The Specificity of General Human Capital: Evidence from College Major Choice

Josh Kinsler, University of Rochester
Ronni Pavan, Royal Holloway, University of London

May 12, 2014

Abstract

College graduates with a science or business related degree earn up to 25% higher wages than other college graduates. However, individuals do not always pursue careers related to their major, generating within-major gaps in wages that are similar in size to the across major gaps. As an example, science majors who work in jobs related to their field of study earn approximately 30% higher wages than those working in non-related jobs. In this paper, we aim to estimate the true returns to college major accounting for the specificity of skill. We develop a structural model of human capital that allows for both skill uncertainty and differential accumulation of human capital across major. Our findings indicate that the average returns to obtaining a business or science degree, although quite large, are smaller than the raw gaps would indicate. The average return to obtaining a science degree and working in a related job remains close to 30%. We also find that individuals are uncertain about their future productivity at the time of the college major decision. The combination of skill uncertainty and the specificity of the return appear to make majoring in a science related field less attractive.

1 Introduction

On average, college graduates majoring in science or business earn significantly higher wages than other college graduates, often on the order of 25% or more. While a portion of the wage gaps across major is likely the result of ability bias, i.e. ex-ante more productive workers cluster in business and science majors, part of the gap likely reflects heterogenous labor market returns to the varied skills students develop across majors. However, are the different skills that students accumulate across majors general, in the sense that they are equally valued regardless of career choice, or are they specific to particular types of jobs? This is an important question since if the human capital accumulated in college is specific, the returns to college major will depend critically on the type of job or career an individual pursues.

Understanding the specificity of human capital obtained in college is important not only for students making their major decision, but also for informing higher education policy. Recent programs in the U.S., such as the National Science and Mathematics Access to Retain Talent Grant (SMART) and the Science, Technology, Engineering, and Mathematics Expansion Program (STEP), indicate a strong desire among policy makers to increase the share of college students entering technical majors. There are at least two perceived benefits from such policies. First, policy makers believe that graduates who obtain a degree in a technical field generate a positive externality in the broader economy, primarily through innovation.[1] Second, because science and engineering majors tend to earn more on average, it is assumed that they will have an easier time paying off student debt, a burden many believe hampers economic growth.[2] However, if the skills obtained in college are specific and students induced to major in technical fields are less likely to work in a job related to their major, both of these potential benefits would be undercut.

In this paper, we first present detailed descriptive evidence indicating that the wage returns to major have a strong specific component to them. We use the Baccalaureate and Beyond Longitudinal Study (B&B), which consists of a representative sample of 11,192 graduating seniors drawn from the 1993 National Postsecondary Student Aid Study (NPSAS). This group of college graduates is subsequently surveyed in 1994, 1997, and 2003 about their schooling and labor market outcomes. Importantly, we observe a student's undergraduate major and self-reports about whether the current job is related to the major field of study. This direct

[1] Murphy *et al.* (1991) provides support for this belief. Note that the focus on math and science skills is not limited to the higher educations sector. The National Math and Science Initiative is another recently developed program that focuses strictly on primary and secondary school students.

[2] Avery & Turner (2012) discuss the interplay between major choice and student debt. In the March 2013 meeting of the Federal Open Market Committee, members of the panel mentioned the high level of student debt as a risk to aggregate household spending over the next three years. Meeting minutes can be found at http://www.federalreserve.gov/monetarypolicy/fomcminutes20130320.htm

measure of major applicability allows us to to identify the specificity of human capital within major by looking at wage variation across job types. Using a series of simple regressions we show that students who work in a job related to their major earn approximately 15% higher wages than those who do not, and that for science majors this gap can be as large as 30%.[3] These gaps are robust to standard observable ability measures, such as SAT scores and college grade point average.

Wage gaps across related and non-related jobs are consistent with a labor market where different types of jobs reward skills differently. The idea that a portion of a worker's accumulated human capital is not universally applicable across all jobs has been incorporated into career and occupational choice models for decades.[4] Yet, in most of these models the only channel through which workers accumulate specific human capital is through work experience and the amount of schooling.[5] Instead, what we have in mind is that individuals accumulate specific human capital through their choice of schooling type.

While the wage differences related to the applicability of one's major estimated by OLS are consistent with human capital specificity, there is a concern that the gaps could be driven by sorting on unobserved productivity. To explore this possibility further, we develop a model of major choice and labor market outcomes that relies on a human capital framework. The model is essentially a dynamic extension of the classic Roy (1951) model where workers first select into major and then select job type. Our framework and identification strategy is related to Carneiro *et al.* (2003) and Cunha *et al.* (2005), as it utilizes a similar factor-based structure. Individuals are endowed with two types of latent human capital, which we label as math and verbal human capital. An individual's math and verbal human capitals evolve as a result of schooling, where the nature of the accumulation varies across major. However, at the time of the major decision students only observe noisy measures of their underlying human capital.

