significant but relatively small, averaging about 4% a year. Working while in college (as opposed to working without being enrolled in college) results in a substantial wage loss, particularly for part-time work in both two-year and four-year colleges. The part-time penalty itself, however, is higher in the skilled sector.

Turning to the skilled sector, graduating from a science (as opposed to non-science) major increases on average the wage in the skilled sector by 14%. While economically significant, it is interesting to note that this estimate is lower than most of the science major premia which have been obtained in earlier studies, including those based on structural models of college major choice (Altonji, Arcidiacono & Maurel 2016). This finding may partly reflect the fact that our estimates control for selection on ability using a choice model that is more flexible in what is included in the individual's information set.[37] The estimates also suggest positive returns to graduate school, though these disappear after the first year.

Finally, all else equal, blacks have significantly lower wages (9.1% penalty) in the unskilled sector, but no penalty in the skilled sector, a finding that is consistent with Arcidiacono (2004) and Arcidiacono et al. (2010). Turning to unobserved type-specific heterogeneity, type 1 individuals have substantially higher wages than type 2 individuals in the skilled sector (17% premium) but lower wages in the unskilled sector (11% penalty). These results are consistent with the existence of a sector-specific comparative advantage, which is part of the information set of the agents as they start making their college decisions.

[36]This finding may partly reflect the fact that, after graduating from college, individuals who have accumulated more experience in the unskilled sector might also be more likely to end up working in a relatively low-paying occupation. Investigating this issue would require modeling the choice of occupation.

[37]Indeed, our specification allows individual choices to depend on the information about their ability that is revealed over time, in addition to the more standard source of unobserved heterogeneity which is permanent.

6.3 Learning

Table 7 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. Indeed, eight out of the ten correlation coefficients are significantly different from one at the 1% level, while the other two correlation coefficients (across science and non-science majors in four-year colleges, and across two-year colleges and four-year colleges non-science majors) are significantly different from one at the 10% level only. A unidimensional skill model, or even a model with two imperfectly correlated skills (schooling ability and labor market productivity), are unambiguously rejected by the data. As such, these results add to a large and growing empirical literature providing evidence that skills are multidimensional in nature (see Heckman & Mosso 2014, and multiple references therein).

Schooling ability is highly correlated across college types and majors. The correlations across college types and majors range from 0.7 (for four-year college science majors and two-year colleges) to 0.9 (for four-year colleges non-science majors and four-year colleges science majors). A similar picture emerges across skilled and unskilled sectors within the labor market, which are strongly correlated (estimated correlation coefficient of 0.78).

The correlations between schooling abilities and labor market productivity are all positive, but markedly lower than the correlations across college types and majors. Notably, the correlations between ability in four-year college non-science or science major, and productivity in the skilled sector are very small (between 0.05 and 0.06). Correlations between schooling abilities in four-year colleges and labor market productivity are larger for the unskilled sector, although both of them remain below 0.21. Taken together, these patterns provide clear indication that grades earned in college, in science as well as in non-science majors, reveal little new information about future labor market performance. Correlations are larger for two-year colleges, possibly reflecting the fact that courses and examinations in two-year colleges tend to be more geared toward professional life than in four-year colleges. However, these correlations remain much smaller than the correlations across schooling options or between skilled and unskilled labor market sectors.

The variances of each of the ability measures are given in the bottom row of Table 7. 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 13% of the variance of grades in two-year college, to as much as 38.7% of the variance of log wages in the skilled sector. Among the schooling options, the share is highest for four-year sciences (35.1%), highlighting the important role of unobserved ability in science majors. Overall, these results provide evidence that it is important in this context to allow for a portion of ability to be imperfectly known by the individuals. Even in the case with the smallest variance—ability in the unskilled sector—a one standard deviation increase in ability would translate into a 27% increase in wages.

While the unknown ability component is large, learning may still take time due to the noise of the signals. Table 8 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. This is particularly true in the first year of college where the signal-to-noise ratios are 0.27, 0.20 and 0.09 only for four-year science, four-year non-science, and two-year respectively.[38] Signals are much more informative in the second year of college with the corresponding signal to noise ratios of 0.59, 0.44, and 0.19. The signal-to-noise ratios for the skilled and unskilled sectors are 0.46 and 0.31, respectively.

6.4 Graduation parameters

Estimates of the parameters governing graduation are given in Table A.6. Recall that no one is at risk of graduating until their third year of undergraduate education, and individuals are only at risk of graduating if they are attending a four-year college. As expected, more time in school makes graduation more likely, especially time at a four-year institution.

More interesting are the major-specific components. All else equal, being a science major delays graduation. The degree to which it delays graduation, however, depends on one's abilities in the sciences and non-sciences. Having higher abilities in one's subject area increases graduation probabilities but does so more in the sciences. Since science ability also has a higher variance, this again underscores the importance of abilities in the sciences.

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

6.5 Flow payoffs

Table 9 reports the structural parameter estimates of the flow utility parameters. Higher SAT scores, both math and verbal, make the four-year non-science option more attractive, while only SAT math scores are important for four-year science enrollment. Higher high school grades make all four-year options more attractive. While none of these measures has a direct effect on the payoff to working (the coefficients are all small and insignificant), they still have an indirect effect on the payoff to working through higher wages.

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. The effects on the four-year options, as well as the effects of SAT scores and high school grades, suggest lower costs of effort when these measures are high. Ability in the workforce enters through log wages. Although the coefficient is positive and large, it is imprecisely estimated.

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. Finally, type 1 individuals are found to have higher preferences for four-year colleges and a relative distaste for two-year colleges. Together with the estimation results obtained for the grade equations, the latter estimates point to a positive correlation between preferences and abilities for both types of colleges.

The bottom panel of Table 9 gives the estimated probability of missing grades being in each quartile of the grade distribution, conditional on the (first) occurrence of a missing grade. If grades were missing completely at random, we would expect these probabilities to be close to a quarter. Since individuals are more likely to leave college when they receive low grades, and individuals are more likely to not report their grades when they leave college, it is not surprising that the conditional distribution is shifted towards lower grades.

6.6 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. Specifically, we begin by drawing an ability vector for each individual from the population distribution (a multivariate normal with mean zero and covariance $\hat{\Delta}$).[39] We then draw an unobserved type for each individual from a Bernoulli distribution with parameter $\hat{\pi}_1$ (estimated unobserved type probability). Next, we draw preference shocks and compute choice probabilities using the observed states (i.e. the demographic characteristics and heterogeneity type and ability drawn at the beginning of the simulation), the structural flow utility estimates and the reduced-form CCPs to represent the future value term.[40] 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 $T = 10$ times.[41] We perform this forward simulation 100 times for each individual in the estimation sample.

Tables 10 and 11 report the fit of the model in several relevant dimensions. Table 10 reports, for each period, the empirical frequency (Data column) and the choice frequency predicted using the model (Model column) for four different events, namely college entry, college attrition, college re-entry, and graduation.

Overall, most of the predicted choice frequencies are reasonably close to the empirical ones, despite some non-trivial discrepancies in a couple of cases. Importantly, our model does a good job in predicting the dynamics of these choices across all periods. Table 11 shows the fit of the model in terms of choice frequencies, pooled across all periods. The fit is generally very good, with the predicted choice frequencies being in most cases very close to the empirical frequencies.