[3]These results are consistent with recent findings from Silos & Smith (2012) who find that wage growth is positively related to how applicable the skills obtained in college are to the current job.

[4]Becker (1962) and Oi (1962) originally developed the notion of firm-specific human capital. Over time, the idea was expanded to include occupation, industry, and location specific human capital. Prominent examples include McCall (1990), Parent (2000), Neal (1995), Pavan (2011), and Kennan & Walker (2011). Recent papers, such as Poletaev & Robinson (2008), Gathmann & Schnberg (2010), and Yamaguchi (2012) have focused more on task specific human capital, stressing that what matters in not the job's label, but the actual tasks a job employs.

[5]Examples include Keane & Wolpin (1997) and Sullivan (2010).

Upon exiting college, individuals enter the labor market, learn their true human capitals, and endogenously work in a job that may or may not be related to their major. Wages are a function of an individual's level of human capital, major, and job type. While math and verbal human capital affect wages in all jobs, their impact can vary according to major and whether the job is related to the field of study.

The structural model of major choice yields a number of benefits relative to our initial OLS regressions. First, in our simple wage regressions we rely on SAT and major GPA to account for sorting into major and related jobs. Both measures are noisy, and in the case of GPA, endogenous to major choice. Thus, the returns to major and relatedness estimated using OLS could be attenuated or inflated. Using the factor structure of our model, however, we are able to increase precision by seamlessly integrating additional measures of ability, such as non-major GPAs, while also accommodating differential skill returns across major. Second, using the model we are able to calculate the returns to major and working in a related job for the average individual as well as for any selected sub-group. A third benefit of the model is that we are able to decompose the within-major variation in wages into general and specific components, illustrating precisely how much of the skills accumulated in each major are non-portable.

Finally, our model can also be used to investigate how human capital specificity in combination with skill uncertainty impacts college major choice. In our model, individuals face two types of uncertainty when choosing a major. First, conditional on their human capitals, individuals do not know whether they will work in a job related to their field of study. Second, because individuals are assumed to have imperfect information about their human capitals at the time of the major choice, the precise probability of finding a related job and the wages associated with each job once in the labor market are also unknown. While completely eliminating occupational or career uncertainty seems implausible, one could envision reducing labor market skill uncertainty, as it is essentially driven by a disconnect between being a good student and being a good worker. Many educational programs are already aimed at reducing this uncertainty, including internships, cooperative education, and student mentoring. However, these programs often don't start until after a student has already made their major decision. Instead, using our model we can easily examine how student major choices change if skill uncertainty is eliminated.

We estimate our model by maximum likelihood and show that we are able to replicate the key findings from our descriptive analysis using data simulated from the model. The parameter estimates indicate that at the time the college major decision is made students are quite uncertain about their underlying human capital. Students with high expected math human capital tend to sort into science related majors, while students with high expected verbal human capital sort away from business and science. The returns to math human capital are quite large for business majors, regardless of whether the individual is in a related job or not. For science majors, however, the return to math human capital is only large if the individual is working in a related job.

The notion that students do not have perfect information about their underlying skills when choosing a major is an assumption of the model. We test this informational assumption directly by examining the relative performance of an alternative model that allows the true human capitals to enter the major choice equation directly.[6] This is similar to an approach utilized in Cunha *et al.* (2005) for testing the agent's information set. We are able to reject the model of perfect certainty in favor of the model that allows for human capital uncertainty.

Using data simulated from our model with human capital uncertainty, we calculate the "average" returns to college major and working in a related job. We find that the returns to obtaining a business or science degree are 0.15 and 0.18 log points respectively, a sizable decrease relative to the returns estimated by OLS, 0.19 and 0.23 log points respectively. The true benefit of working in a related job for the average worker is quite similar across the OLS and model estimates. As an example, both indicate that wages are 0.27 log points higher in related jobs following a science degree. Workers who select into related jobs, though, tend to have much larger returns.

Finally, we examine how important the specificity of human capital is across majors through a simple wage decomposition exercise. The specific nature of human capital as it relates to job type explains at least 19% of the wage variability associated with human capital in the science field. For business and all other fields, specificity is responsible for only 4% and 9% of the wage variation related to human capital. These results, in conjunction with the

[6]Other papers have looked at different dimensions of imperfect information for major choice, focusing on its role for major switching, time to degree, and dropping out. See for example Altonji (1993), Arcidiacono (2004), Beffy *et al.* (2012), and Montmarquette *et al.* (2002).

evidence supporting human capital uncertainty at the time of the major choice, suggest that science related majors may be a riskier human capital investment for students relative to other fields of study. This risk could aid in explaining why students tend to shy away from math and science related majors. In a final exercise we show that indeed, students would choose to major in science more frequently if they knew their human capital with precision.

The remainder of the paper is as follows. Section 2 describes the B&B data in detail and provides descriptive evidence of the specificity of human capital through schooling. In section 3 we describe our model of major choice and discuss identification. Section 4 discusses estimation and presents the structural estimates along with some simple validation exercises. In section 5 we perform several counterfactual experiments to understand the relative importance of the various mechanisms driving major choices. Section 6 concludes.

2 Data and Descriptive Analysis

2.1 Data

We use data from the first cohort of the B&B to investigate the links between human capital, major choice, and wages. The initial B&B cohort consists of a representative sample of 11,192 graduating seniors drawn from the 1993 National Postsecondary Student Aid Study (NPSAS). This group of college graduates is subsequently surveyed in 1994, 1997, and 2003. While the sample is representative, the B&B provide sampling weights that we use in both the descriptive and structural analysis. Additional details regarding the sample are provided below.

2.1.1 Student Background and Schooling Data

The 1993 NPSAS and 1994 wave of the B&B collected detailed background data on each student. Using this data we are able to construct measures of respondent race, gender, and age. We limit the sample to males to avoid the complications fertility expectations have on female human capital accumulation and labor supply. This reduces the number of students to 4,834. In addition, there is a significant number of graduating seniors in 1993 who are older than 30. Older graduates tend to have lower (higher) math (verbal) SAT scores, choose business majors more often, and earn more than their fellow graduates who are significantly younger. In order to keep a relatively homogenous sample we drop anyone above the age of

30 in 1993, reducing the sample to 4,264 individuals. In addition, any individual with missing parental information or missing SAT scores is excluded from the sample, resulting in a final sample of 2,476 individuals.[7]

Detailed data on college major choice, as well as major specific grade point average (GPA) are readily available.[8] In order to keep the model tractable, we collapse major choice into three broad categories: business, science, and other. The aggregate business major includes fields such as economics, accounting, and general business. The science major includes engineering, the physical and natural sciences, and computer science. All remaining fields, such as social sciences, education, psychology, humanities, etc., fall into the other category. For the estimation of the structural model, we also construct variables that are assumed to be exogenous to the model. Using information from the IPEDS, we calculate the fraction of college students enrolled in each one of the three major categories for each state in 1993.[9] We then merge this information with the legal state of residence of the student's parents in 1993. This variable will act as an exogenous shifter in the choice of college major.[10]

In addition to major choice, information regarding student participation and performance across the various majors is also available. Total credits and GPA are separately available for business, social science, science and engineering, education, math, and foreign language courses, among others. We use these disaggregated variables to construct individual specific GPA measures for each of our broad major categories.[11] Note that we can only construct these

[7]Many individuals in our sample take the ACT rather than the SAT. We are unable to incorporate these individuals since only the composite ACT score is available and in our empirical analysis we treat separately the SAT math and verbal tests. We eliminate 1,552 individuals because of missing SAT scores. 1,020 of these individuals have a valid ACT score while 532 individuals have no valid SAT or ACT score.

[8]One disadvantage of the B&B is that we only observe each student's graduating major. As a result, we cannot incorporate major switching into our model.

[9]IPEDS is the Integrated Postsecondary Education Data System. Further details can be found at http://nces.ed.gov/ipeds/.

[10]There is quite a bit of variation in the fraction of college major across states, mostly driven by the small sample size. Science graduates range from 17% to 50%, business graduates range from 5% to 40%, and the residual category ranges between 20% and 74%.

[11]We trim the subject-specific GPAs according to the following procedure. We find the percentile x at which all individuals above this percentile receive a 4.0. We then find the GPA associated with the 1-x percentile, and replace all lower GPAs with this value. Thus, the top and bottom of the GPA distribution are trimmed in a similar fashion. The trimming does not affect the reduced form analysis and we do it to reduce the importance

measures if a student ever took a course in one of these subjects. GPA measures for business, science, and other majors are available for 45%, 95%, and 98% of our sample respectively.

Table 1 provides basic summary statistics describing schooling outcomes. Overall we see that science and business majors account for approximately 50% of the sample, with students split evenly across the two categories. A quarter of the sample eventually obtains a post-graduate degree.[12] When we examine the characteristics of the students across each major significant differences appear. Science majors have higher SAT math and verbal scores than either the business or residual group. In addition, their science GPA is also significantly higher. Note that both business and other majors take a significant amount of science credits, approximately 20 and 18 respectively. Across fields, GPA is always highest on average for those who chose to major in that field. This likely reflects both selection and the accumulation of additional human capital most relevant for the chosen field. Finally, across majors there also appear to be important differences in family background. Individuals who major in science are more likely to come from households where both parents are foreign born.