Turning to the ability sorting patterns, Table 12 shows the posterior mean of each unobserved ability, for different choice paths. These results are obtained by forward simulating 100 times, for each individual in the sample, the outcomes and sequences of choices.[42]

Though sorting effects are relatively small, the signs are generally in the expected direction. Those who go continuously to college and graduate with a degree in science have relatively high (posterior) ability in science, but also, to a lesser extent, in non-science ma-

[39] The parameter values of the correlation matrix associated with $\hat{\Delta}$ are listed in Table 7.

[40] 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.

[41] 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.

[42] Note that one could alternatively construct a similar table by using the learning estimates only and then compute the average posterior abilities for those who chose particular paths. Using forward simulations enables us to work with larger cell sizes.

jors and two-year colleges. Individuals who are continuously enrolled in college and graduate with a non-science degree also have larger abilities in all types of schools and majors than dropouts and stopouts, as well as individuals who never attend college. However, it is interesting to note that, among the group of individuals who do not work while in college, science college graduates are on average not only better than non-science college graduates in science, but they also have higher non-science posterior ability. This pattern points to four-year college graduates in sciences having an absolute advantage in both types of majors.

Those who go continuously to college and work while in school at all periods have higher unskilled (and skilled) labor market productivities than those who never work while in school. Individuals who stop out, but then graduate from college in science have lower schooling ability in four-year sciences than the continuous enrollees who work in college and obtain a degree in science. On the other hand, these stopouts who graduate have substantially larger schooling abilities, for all types of college and majors, than the individuals who stopped out but then left college again without graduating. Individuals for whom dropping out was an absorbing state have on average higher schooling abilities than those who stop out and drop out from science or non-science majors, but substantially smaller abilities than those who graduate after stopping out.

In order to investigate how ability uncertainty gets resolved, Table 13 reports the posterior ability variances for different choice paths at either (*i*) time of college graduation (for those who do graduate) or (*ii*) in the last survey period (for those who do not). Focusing on the first two panels (continuous enrollees graduating in science or non-science), the results show that a fair amount of ability learning takes place while in college. On average among all college graduates in science, the posterior ability variance in science is about six times smaller upon graduation than when they graduated from high school. Consistent with the existence of a smaller signal-to-noise ratio in non-science majors, learning is somewhat slower but nonetheless sizable in non-science fields, with the posterior ability variance being more than four times smaller upon graduation than those who do not attend college. Individuals also learn about their abilities in other types of colleges and majors than the one they graduate from. Learning across different schooling options is a non-trivial source of information revelation. For instance, the posterior non-science ability variance is three and a half times smaller upon graduation from a science major than for those who do not attend college.[43]

[43]Table A.7 in Appendix A reports the posterior ability variances at the end of the panel (2011), broken out by background characteristics. The speed of ability learning is quite heterogeneous across those different subgroups, particularly so for the academic abilities. This finding reflects the fact that individuals with

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 the following, we simulate individual choices and labor market outcomes under a counterfactual scenario where we assume that individuals have perfect information on their abilities once they leave high school. To do this, 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, unskilled productivity, and skilled productivity—which individuals are now assumed to know when making their educational and labor supply decisions. Three sources of uncertainty remain: the probability of graduating from a four-year college conditional on attendance, aggregate labor market shocks to both sectors, and individual preference shocks. 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.

7.1 Information and educational choices

Table 14 reports the college completion status frequencies in the baseline and counterfactual scenario (referred to as full- or perfect-information scenario in the following). The fraction of individuals who never attend college is virtually unchanged. However, the predicted share of individuals graduating from four-year college would increase by about nine percentage points (a 38% increase) if they had perfect information on their ability by the end of high school.[44] Of the nine percentage point increase, over eight percentage points is due to the

different background characteristics tend to choose different paths in college and in the labor market, some of these paths carrying significantly more information about individual abilities than others. As such, the posterior variances reported in this table point to sizable informational inequalities by race, parental education and ability (as measured by SAT scores and high school GPA). In particular, blacks and Hispanics, individuals below the median of the SAT math, SAT verbal and high school GPA distributions, as well as those whose parents have not graduated from college are on average more uncertain about their abilities.

[44] Stinebrickner & Stinebrickner (2014*a*) show that learning about grades accounts for 45% of dropouts in the first two years, a finding based on predictions had students received their prior abilities as their grade signals. The corresponding numbers from providing full information in our case result in smaller percentage changes in the drop out rates. This is partly due to differences driven by receiving full information rather than one's prior, focusing on the first two years of college as opposed to all years of enrollment, as well as differences in the sample composition (the full population versus enrollees at Berea). See also Stinebrickner & Stinebrickner (2014*b*) who show that ability learning plays a particularly important role in explaining

increase in the share of individuals who attend college continuously.

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 skilled productivity and choose to enroll in college continuously. These countervailing forces result in the number of individuals never attending college remaining the same but the number of college graduates substantially increasing.

We next investigate how the counterfactual scenario changes the ability compositions across the different choice paths. Table 15 replicates Table 12, but with the abilities now calculated based on the counterfactual simulations. Comparing how much ability sorting there is in this counterfactual full information scenario with the baseline sorting patterns discussed above (Table 12) speaks to the cost of imperfect information in this context. In most cases, the sorting effects go in the same direction as the ones obtained earlier without assuming perfect information. However, the magnitudes of these effects are generally much larger. In particular, we find much stronger evidence of sorting on comparative advantage in the perfect information scenario. For instance, individuals who do not work while in school, are continuously enrolled, and graduate from a science major in a four-year college have on average a 0.86 standard deviation higher science ability than those who never enroll in college. The ability differences are much smaller in the baseline scenario, where the continuous enrollees in four-year science have a 0.25 standard deviation higher science ability than those who never enroll in college.

It is also interesting to compare sorting on science and non-science abilities across both types of majors, in the baseline and in the counterfactual scenario. In the full information scenario, among those who do not work while in school, continuous enrollees in science majors have a 0.18 standard deviation higher science ability than the continuous enrollees in non-science majors. Those differences are smaller in the baseline scenario (0.09 standard deviation), which provides evidence that imperfect information limits the extent to which individuals sort across college majors based on their comparative advantage.

---

attrition in science majors.

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. Panel (a) of Table 16 shows mean log wages in the skilled and the unskilled sectors under two different scenarios. Column (1) of Table 16 reports our baseline case, which serves as a benchmark. Columns (2) and (3) provide the counterfactual full information cases with the only difference being that in the latter case the level of labor market experience is held constant with respect to the baseline. The results reported in columns (1) and (2) indicate that, in the presence of full information, wages in the skilled sector increase on average by approximately 16.7%, while in the unskilled sector wages decrease by 16%.