2.1.2 Labor Market Data

In survey years 1994 and 1997, respondents were asked about their primary employment during the month of April.[13] In 2003, respondents were asked about their current job. Across all surveys, individuals provided information regarding their hours, wages, and whether their job is related to their field of study.

Individuals can report wages either hourly, daily, weekly, monthly, or annually. We convert all wages to full-time yearly equivalents for those individuals who report working at least 30 hours per week. For anyone working fewer than 30 hours per week, we treat their annual salary as missing, since we do not model labor supply explicitly. All salaries are measured in

of the outliers which would be problematic once we estimate the structural model.

[12]In survey years 1994, 1997, and 2003, respondents are asked about post-BA degree receipt. For tractability, we treat all graduate degrees identically and do not allow individuals to switch their major at this point. This is largely consistent with the fact that 70% of the individuals who eventually obtain a graduate degree choose a graduate field of study that falls in the same broad major category as their undergraduate field of study. Note that if an individual reports obtaining a graduate degree by 1994, we utilize the graduate degree major rather than the undergraduate major. This occurs for 31 individuals.

[13]Note that in 1994, individuals were also asked about their primary job. When possible, missing information for the April job is replaced with information from the primary job.

2000$. The most important labor market variable for our purposes is whether an individual's job is related to their field of study. For the 1994 and 1997 surveys, a job is defined as being related to an individual's field of study if the respondent reported that the April job was either closely or somewhat related to their field of study. In 2003, the relationship question is altered slightly. A respondent is considered to be working in a job related to their field of study if their undergraduate education is very important in their current job or if their graduate education is very important in their current job and the respondent reports obtaining a graduate degree. Because of the change in the wording of the question, if an individual reports being in the same job as they were in 1997, we use the relationship variable from 1997. Approximately 13% of the valid relationship entries in 2003 are changed as a result. If information about whether the job is related to the field of study is missing, to the greatest extent possible we use information from the subsequent surveys to fill it in. For example, if the respondent reports starting their 2003 job prior to 1997, we replace the missing relationship variables in 1997 with their 2003 values.

Table 2 provides summary statistics for the labor market outcomes for the sample as a whole and by field of study. Overall, we see that approximately 70% of college graduates are working in jobs that are related to their college major. Not surprisingly, annual salaries are significantly different according to whether an individual is working in a job that is related to their field of study. This difference could reflect sorting, meaning that higher ability individuals are more likely to work in a related job, or it could reflect the idea that human capital is priced differently across different types of jobs.

Looking at labor market outcomes across fields of study illustrates important differences in outcomes by major. First, business and science majors earn significantly higher salaries relative to the residual group. This pattern is not unique to the B&B and can be found in the National Longitudinal Surveys of Youth (1979 and 1997) and more recently in the 2009 American Community Survey.[14] Second, and more important for our purposes, is the huge impact that working in a related job has on the salary of science majors. Science majors who work in a job related to their field of study earn close to a 30% premium relative to science majors who work in an unrelated job. The gaps for business and the residual major are only 3% and 11% respectively. These patterns are consistent with varying returns to skill, but

[14] Additional details available upon request.

could also reflect differing degrees of sorting into related jobs across majors or issues with aggregation.

2.2 Descriptive Evidence of the Specificity of Schooling Human Capital

Table 2 illustrates that not only do wages vary considerably across college major, but also within major according to whether an individual works in a related job. While these patterns are consistent with heterogenous returns to college through major choice and the existence of skill price differentials by job type, sorting or aggregation bias introduced by our course characterization of majors could also rationalize the data. In this section we run some simple regressions to shed light on the underlying mechanisms that drive the observed patterns and to help motivate our modeling choices regarding human capital and college major choice.

To study the wage differences across majors, we begin with some simple OLS regressions presented in Table 3. Column 1 indicates that conditional on year effects and graduate degree receipt, business and science majors earn 0.19 and 0.23 log point higher wages relative to the residual major category. Column 2 illustrates that the estimated returns to majoring in business or science decline by 5 and 10% respectively when we add controls for SAT, an observable measure of student ability.[15] In column 3, we include an additional observable measure of ability, major specific GPA, and find little change in the estimated return to majoring in business or science. We hesitate to interpret this as evidence that SAT scores are capable of fully accounting for ability sorting into major since GPA is endogenous to major choice through both the accumulation of major-specific skill and differential grading standards across fields of study.[16] The structural model presented in the next section provides

[15]An interesting result that emerges in Table 3 is that conditional on math SAT scores, verbal SAT scores negatively impact wages. This result is robust to controls for major, relatedness, and occupation. Further, the negative impact of verbal test scores on wages can be replicated using other data sources, such as the National Longitudinal Surveys of Youth. While we do not pursue in detail the root of this negative relationship, we choose a specification of our human capital model that is flexible enough to replicate this empirical regularity.