As wages in the skilled sector increase while wages in the unskilled sector decrease, the college wage premium increases by 32.7 percentage points in the full-information scenario.[45] In order to understand to what extent part of this increase can be explained by differences in the experience profiles between the baseline and counterfactual cases, as opposed to the direct effect of information through ability sorting across sectors, column (3) presents counterfactual log wages when experience is held constant. Results show that the college wage premium increases by 19.8 percentage points in this case, thus indicating that a significant portion of the increase in the college wage premium in the full information scenario occurs through the accumulation of sector-specific labor market experience. Taken together, these results provide evidence that imperfect information about ability plays an important role in the composition of the workforce in both sectors. Eliminating these informational frictions would result in a large increase in the average college wage premium.[46]

[45] Recall that the unskilled sector includes all the individuals who have enrolled in two- and four-year colleges but have not obtained a bachelor's degree, while the skilled sector also includes individuals who enrolled in graduate school after graduating from college. As such, the college wage premia reported in Table 16 are also affected by the returns to graduate school and to each year of college. If instead we were to compute the college wage premium as the effect of graduating from a four-year college only (without graduate education) relative to high school only (without college education), this would result in a baseline premium of 26.9% instead of 24.7% which is reported in the table. The fact that the estimated college premium is relatively low is mainly due to the early point of evaluation in the life cycle, with workers in the skilled sector averaging less than 3.5 years of skilled experience. In particular, the estimated premium goes up to about 37% after 4 additional years of experience.

[46] The predicted decline in the wages in the unskilled sector is compatible with individuals making more informed choices in the counterfactual scenario. Even in the case of the standard Roy model where sorting across sectors is purely based on monetary gains, mean wages can decrease in one of the two sectors if agents

Panel (b) of Table 16 further examines the extent to which wages in the two sectors are different in the baseline versus full information scenario as a result of differences in observable characteristics. First, we calculate, for the baseline and the full information scenarios, the mean (counterfactual) wages in the skilled sector for high school graduates, conditional on having on average the same experience, choice of major, and graduate school enrollment decisions as those who graduated from college. Similarly, we also compute the wages in the unskilled sector for college graduates, conditional on having on average the same experience and college enrollment decisions as those who only graduate from high school. Results indicate that, in the baseline scenario, high school graduates would earn on average in the skilled sector 17.6% less than college graduates working in that sector. In the counterfactual scenario, the gap goes up to 39%. The larger difference in the counterfactual scenario highlights how providing more information to individuals would amplify sorting into the different sectors. In a similar vein, the decomposition shows that wages for high school graduates, had they instead chosen the skilled sector but holding their characteristics fixed, would have been higher in the baseline than in the counterfactual scenario. The opposite holds for college graduates in the unskilled sector. This asymmetry captures the essence of the sorting mechanism which is driven by the fact that ability is highly correlated across sectors, and that individuals who graduate from college tend to be more productive in both sectors, particularly so in the full information case.

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 would result in a sizable increase in four-year college graduation rates. 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, leading in turn to a substantial increase in the college wage premium.

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 longitudinal data from the NLSY97, we estimate

sort across sectors according to their comparative advantage, as opposed to a random allocation (Heckman & Honoré 1990). In our model, non-monetary components of individual preferences is an additional channel through which mean wages can decrease in a given sector.

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 that four-year college graduation rates would increase substantially (by 38%) relative to the baseline imperfect information scenario, mostly through a decrease in dropout rates. Imperfect information on ability also has significant implications regarding the composition of college graduates in science and non-science majors, dropouts and stopouts. In particular, we find evidence that imperfect information on ability acts as a barrier to the pursuit of comparative advantage through schooling choices. Providing full information to the students would result in significant changes in the average productivity levels of workers within each sector, leading in turn to a substantial increase in the college wage premium. As such, our results suggest that providing more opportunities to learn about own ability before starting college,[47] as has been recently implemented in different contexts (Bobba & Frisancho 2016, Pistolesi 2016), would have significant implications in terms of schooling decisions and on the relative skill composition of the college versus non-college educated workforce.

[47]This could include such policies as encouraging more enrollment in advanced placement or concurrent enrollment courses, arranging for high school students to audit a sample of college courses, as well as encouraging students to work before enrolling in college.

References

Adda, J., Dustmann, C., Meghir, C. & Robin, J.-M. (2010), Career progression and formal versus on-the-job training. Working paper.

Altonji, J., Arcidiacono, P. & Maurel, A. (2016), The analysis of field choice in college and graduate school: Determinants and wage effects, *in* E. Hanushek, S. Machin & L. Wößmann, eds, 'Handbook of the Economics of Education', Vol. 5, Elsevier.

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

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

Antonovics, K. & Golan, L. (2012), 'Experimentation and job choice', *Journal of Labor Economics* **30**, 333–366.

Arcidiacono, P. (2004), 'Ability sorting and the returns to college major', *Journal of Econometrics* **121**, 343–375.

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

Arcidiacono, P., Bayer, P. & Hizmo, A. (2010), 'Beyond signaling and human capital: Education and the revelation of ability', *American Economic Journal: Applied Economics* **2**(4), 76–104.

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

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

Arcidiacono, P. & Miller, R. (2015*a*), Identifying dynamic discrete choice models off short panels. Working paper.

Arcidiacono, P. & Miller, R. (2015*b*), Nonstationary dynamic models with finite dependence. Working paper.

Beffy, M., Fougere, D. & Maurel, A. (2012), ‘Choosing the field of study in postsecondary education: Do expected earnings matter?’, *Review of Economics and Statistics* **94**, 334–347.

Belzil, C. & Hansen, J. (2002), ‘Unobserved ability and the return to schooling’, *Econometrica* **70**(5), 2075–2091.

Bobba, M. & Frisancho, V. (2016), Learning about oneself: The effects of signaling academic ability on school choice. Working paper.

Bound, J., Lovenheim, M. & Turner, S. (2010), ‘Why have college completion rates declined? an analysis of changing student preparation and collegiate resources’, *American Economic Journal: Applied Economics* **2**, 129–157.

Bound, J. & Turner, S. (2011), Dropouts and diplomas: The divergence in collegiate outcomes, *in* E. Hanushek, S. Machin & L. Wößmann, eds, ‘Handbook of the Economics of Education’, Vol. 4, Elsevier.

DeGroot, M. (1970), *Optimal Statistical Decisions*, New York: McGraw Hill.

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

Goldin, C. & Katz, L. F. (2008), *The Race Between Education and Technology*, Belknap/Harvard University Press, Cambridge, Mass.

Hansen, K. T., Heckman, J. J. & Mullen, K. J. (2004), ‘The effect of schooling and ability on achievement test scores’, *Journal of Econometrics* **121**, 39–98.

Hastings, J., Neilson, C. A., Ramirez, A. & Zimmerman, S. (2016), ‘(Un)informed college and major choice: Evidence from linked survey and administrative data’, *Economics of Education Review* **51**(1), 136–151.

Heckman, J. J. & Honoré, B. E. (1990), ‘The empirical content of the roy model’, *Econometrica* **58**, 1121–1149.

Heckman, J. J., Layne-Farrar, A. & Todd, P. (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, J. J., Lochner, L. J. & Todd, P. E. (2006), Earnings functions, rates of return and treament effects: the Mincer equation and beyond, *in* E. Hanushek & F. Welch, eds, 'Handbook of the Economics of Education', Vol. 1, Elsevier.

Heckman, J. J. & Mosso, S. (2014), 'The economics of human development and social mobility', *Annual Review of Economics* **6**, 689–733.

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

Heckman, J. J. & Urzúa, S. (2009), The option value of educational choices and the rate of return to educational choices. Working paper.

James, J. (2011), Ability matching and occupational choice. Working paper.

Joensen, J. S. (2009), Academic and labor market success: The impact of student employment, abilities and preferences. Working paper.

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

Keane, M. P. & Wolpin, K. I. (2000), 'Eliminating race differences in school attainment and labor market success', *Journal of Labor Economics* **18**, 614–652.

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

Kinsler, J. & Pavan, R. (2015), 'The specificity of general human capital: Evidence from college major choice', *Journal of Labor Economics* **33**(4), 933–972.

Magnac, T. & Thesmar, D. (2002), 'Identifying dynamic discrete decision processes', *Econometrica* **70**, 801–816.

Manski, C. F. (1989), 'Schooling as experimentation: A reappraisal of the postsecondary dropout phenomenon', *Economics of Education Review* **8**, 305–312.

Manski, C. F. & Wise, D. A. (1983), *College Choice in America*, Harvard University Press.

McLachlan, G. & Peel, D. (2004), *Finite Mixture Models*, John Wiley and Sons.

Miller, R. (1984), 'Job matching and occupational choice', *The Journal of Political Economy* **92**, 1086–1120.

Pavan, R. (2011), 'Career choice and wage growth', *Journal of Labor Economics* **29**(3), 549–587.

Pistolesi, N. (2016), Advising students on their field of study: Evidence from a French university reform. Working paper.

Pugatch, T. (2012), Bumpy rides: School to work transitions in south africa. IZA Discussion Paper No. 6305.

Robin, J.-M. (2011), 'On the dynamics of unemployment and wage distributions', *Econometrica* **79**, 1327–1355.

Rust, J. (1994), Estimation of dynamic structural models, problems and prospects: Discrete decision processess, *in* C. Sims, ed., 'Advances in Econometrics: Sixth World Congress', Vol. 2, Cambridge University Press, New York, pp. 5–33.

Sanders, C. (2010), Skill uncertainty, skill accumulation, and occupational choice. Working paper.

Sanders, C. (2015), Reading skills and earnings: Why do doing words good hurt you're wages? Working paper.

Stange, K. M. (2012), 'An empirical investigation of the option value of college enrollment', *American Economic Journal: Applied Economics* **4**, 49–84.

Stinebrickner, T. & Stinebrickner, R. (2003), 'Working during school and academic performance', *Journal of Labor Economics* **21**, 473–491.

Stinebrickner, T. & Stinebrickner, R. (2012), 'Learning about academic ability and the college drop-out decision', *Journal of Labor Economics* **30**, 707–748.

Stinebrickner, T. & Stinebrickner, R. (2014*a*), 'Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model', *Journal of Labor Economics* **32**, 601–644.

Stinebrickner, T. & Stinebrickner, R. (2014*b*), 'A major in science? initial beliefs and final outcomes for college major and dropout', *Review of Economic Studies* **81**, 426–472.

Wiswall, M. & Zafar, B. (2015), 'Determinants of college major choice: Identification using an information experiment', *Review of Economic Studies* **82**(2), 791–824.

Wiswall, M. & Zafar, B. (2016), Preference for the workplace, human capital, and gender. NBER Working Paper No. 22173.

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	0.212	0.129	0.208	0.156	0.299	0.228
	(0.409)	(0.336)	(0.406)	(0.363)	(0.458)	(0.420)
Hispanic	0.211	0.112	0.108	0.128	0.208	0.177
	(0.408)	(0.317)	(0.311)	(0.334)	(0.406)	(0.381)
SAT math	-0.252	0.488	0.121	0.216	-0.47	-0.157
	(0.635)	(0.789)	(0.736)	(0.758)	(0.640)	(0.751)
SAT verbal	-0.225	0.252	0.131	0.142	-0.477	-0.179
	(0.646)	(0.750)	(0.787)	(0.783)	(0.715)	(0.770)
HS Grades	0.035	0.889	0.626	0.689	-0.259	0.168
	(0.856)	(0.833)	(0.807)	(0.802)	(0.893)	(0.951)
Parent graduated college	0.263	0.579	0.527	0.488	0.102	0.296
	(0.441)	(0.495)	(0.500)	(0.500)	(0.303)	(0.457)
Family Income ($1996)	36,789	53,070	54,204	54,949	28,394	40,036
	(36,778)	(51,502)	(56,847)	(54,878)	(30,241)	(43,741)
Total N	760	178	332	461	982	2,713

Notes: This table reports summary statistics for the subsample of the NLSY97 that is used to estimate our structural model. Test scores are standardized to the NLSY97 population. Grades are standardized to the NLSY97 male population. Standard deviations are listed directly below the mean (in parentheses) for each entry. See Table A.2 for complete details on sample selection.

Table 2: 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	1.97	43.26	4.22	13.45	9.71
Continuous completion (CC), grad. Hum	7.11	13.48	45.48	38.83	23.57
Stopped out (SO), graduated Sci	1.84	2.25	1.51	0.87	1.56
Stopped out (SO), graduated Hum	4.08	3.93	6.63	6.72	5.26
Stopped out (SO) then dropped out	16.84	4.49	7.23	8.24	11.44
Dropped out (DO)	56.71	20.22	24.70	22.56	37.72
CC right censored	0.79	5.06	4.52	4.77	3.00
SO right censored	10.66	7.30	5.72	4.56	7.74
Total N	760	178	332	461	1,731

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 A.2 for complete details on sample selection.

Table 3: Period-t GPA outcomes (by $t+1$ period college decision)

(a) 4-year Students, GPA levels

	Mean GPA	Std Dev	N	$\|t\text{-stat}\|$
Drop out from 4-year college	2.801	0.716	294	7.65
Stay	3.099	0.619	2,433	
Switch majors (or to 2-year college)	2.925	0.690	176	3.36
Stay	3.089	0.615	2,340	

(b) 2-year Students, GPA levels

	Mean GPA	Std Dev	N	$\|t\text{-stat}\|$
Drop out from 2-year college	2.834	0.878	317	4.37
Stay	3.052	0.678	748	
Switch to 4-year college (any major)	3.135	0.627	99	2.06
Stay	2.972	0.759	966	

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

	Mean residual	Std Dev	N	$\|t\text{-stat}\|$
Drop out from 4-year college	-0.188	0.686	263	5.64
Stay	0.021	0.555	2,253	
Switch majors (or to 2-year college)	-0.088	0.667	176	2.10
Stay	0.007	0.565	2,340	

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

	Mean residual	Std Dev	N	$\|t\text{-stat}\|$
Drop out from 2-year college	-0.136	0.839	317	4.13
Stay	0.058	0.633	748	
Switch to 4-year college (any major)	0.086	0.573	99	1.27
Stay	-0.009	0.718	966	

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, and work intensity dummies.