[16]For the set of students who take at least one course in each major category we can estimate a wage regression that includes separate GPA measures for each field of study. When we do this, the estimated returns to business and science decline by about a third. Again, we hesitate to interpret these as more precise measures of the returns to business and science since the GPAs measure individual skill post treatment. Science majors will have higher science GPAs since they build up more science skills and thus part of the return we seek to measure is likely captured by the separate GPA measures.

a framework for extracting skill signals from field specific GPAs while also dealing with the endogenous nature of these measures. Lastly, column 4 shows that the estimated major returns are not sensitive to whether we use hourly wages rather than our measure of annualized income.

The primary advantage of the B&B in examining the returns to college major is that it contains direct information on whether an individual's job is related to their field of study. In Table 4 we look at the relationship between this variable and the earnings of a worker. In the first column we note that even after controlling for observable measures of human capital using SAT and GPA, wages are significantly larger in jobs that are related to the field of study, and this relationship is significantly stronger for science majors.[17] In column 2 we show that this pattern is not generated by our aggregation of majors. Including dummy variables for each of the 28 majors observed in the data, we see that the impact of being in a related job for the residual major decreases, yet we still observe large returns to working in a related job for business and science majors.[18]

The third column of Table 4 examines how the inclusion of worker fixed effects alters the estimated returns to working in a related job. The importance of the job's relationship with the field of study decreases significantly for business and science majors, indicating that sorting across majors on unobserved dimensions can help explain the relationship between wages and labor market outcomes. However, even in this case science majors receive a much larger increase in their wage in related jobs when compared to workers with different majors. The inclusion of worker fixed effects ensures that the returns to working in a related job are identified by workers who switch job types. The fact that these workers switch, however, suggests that being in a related job is less salient for them as compared to workers who do not switch. As a result, the estimates in column 3 reflect the return to relatedness for a set

[17]The endogeneity of major specific GPA is less of a concern in this regression since identification of the returns to relatedness comes from within major variation in wages. This also suggests that we could incorporate additional GPA measures to help alleviate selection issues. If we include non-major GPA in addition to major specific GPA the estimated returns to relatedness are essentially unchanged.

[18]We also examined whether the large return to relatedness for science majors is primarily driven by one sub-major. We split the relatedness dummy for science into relatedness for engineers and relatedness for all other science graduates. The OLS return to relatedness for engineers is 0.20 log points while the OLS return to relatedness for all other science graduates is 0.28 points. Although the returns across the two groups of science majors are slightly different, they are both significantly larger than the return for the residual category.

of marginal workers who are close to being indifferent between related and non-related jobs. If the returns to relatedness are heterogeneous, then these estimates likely understate the average returns as the majority of workers work in a related job.

Column 4 of Table 4 shows that the returns to working in a related job decline considerably when we control for detailed occupation effects in the regression. With occupation effects, the impact of working in a related job is identified by differences in wages among workers with the same major and occupation but who report differently their relatedness status. The source of the differences in reported relatedness could reflect variability in training conditional on major or variability in job tasks conditional on occupation. Wage differences associated with either of these sources of variation is likely smaller than wage variation across occupation conditional on major, which is why the returns to relatedness decline. However, we don't necessarily want to discard all the across occupation variation in wages, since one could view occupation itself an indicator of relatedness. In fact, the decline in the relatedness coefficients after the inclusion of the occupational dummies indicates that related jobs tend to cluster in highly paid occupations.

To further investigate the relationship between occupation and relatedness we examine how workers are distributed across fourteen occupation categories conditional on major and relatedness. The results are displayed in Table 5 and indicate that indeed occupation and relatedness are tightly linked. For example, approximately 90% of business majors who work as a finance professional report working in a related job. Similarly, 95% of science majors working as engineers report being in a related job. Alternatively, across all majors 64% of mechanics/laborers report working in a non-related job. Thus, one could view relatedness as partly capturing the occupation-major match. The advantage of using the relatedness measure instead of occupation when examining the specificity of schooling human capital is its simplicity. We can avoid explicitly modeling occupation, which significantly reduces the computational burden.

The descriptive evidence suggests that the human capital developed through schooling is specific to particular types of jobs, however, no definitive conclusions can be drawn since our observable measures of ability are noisy, and in the case of GPA, endogenous to major choice. Thus, we cannot rule out the possibility that the important patterns we find in our descriptive analysis are either inflated or attenuated by the presence of sorting on unobservables and differential skill accumulation across major. For this reason, we present and then estimate a model of college major choice and relatedness that explicitly incorporates these features. As noted in the introduction, an additional benefit of the structural model is that it allows us to examine how skill specificity and uncertainty combine to influence major choice.