Table 4: 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	2.424	0.557	3,152	4.41
Return to school	2.242	0.418	188	
Total	2.414	0.552	3,340	

(b) Log wage *residuals*

	Mean residual	Std Dev	N	$\|t\text{-stat}\|$
Stay in work	0.061	0.520	3,152	2.10
Return to school	-0.020	0.403	188	
Total	0.057	0.514	3,340	

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; experience; age; in-school work dummies; and work intensity dummies.

Table 5: 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
Constant	3.070	(0.141)	2.908	(0.083)	2.779	(0.063)
Black	-0.136	(0.119)	-0.073	(0.075)	-0.200	(0.056)
Hispanic	0.013	(0.123)	-0.093	(0.069)	-0.027	(0.053)
SAT math	0.135	(0.067)	0.081	(0.047)	0.003	(0.039)
SAT verbal	0.121	(0.072)	0.125	(0.045)	0.087	(0.043)
HS Grades	0.212	(0.059)	0.208	(0.031)	0.168	(0.027)
Parent graduated college	0.011	(0.092)	0.066	(0.050)	0.017	(0.051)
Work FT	-0.100	(0.069)	-0.072	(0.041)	-0.017	(0.051)
Work PT	-0.040	(0.054)	-0.046	(0.030)	-0.076	(0.044)
Age 18 and under	-1.005	(0.088)	-0.652	(0.055)	-0.391	(0.056)
Age 19	-0.592	(0.078)	-0.300	(0.052)	-0.255	(0.067)
Age 20	-0.252	(0.070)	-0.192	(0.041)	-0.328	(0.061)
Age 21	-0.113	(0.057)	-0.139	(0.038)	-0.086	(0.060)
Year 2+					0.430	(0.041)
λ_0 (ability index intercept)	0.118	(0.217)	0.251	(0.161)	0.000	(—)
λ_1 (ability index loading)	0.901	(0.063)	0.928	(0.051)	1.000	(—)
Unobserved type 1	0.058	(0.151)	0.052	(0.094)	-0.078	(0.063)
Person-year obs.	1,045		2,248		1,570	

Notes: Bootstrap standard errors in parentheses. Age 22 and older is the reference category for the age dummies. Not working while in school is the reference category for the work intensity dummies.

Table 6: Estimates of Skilled and Unskilled Wage Parameters

	Skilled		Unskilled	
	Coeff.	Std. Error	Coeff.	Std. Error
Constant	2.083	(0.134)	1.981	(0.057)
Black	0.032	(0.042)	-0.091	(0.018)
Hispanic	-0.020	(0.053)	-0.010	(0.019)
SAT math	0.109	(0.026)	0.062	(0.013)
SAT verb	-0.047	(0.027)	-0.015	(0.013)
HS Grades	0.079	(0.022)	-0.008	(0.008)
Parent graduated college	-0.009	(0.033)	-0.004	(0.018)
Age	0.002	(0.009)	0.008	(0.005)
Unskilled Experience	0.001	(0.009)	0.056	(0.003)
Skilled Experience	0.081	(0.008)		
PT work	-0.072	(0.034)	-0.018	(0.011)
PT 2 year			-0.111	(0.021)
PT 4 year			-0.159	(0.018)
FT 2 year			-0.050	(0.018)
FT 4 year			-0.073	(0.019)
PT graduate school	0.001	(0.075)		
FT graduate school	-0.081	(0.046)		
1 year graduate school	0.112	(0.050)		
2 years graduate school	0.057	(0.050)		
3 years graduate school	0.100	(0.078)		
4+ years graduate school	0.090	(0.095)		
1 year college			0.055	(0.012)
2 years college			0.055	(0.015)
3 years college			0.107	(0.016)
4+ years college			0.150	(0.019)
Science major	0.139	(0.035)		
Unobserved type 1	0.166	(0.067)	-0.113	(0.022)
person-year obs.	1,700		12,372	

Notes: Bootstrap standard errors in parentheses. Full-time outside of school is the reference category for the work intensity dummies. Controls for calendar year dummies were also included.

Table 7: Correlation Matrix and Variances for Unobserved Abilities

	Skilled	Unskilled	Science	Non-Science	2-year
Skilled	1.000				
	(—)				
Unskilled	0.781	1.000			
	(0.053)	(—)			
4 year Science	0.057	0.212	1.000		
	(0.067)	(0.066)	(—)		
4 year Non-Science	0.053	0.163	0.902	1.000	
	(0.059)	(0.049)	(0.052)	(—)	
2 year	0.321	0.180	0.703	0.883	1.000
	(0.116)	(0.079)	(0.115)	(0.066)	(—)
Variances	0.127	0.074	0.235	0.165	0.085
	(0.008)	(0.003)	(0.045)	(0.020)	(0.013)

Note: Bootstrap standard errors in parentheses.

Table 8: Idiosyncratic Variances

Period	Skilled	Unskilled	Science	Non-Science	2-year
1	0.148	0.161	0.646	0.670	0.897
	(0.004)	(0.002)	(0.057)	(0.039)	(0.050)
2			0.160	0.212	0.352
			(0.026)	(0.017)	(0.029)
3			0.140	0.135	0.372
			(0.044)	(0.024)	(0.028)
4			0.120	0.111	
			(0.041)	(0.022)	
5+			0.321	0.199	
			(0.085)	(0.033)	

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

Table 9: Flow Utility Estimates

	2-year	4-year Sci	4-year Non-Sci	Work PT	Work FT	Grad Sch.
Constant	-3.594	-6.643	-5.193	-4.159	-3.318	-3.470
	(0.155)	(0.265)	(0.162)	(1.056)	(1.068)	(0.853)
Black	-0.024	0.108	0.165	-0.174	-0.186	0.134
	(0.123)	(0.133)	(0.065)	(0.079)	(0.053)	(0.391)
Hispanic	0.071	0.018	-0.049	-0.066	-0.037	0.020
	(0.099)	(0.085)	(0.074)	(0.099)	(0.047)	(0.480)
SAT math	0.066	0.456	0.181	-0.070	-0.006	-0.007
	(0.081)	(0.094)	(0.166)	(0.129)	(0.063)	(0.138)
SAT verbal	0.016	-0.005	0.209	0.106	-0.017	0.018
	(0.088)	(0.139)	(0.097)	(0.053)	(0.057)	(0.130)
HS grades	0.098	0.436	0.326	0.038	0.019	0.009
	(0.076)	(0.092)	(0.054)	(0.022)	(0.014)	(0.064)
Parent graduated college	0.215	0.554	0.614	0.102	-0.117	0.101
	(0.126)	(0.068)	(0.070)	(0.048)	(0.113)	(0.059)
Prior Academic Ability	0.115	0.835	0.948			0.063
	(0.172)	(0.467)	(0.256)			(0.211)
Expected Log Wage				0.404	0.404	
				(0.452)	(0.452)	
Previous HS	1.595	2.794	2.234	1.131	0.887	
	(0.073)	(0.142)	(0.096)	(0.069)	(0.059)	
Previous 2-year	2.932	1.770	1.487	0.320	0.305	
	(0.062)	(0.174)	(0.115)	(0.079)	(0.070)	
Previous 4-year Sci	1.314	5.808	3.185	0.796	0.580	0.049
	(0.218)	(0.156)	(0.140)	(0.105)	(0.107)	(0.352)
Previous 4-year Non-Sci	0.608	2.671	4.425	0.664	0.614	0.324
	(0.157)	(0.166)	(0.094)	(0.078)	(0.071)	(0.263)
Previous Work PT	0.083	0.225	0.206	1.869	1.443	0.116
	(0.087)	(0.136)	(0.088)	(0.057)	(0.058)	(0.269)
Previous Work FT	-0.214	-0.008	0.095	0.960	2.431	0.299
	(0.081)	(0.133)	(0.090)	(0.057)	(0.051)	(0.232)
Previous Grad School				0.259	0.438	3.437
				(0.329)	(0.230)	(0.257)
Graduated 4-year college				-0.320	0.567	
				(0.272)	(0.342)	
Work PT	0.464	-0.373	-0.160			-0.351
	(0.233)	(0.296)	(0.239)			(0.490)
Work FT	-0.925	-1.403	-1.395			-3.671
	(0.230)	(0.236)	(0.230)			(0.596)
Unobserved type 1	-0.121	0.368	0.472	0.835	0.220	-0.107
	(0.173)	(0.146)	(0.225)	(0.081)	(0.060)	(0.615)
Pr (Unobs. type = 1)	0.5033					
	(0.028)					
Pr (Unobs. major = science)	0.3139					
Pr (Unobs. GPA $\in [0.0, 2.5]$)	0.4660					
Pr (Unobs. GPA $\in (2.5, 3.0]$)	0.2211					
Pr (Unobs. GPA $\in (3.0, 3.6]$)	0.2223					
Pr (Unobs. GPA $\in [3.6, 4.0]$)	0.0905					
log likelihood	-25,836					
Person-year obs.	21,343					