3 The Model

In this section we present a model of major choice and labor market outcomes based on an underlying human capital framework. The model is essentially a dynamic extension of the classic Roy (1951) model. Workers first select into major and then job type according to their unobserved skills. This self-selection implies that OLS will yield biased estimates of the average return to major and job type.

The remainder of this section presents the details of our sorting model. First, we characterize the factor structure of human capital since this binds the schooling and labor market outcomes together. We then discuss agent choices and outcomes in a recursive fashion, starting with outcomes in the labor market and moving backwards to the decision regarding college major. Finally, we end the section with a brief discussion of our measurement equations and identification. For simplicity, we exclude individual subscripts when presenting the model.

3.1 Human Capital and Scholastic Ability

Individuals enter college endowed with a vector of total ability A_0, which is composed of initial mathematical and verbal ability $(A_{m,0}, A_{v,0})$. We assume that individuals observe these random effects, but the econometrician does not.[19] Each of the endowment vector's elements can be further decomposed into an initial human capital $(H_{j,0})$ and an initial pure scholastic ability $(\nu_{j,0})$, where $A_{j,0} = H_{j,0} + \nu_{j,0}$ and $j = \{m, v\}$.[20] Human capital will be valued in the labor market while scholastic ability will not, allowing us to create a wedge between being a good worker and being a good student. In our baseline framework, we assume that agents observe total ability, $A_{j,0}$, but not its components. Therefore, agents are unsure about how much of their observed skill will translate to the labor market when making decisions on the

[19] In Section 3.4 we discuss how to incorporate noisy measures of student skill to identify the distribution of unobserved student abilities.

[20] We assume that all human capital elements are drawn from independent distributions.

type of degree to pursue. We later introduce a strategy to test whether this assumption is consistent with individual behavior.

While in college, each individual's math and verbal human capital evolve as a function of the chosen major according to

$$H_{j,1} = H_{j,0} + \mu^{H}_{j,f^*}\ , \ \text{ for } j = \{m, v\} \tag{1}$$

where f^* indicates the chosen field of study. Similarly, scholastic ability evolves with major choice according to

$$\nu_{j,1} = \nu_{j,0} + \mu^{\nu}_{j,f^*}, \ \text{ for } j = \{m, v\}\,. \tag{2}$$

The accumulation of human capital and scholastic ability through major choice is assumed to be independent of the initial levels of the math and verbal skill.[21]

3.2 Labor Market

Participation in the labor market is the final stage of an individual's life. Upon entering the labor market, individuals learn their human capitals, H_1, choose whether to work in a related job, and earn a wage. We first describe the wage equation and then consider the relatedness choice.

3.2.1 Wages

Each period individuals receive a wage that depends on their major field of study, their post-schooling human capitals, and whether the chosen job is directly related to the field of study of the worker:

$$\ln w_{r,f^*,t} = p_{r,f^*,m} H_{m,t} + p_{r,f^*,v} H_{v,t}. \tag{3}$$

The t subscript indicates calendar time, while the r subscript indicates whether the job is related to the studies of the individual. The coefficient on each type of human capital depends on both the individual's major and the relatedness status of the job. This reflects the idea

[21] While it is theoretically possible to identify heterogenous returns given the available data, in practice identification would be rather weak since we only observe major and noisy measures of post collegiate ability, namely GPA, at the end of a student's undergraduate career. Another implication of this data restriction is that we cannot investigate the causes and consequences of major switching or dropping out of college, since we do not observe these events in our data.

that workers from different schooling backgrounds sort into different occupations that possibly reward each type of human capital differently. Similarly, related jobs may reward skills differently than non-related jobs. The two types of human capital are "general" in the sense that workers bring these skills with them to all types of jobs, but at the same time they are "specific" since their value in the labor market depends on the particular job chosen.

Although it is reasonable to expect that human capital evolves once in the labor market as a result of accrued experience and/or post-secondary education investment, we assume that human capital remains constant since we have relatively few wage observations in the time dimension. Thus, $H_{j,t} = H_{j,1} \ \forall \ j$. However, in order to allow for the fact that wages increase both with time and the acquisition of a graduate degree, we also include in our wage specification a time dummy and a graduate degree dummy that depends on the major chosen.