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 log wage term.

Table 10: Model fit: College entry, attrition, re-entry, and graduation rates by period

	College Entry		College attrition		College re-entry		Ever graduated	
Period	Data	Model	Data	Model	Data	Model	Data	Model
1	43.27	46.14	0.00	0.00	0.00	0.00	0.00	0.00
2	8.51	4.50	8.71	11.93	0.00	0.00	0.00	0.00
3	4.03	3.63	8.37	8.65	1.34	1.04	0.00	0.00
4	3.04	3.15	7.17	6.90	2.33	1.59	0.19	2.86
5	1.80	2.79	6.26	5.72	2.57	2.00	7.39	7.41
6	1.48	2.46	6.02	4.85	2.14	2.27	12.65	12.82
7	1.93	2.26	2.93	4.19	2.63	2.34	16.50	17.27
8	2.23	2.02	3.29	3.81	1.92	2.44	18.55	20.15

Notes: Model frequencies are constructed using 100 simulations of the structural model for each individual included in the estimation.

Table 11: Model fit: Choice frequencies

Choice alternative	Data Frequency (%)	Model Frequency (%)
2-year & work FT	2.35	2.13
2-year & work PT	2.43	2.12
2-year only	2.57	2.22
4-year Science & work FT	0.68	0.58
4-year Science & work PT	1.39	1.06
4-year Science only	2.83	2.15
4-year Non-Science & work FT	1.56	1.41
4-year Non-Science & work PT	3.03	2.49
4-year Non-Science only	5.94	4.75
Work PT only	7.50	7.29
Work FT only	46.35	51.45
Home production	22.29	20.44
Grad school & work FT	0.42	0.67
Grad school & work PT	0.23	0.34
Grad school only	0.43	0.89

Note: Model frequencies are constructed using 100 simulations of the structural model for each individual included in the estimation.

Table 12: Average posterior abilities after last year of college for different choice paths

Choice Path	Skilled	Unskilled	Science	Non-Science	2-year	Share(%)
Continuous enrollment, graduate in science with x years of in-school work experience						
$x = 0$	0.009	0.045	0.252	0.233	0.185	1.3
$x \geq 1$	0.014	0.042	0.204	0.191	0.154	4.8
Continuous enrollment, graduate in non-science with x years of in-school work experience						
$x = 0$	-0.002	0.013	0.160	0.179	0.159	2.9
$x \geq 1$	0.007	0.021	0.149	0.167	0.150	11.4
Stop out (SO)						
SO, graduate in science	0.007	0.026	0.182	0.171	0.140	0.7
SO, graduate in non-science	-0.002	-0.001	0.134	0.158	0.149	2.3
SO then DO, start in 2yr	-0.015	-0.027	-0.079	-0.083	-0.072	3.8
SO then DO, start in science	-0.010	-0.044	-0.261	-0.257	-0.213	0.8
SO then DO, start in non-science	-0.010	-0.032	-0.187	-0.202	-0.177	2.2
Drop out (DO) after x years of school						
$x = 1$	-0.005	-0.007	-0.040	-0.044	-0.040	13.8
$x = 2$	-0.006	-0.014	-0.109	-0.117	-0.105	7.2
$x = 3$	0.005	-0.006	-0.138	-0.147	-0.128	4.2
$x = 4$	-0.005	-0.021	-0.181	-0.192	-0.168	2.5
$x \geq 5$	-0.006	-0.030	-0.213	-0.218	-0.185	2.4
Never attended college						
Never attend college	0.002	0.003	0.001	0.001	0.001	29.6

Notes: Abilities are reported in standard deviation units. This table is constructed using 100 simulations of the structural model for each individual included in the estimation. For those who never attended college, we use the posterior in the 10th period.

Table 13: Average posterior variance after last year of college for different choice paths

Choice Path	Skilled	Unskilled	Science	Non-Science	2-year	Share(%)
Continuous enrollment, graduate in science with x years of in-school work experience						
$x = 0$	0.124	0.070	0.042	0.050	0.047	1.3
$x \geq 1$	0.086	0.035	0.041	0.047	0.045	4.8
Continuous enrollment, graduate in non-science with x years of in-school work experience						
$x = 0$	0.124	0.070	0.079	0.038	0.034	2.9
$x \geq 1$	0.085	0.034	0.073	0.034	0.033	11.4
Stop out (SO)						
SO, graduate in science	0.082	0.032	0.050	0.044	0.040	0.7
SO, graduate in non-science	0.081	0.031	0.076	0.036	0.032	2.3
SO then DO, start in 2yr	0.078	0.030	0.159	0.098	0.049	3.8
SO then DO, start in science	0.080	0.031	0.096	0.069	0.045	0.8
SO then DO, start in non-science	0.080	0.031	0.126	0.072	0.043	2.2
Drop out (DO) after x years of school						
$x = 1$	0.101	0.050	0.211	0.146	0.075	13.8
$x = 2$	0.097	0.047	0.169	0.110	0.058	7.2
$x = 3$	0.094	0.043	0.136	0.083	0.047	4.2
$x = 4$	0.090	0.040	0.112	0.065	0.040	2.5
$x \geq 5$	0.083	0.034	0.099	0.056	0.035	2.4
Never attended college						
Never attend college	0.071	0.021	0.227	0.162	0.083	29.6
Time 0 population variance	0.127	0.074	0.235	0.165	0.085	

Notes: This table is constructed using 100 simulations of the structural model for each individual included in the estimation. For those who never attended college, we use the posterior in the 10th period.