3.2.2 Working in a Related Job

Prior to earning wages each period, workers must choose whether to work in a job related to their field of study.[22] We assume that the choice of working in a related job at time t depends on whether the latent variable R^* is positive, i.e. the worker works in a related job if $R^* > 0$, where R^* is defined as:

$$R^* \quad = \quad \Upsilon_t(f^*, H_{m,1}, H_{v,1}, Z_t) + u_t^{\Upsilon}. \tag{4}$$

The probability of working in a related job depends on the worker's chosen field of study (f^*), human capital ($H_{m,1}$ and $H_{v,1}$), and an exogenous shifter (Z_t). For the exogenous shifter, Z_t, we use state specific deviations from the national average of the fraction of workers in related jobs, controlling for average wages and demographic composition.[23] The random

[22] Although the choice of acquiring a graduate degree is clearly related to the human capital of an individual, we decided to treat graduate studies as an exogenous variable known to the agent since the beginning of life, and it is suppressed in the presentation of the model for expositional clarity. This choice simplifies greatly the computation of the likelihood function and reduces the number of the parameters to be estimated. We believe that the results are not strongly affected by this simplification given that in our sample workers with graduate degrees do not earn appreciably larger wages than workers without them (around 7%) and the sorting into those studies does not look very different across majors. Previous preliminary estimations of a version of this model with endogenous graduate studies support this hypothesis.

[23] The use of panel data can actually mitigate the need for an instrument in this context, as long as there is enough time variation in the selection rule (see for example theorem 11 of Heckman & Honore (1990)).

variable u_t^{Υ} is meant to capture in a reduced form way several features of the labor market potentially affecting relatedness outcomes. For example, workers might not receive an offer for a related job or there may be unobserved non-pecuniary benefits associated with related jobs that are orthogonal to a worker's skills. We believe that some of this randomness can actually be part of the information set of the agent at the time of the college major choice. As an example, some workers might be particularly skilled at finding a related job, independent of their human capital, altering the relative attractiveness of each major. To incorporate this type of heterogeneity into the model we split the "relatedness" shock into two components:

$$u_t^{\Upsilon} = \alpha_{f^*}^{\Upsilon}\theta + e_t^{\Upsilon}, \tag{5}$$

where the random component e_t^{Υ} is exogenous, independent over time, and unknown to the agent at the time of college decision. The random variable θ is instead assumed to be part of the information set of the agent at the time of the college decision and independent of other skills. If $\alpha_{f^*}^{\Upsilon}$ (or the parameters associated with θ in the college major choice decision) are not statistically different from zero, then we can conclude that agents do not have information about their future relatedness outcomes in addition to what is contained in their abilities.

As noted above, the reduced-form choice equation for working in a related job is meant to to capture two distinct ideas. First, workers may not always receive a job offer that utilizes the skills specific to their field of study. The probability of receiving such an offer likely depends on the skill level of the worker, captured by $H_{m,1}$ and $H_{v,1}$. It will also depend on demand factors which are field (f) and location specific (Z_t). Second, even workers who receive an offer to work in a related job may choose not to do so, and the non-monetary utility over job type would also depend on human capital, experience, and idiosyncratic components.

The actual realization of relatedness is a mix of labor demand and supply and without additional data would be difficult to disentangle. Therefore, we choose not to fully specify the model matching workers to job type. Of course this simplification comes at a cost since by taking this approach we cannot understand whether individuals work in jobs unrelated to their field of study because they are happy to do so or because they could not find more desirable jobs. As a result, any counterfactual analysis using our model will be focused on sorting since it is not possible to compare individual welfare across related and non-related jobs.

3.3 Major Choice

Prior to entering the labor market, individuals must decide which field of study to pursue. We assume that this choice is made under uncertainty, i.e., individuals do not observe their initial human capitals $(H_{m,0}, H_{v,0})$, but instead observe their initial total abilities $(A_{m,0}, A_{v,0})$. Because only human capital matters for labor market outcomes, individuals will infer from their total abilities what their future human capitals will be.

The value of choosing field of study f can be summarized by the following function:

$$V_f + \varepsilon_f = E_{H_0|A_0}\left[\Psi_f(H_0, \theta, gpa_{f,f}, Z_f)\right] + \varepsilon_f \quad (6)$$

This specification assumes that students' major choice depends upon the human capitals that they will carry to the labor market, skill in finding a related job, the potential GPA that they will have within the field if they chose that major, some observed preference shifters Z_f, and an idiosyncratic preference shifter ε_f which is unobserved to the econometrician. We include potential GPA in major f conditional on f being chosen, $gpa_{f,f}$, to capture non-monetary preferences over major that are driven by academic success in each field.[24] For the observed preference shifter, we utilize the fraction of students active in each major in the state of parental residence in 1993. Given that students do not directly observe human capital, they take expectations over the function Ψ_f using the information contained in their total abilities. This function can be seen as the value function associated with a particular major, where the agent has some flow non-monetary utility for choosing a particular major and discounts the future utility that he will receive from the labor market. Rather then deriving the actual value function implied by the labor market, we directly estimate V_f, reducing greatly the computational complexity of the estimation procedure. Individuals will choose major f if $\varepsilon_f - \varepsilon_{f'} \geq V_{f'} - V_f$ for all f'.