Table 14: College completion status frequencies: baseline and counterfactual

	Basline (%)	Counterfactual (%)
Continuous completion (CC), Science	6.11	8.57
Continuous completion (CC), Non-Science	14.29	19.94
Stop out (SO) but graduated Science	0.74	0.91
Stop out (SO) but graduated Non-Science	2.29	2.90
Graduate from four-year college	23.42	32.31
Graduate from four-year college & ever attended two-year college	4.05	4.85
Stop out (SO) then drop out	6.80	5.03
Drop out (DO)	30.02	23.91
Never went to college	29.57	29.77

Notes: Figures constructed using 100 simulations of the structural model for each individual included in the estimation.

Table 15: Average abilities for different choice paths in full-information counterfactual scenario

Choice Path	Skilled	Unskilled	Science	Non-Science	2-year	Share(%)
Continuous enrollment, graduate in science with x years of in-school work experience						
$x = 0$	0.023	0.039	0.604	0.460	0.350	2.3
$x \geq 1$	0.200	0.223	0.470	0.333	0.267	6.3
Continuous enrollment, graduate in non-science with x years of in-school work experience						
$x = 0$	0.039	-0.028	0.426	0.529	0.543	4.5
$x \geq 1$	0.218	0.181	0.335	0.415	0.452	15.5
Stop out (SO)						
SO, graduate in science	0.109	0.136	0.451	0.355	0.285	0.9
SO, graduate in non-science	0.129	0.087	0.289	0.372	0.401	2.9
SO then DO, start in 2yr	-0.127	-0.112	-0.212	-0.211	-0.213	3.2
SO then DO, start in science	-0.195	-0.118	0.047	-0.013	-0.111	0.5
SO then DO, start in non-science	-0.194	-0.118	-0.004	0.036	-0.020	1.4
Drop out (DO) after x years of school						
$x = 1$	-0.092	-0.079	-0.208	-0.223	-0.225	12.2
$x = 2$	-0.118	-0.088	-0.160	-0.166	-0.182	5.7
$x = 3$	-0.141	-0.094	-0.074	-0.078	-0.116	3.1
$x = 4$	-0.168	-0.128	-0.079	-0.072	-0.102	1.6
$x \geq 5$	-0.136	-0.100	-0.146	-0.140	-0.159	1.3
Never attended college						
Never attend college	-0.069	-0.063	-0.259	-0.279	-0.268	29.8

Notes: Abilities are reported in standard deviation units. This table is constructed using 100 simulations of the structural model for each individual included in the estimation.

Table 16: Wage decomposition results

(a) College wage premium

	Baseline	Counterfactual	Counterfactual with baseline experience
Skilled ln(wages) for college grads ($X^g\gamma^g$)	2.770	2.937	2.896
Unskilled ln(wages) for HS grads ($X^n\gamma^n$)	2.523	2.363	2.452
Gap ($X^g\gamma^g - X^n\gamma^n$)	0.247	0.574	0.445

(b) Decomposition of characteristics

	Baseline	Counterfactual
Skilled ln(wages) for HS grads ($X^n\gamma^g$)	2.595	2.547
Difference due to characteristics ($X^g\gamma^g - X^n\gamma^g$)	0.176	0.390
Unskilled ln(wages) for college grads ($X^g\gamma^n$)	2.547	2.603
Difference due to characteristics ($X^g\gamma^n - X^n\gamma^n$)	0.024	0.240

Notes: In panel (a), "Counterfactual with baseline experience" refers to the counterfactual expected wage, but holding fixed the work experience levels from the baseline. The results in all panels of this table assume that all individuals are working full time. Additionally, when computing the counterfactual wages $X^n\gamma^g$, we condition the characteristics such that the HS graduates have the same grad school, choice of major, and work experience as those who went to college. Similarly, when computing the counterfactual wages $X^g\gamma^n$, we condition the characteristics such that the college graduates have the same college and work experience as those who did not graduate from college.

A Additional data details and estimation results

This appendix section details the criteria we use to define science majors, as well as how we select our estimation subsample. It also presents further estimation results that may be useful to the reader. Table A.1 lists the majors in each category. Table A.2 outlines each of the criteria used to construct our estimation subsample. Table A.3 shows cell sizes that our ability covariance matrix estimates are based on. Table A.4 reports estimates of the mapping between ASVAB subject tests and SAT scores. Table A.5 presents the wage market shock equation estimates, while Table A.6 shows estimates of the probability of graduation equation. Table A.7 lists the posterior variance of ability at the end of the panel, broken out by demographic characteristics.

Table A.1: 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 A.2: Sample Selection

Selection criterion	Resultant persons	Resultant person-years
Full NLSY97 sample	8,984	134,760
Drop females	4,599	68,985
Drop other race	4,559	68,385
Drop missing test scores	3,502	52,530
Drop missing HS grades or Parental education	3,327	49,905
Drop HS Dropouts (or those not receiving GED)	2,959	44,385
Drop observations before HS graduation	2,843	31,573
Drop right-censored missing interview spells	2,841	29,749
Drop any who attend college at a young age or graduate college in 2 or fewer years	2,841	29,097
Drop any who are not in HS at age 15 or under or have other outlying data	2,841	29,061
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[a]	2,713	21,343
Final estimation subsample	2,713	21,343

[a] Our structural estimation procedure incorporates integration of missing GPA and major observations, as discussed in Section 5.

Table A.3: Cell sizes of ability covariance matrix

	Skilled	Unskilled	Science	Non-Science	2-year
Skilled	352				
Unskilled	273	2,257			
4 year Science	233	555	796		
4 year Non-Science	300	731	660	990	
2 year	67	767	143	197	843

Note: Numbers reflect the number of individuals that ever participate in the given combinations of sectors.

Table A.4: Estimates of SAT scores using ASVAB subject tests as predictors

	SAT Math		SAT Verbal	
Subject test	Coeff.	Std. Error	Coeff.	Std. Error
Constant	-0.186	(0.054)	-0.300	(0.036)
Arithmetic Reasoning	0.404	(0.037)	0.135	(0.052)
Coding Speed	0.055	(0.063)	0.064	(0.036)
Mathematical Knowledge	0.187	(0.043)	-0.097	(0.061)
Numerical Operations	-0.018	(0.055)	-0.069	(0.042)
Paragraph Comprehension	-0.142	(0.057)	0.165	(0.054)
Word Knowledge	0.117	(0.037)	0.542	(0.056)
Observations	842		827	
R^2	0.2928		0.3544	

Note: Each of the test scores (both SAT sections and all ASVAB subject tests) is normalized to be mean zero, unit variance.

Table A.5: Wage market shock forecasting estimates

Parameter	Coeff.	Std. Error
Autocorrelation (non-graduates)	0.9597	(0.0849)
Autocorrelation (graduates)	0.8793	(0.2036)
SD of shock (non-graduates)	0.0360	(0.0062)
SD of shock (graduates)	0.0334	(0.0096)

Notes: Estimates come from separate AR(1) regressions for each sector. Bootstrap standard errors in parentheses.

Table A.6: Estimates of Probability of Graduation

	Coeff.	Std. Error
Constant	-3.882	(0.239)
Black	-0.673	(0.168)
Hispanic	-0.671	(0.205)
SAT math	0.087	(0.093)
SAT verbal	0.119	(0.104)
HS grades	0.192	(0.075)
Parent graduated college	0.045	(0.113)
Years in 2-year college	0.474	(0.058)
Years in 4-year college	0.859	(0.052)
Science major	-0.303	(0.114)
Prior ability science × Science major	1.249	(0.206)
Prior ability non-sci. × Non-Sci. major	1.004	(0.188)
Currently working part-time	0.146	(0.112)
Currently working full-time	-0.197	(0.141)
Unobserved type 1	0.180	(0.107)
Person-year obs.	13,390	

Notes: Parameter estimates from a logit predicting probability of graduating in the following period. Estimated only on students in their junior year and above. Bootstrap standard errors in parentheses.

Table A.7: Average posterior variances at end of panel by demographic characteristics

Characteristic	Skilled	Unskilled	Science	Non-Science	2-year
Race					
Black	0.070	0.025	0.174	0.118	0.063
Hispanic	0.065	0.021	0.174	0.118	0.063
White	0.058	0.022	0.145	0.097	0.054
Parental college status					
Parent did not graduate college	0.065	0.021	0.175	0.119	0.063
Parent graduated college	0.052	0.025	0.112	0.072	0.044
SAT math quartile					
1st quartile	0.068	0.022	0.187	0.127	0.066
2nd quartile	0.066	0.022	0.178	0.121	0.064
3rd quartile	0.060	0.022	0.147	0.098	0.055
4th quartile	0.051	0.024	0.112	0.074	0.046
SAT verbal quartile					
1st quartile	0.067	0.022	0.181	0.124	0.065
2nd quartile	0.067	0.022	0.178	0.121	0.064
3rd quartile	0.060	0.022	0.147	0.098	0.055
4th quartile	0.053	0.024	0.120	0.078	0.047
HS GPA quartile					
1st quartile	0.068	0.022	0.188	0.128	0.067
2nd quartile	0.061	0.022	0.148	0.098	0.055
3rd quartile	0.055	0.023	0.121	0.079	0.048
4th quartile	0.043	0.025	0.083	0.055	0.038
Time 0 population variance	0.127	0.074	0.235	0.165	0.085

Notes: This table is constructed using 100 simulations of the structural model for each individual included in the estimation.

B Integration of missing outcomes

This appendix section 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 Section 5.4.

The notation used throughout this section mirrors that which is used in Subsection 5.5. Additionally, we introduce two time-invariant indices: *(i)* $m \in \{\text{science}, \text{non-science}\}$, which indexes missing major; and *(ii)* $g \in \{1, \ldots 4\}$, which indexes missing GPA quartile.

B.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 integrating 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).

$$\Pr(r|i) = q_{ir} = \frac{\pi_r L^*_{idr} L_{ibr} L_{iyr}}{\sum_{r'=1}^{R} \pi_{r'} L^*_{idr'} L_{ibr'} L_{iyr'}} \tag{B.1}$$

$$\Pr(r,m|i) = q_{irm} = \frac{\pi_r L^*_{idr} L_{ibr} L_{iyr}\left(G_m, w_u, w_s\right)}{\sum_{r'=1}^{R} \pi_{r'} L^*_{idr'} L_{ibr'} L_{iyr'}\left(G_m, w_u, w_s\right)} \tag{B.2}$$

$$\Pr(r,g|i) = q_{irg} = \frac{\pi_r \int_{g'\in g} L^*_{idr} L_{ibr} L_{iyr}\left(G_g, w_u, w_s\right) \mathrm{d}g'}{\sum_{r'=1}^{R} \pi_{r'} \int_{g'\in g} L^*_{idr'} L_{ibr'} L_{iyr}\left(G_g, w_u, w_s\right) \mathrm{d}g'} \tag{B.3}$$

$$\Pr(r,m,g|i) = q_{irmg} = \frac{\pi_r \int_{g'\in g} L^*_{idr} L_{ibr} L_{iyr}\left(G_{mg}, w_u, w_s\right) \mathrm{d}g'}{\sum_{r'=1}^{R} \pi_{r'} \int_{g'\in g} L^*_{idr'} L_{ibr'} L_{iyr}\left(G_{mg}, w_u, w_s\right) \mathrm{d}g'} \tag{B.4}$$

$$\tag{B.5}$$

where (B.1) holds for those who have no missing data and is the same as (31), (B.2) holds for those who only have a missing major, (B.3) holds for those who only have a missing GPA, and (B.4) holds for those who have both outcomes missing. Note that, for (B.3) and (B.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.

We then obtain the population unobserved type probabilities π_r by marginalizing the above joint probabilities by a simple summation:

$$\pi_r = \frac{1}{N} \sum_i \sum_g \sum_m q_{irmg} \tag{B.6}$$

where $q_{irmg} = 0$ for the elements that correspond to non-missing outcomes.

B.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

B.2.1 Choice M-step

For the choice M-step, we construct 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.

B.2.2 Learning M-step

For the learning M-step, we perform weighted M-estimation to obtain estimates of the γ_{1j}'s and γ_{1l}'s, where again we use a different weight depending on the missingness of the key outcomes.

To the extent that missing major and grades affect our estimates of the population variance of ability (Δ) and idiosyncratic noise of the signals (σ^2), we also adjust the formulas accordingly by weighting with the appropriate weights as described above.

B.3 Second stage estimation

The estimation steps described above pertain only to the first stage of our model, in which we recover the parameters of the learning process while accounting for selection through the permanent unobserved heterogeneity types.

In the second stage of our model, we take the first stage parameters as given and estimate the CCPs and the structural flow utility parameters. In this second stage, we continue to use the weights described above for each individual, depending on his missing outcome status. Thus, the second stage estimation is closely similar to the choice M-step of the first-stage estimation in that we estimate a weighted multinomial logit to recover the CCPs and the structural flow utility parameters.

C 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. The steps to create each bootstrap replication are similar to the steps used to simulate the data when computing the fit of our model in Subsection 6.6. Namely:

1. Sample with replacement N individuals from the estimation subsample of the data, and collect the vector of personal background covariates associated with this sample of individuals.

2. For each individual, draw the ability vector from the estimated population distribution $\mathcal{N}\left(0, \hat{\Delta}\right)$. Similarly, draw an unobserved type $r \in \{1, 2\}$ from the estimated population distribution of types, which is a Bernoulli distribution with parameter $\hat{\pi}$ (i.e. probability of being unobserved type 1).

3. For each individual and each time period, generate the choice dummies using the predicted reduced-form choice probabilities described in Subsection 5.5.

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

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

6. Compute the implied posterior ability beliefs given the outcomes and choices generated previously as discussed in Subsection 3.4.1, and update the state space accordingly.

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

8. Repeat the procedure $B = 150$ 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:

$$\widehat{Var}\left(\widehat{\theta}\right) = \frac{1}{B-1} \sum_{b=1}^{B} \left(\widehat{\theta}_b - \overline{\widehat{\theta}}\right) \left(\widehat{\theta}_b - \overline{\widehat{\theta}}\right)' \tag{C.7}$$

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 detailed in the previous appendix section. 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.