At the time of the major decision, individuals face two types of uncertainty. First, conditional on their human capitals, individuals do not know whether they will find a job related to their field of study. Second, because individuals do not know their labor market human capitals at the time of the major choice, the precise probability of finding a related job once

[24]Note that potential GPAs, $gpa_{f,f'}$, differ from the observed GPAs, GPA_{f,f^*}, since only three GPAs are realized while there exist 9 potential GPAs. The potential GPAs are inferred using our estimates of the GPA measurement equations discussed in the next section.

in the labor market is also unknown. By examining the importance of the random unobserved variable θ in the equation above and formally testing whether indeed only total ability A_0 enters the individuals' information set (and not its components H_0 and ν_0 separately), we can verify whether these types of uncertainty are present and analyze what their impact is for major choice.[25] For example, we can study whether individuals would change their major field of study if they knew their human capital, even if they still do not know precisely whether they will find a related job. This experiment can be achieved by simulating how choices would change if agents used $\Psi_f(\cdot)$ rather than $E_{H_0|A_0}[\Psi_f(\cdot)]$ in their decision process.

3.4 Measurement and Identification

In the model described above, individual schooling and labor outcomes are driven by human capitals and scholastic abilities which are unobserved to the econometrician. A crucial component of our empirical analysis and identification strategy is the availability of observable measures of individual ability, such as SAT scores and college GPAs. These measurements allow the econometrician to have a direct, although imperfect, look at an individual's human capital and scholastic ability. In the following sections we first present our measurement equations and then discuss how these measures help identify the key parameters of the model.

3.4.1 Measurements

Prior to college entry individuals complete the SAT entrance exam, which provides a measure of the initial total math and verbal abilities. We assume that the observed math and verbal SAT scores can be decomposed according to:

$$\begin{aligned} SAT_m &= \eta_{m,c} + \eta_{m,m}A_{m,0} + u_m \\ SAT_v &= \eta_{v,c} + \eta_{v,v}A_{v,0} + \eta_{v,m}A_{m,0} + u_v. \end{aligned} \tag{7}$$

[25] Unfortunately we cannot asses in our set up the full extent of the importance of uncertainty in relatedness choices as we do not solve for the exact value function associated with each major choice. In particular, we cannot asses the importance of the idiosyncratic relatedness shock, e_t^{Υ}, described in Equation 5. While this is a significant limitation, the computational gain to estimating Ψ_f directly is great because we avoid solving for the expected future labor market returns associated with each major, which would entail integrating over the human capitals.

The residual components, u_m and u_v, are measurement error and are independent from each other and the total abilities. The math SAT score is a dedicated measure of the student's total math ability, while the SAT verbal score is a function of both total verbal and total math ability. We include total math ability in the verbal score since SAT scores are highly correlated in the data.[26]

For post-college measures of total ability, we utilize field specific GPAs. Each individual can potentially report three GPAs, one for science related courses, one for the business related courses, and one for all other courses. The GPA in field f for an individual choosing major f^* is defined as:

$$GPA_{f,f^*} = \eta_{f,f^*} + \eta_{f,m}A_{m,1} + \eta_{f,v}A_{v,1} + u_f^{gpa}. \tag{8}$$

where $A_{j,1} = H_{j,1} + \nu_{j,1}$ and $j = \{m, v\}$. Note that we allow the constant of the GPA equation to depend on the chosen major. This feature can capture the fact that grade inflation is more severe in advanced courses within a major. Our assumption that human capital and scholastic ability evolve as a function of the major field of study may not be ideal in the presence of double majors. Fortunately in our data set only 5% of students have multiple majors. The GPA measurements for science and other majors are assumed to be missing at random, while we allow for selection into business courses on the basis of math and verbal total ability.[27] The random components u_f^{gpa} are assumed to be mean zero, independent from each other and the human capitals, but with potentially different distributions across fields.

While SAT scores and field specific GPAs yield information regarding total ability, wages provide a measure of human capital directly. However, we assume that the econometrician does not observe the true wage, but instead observes $\ln w_{r,f^*,t}^{obs} = \ln w_{r,f^*,t} + \epsilon_{r,f^*,t}$, where $\epsilon_{r,f^*,t}$ is independent of all other random variables but its distribution is allowed to differ across majors and relatedness.[28] As the next section discusses, the covariance between wages and

[26] The exclusion of verbal human capital from the SAT math measurement is without loss of generality. Its inclusion would simply require a relabeling of the "math" human capital. Note that the math and verbal labels are themselves arbitrary.

[27] Essentially all students take at least one course in science and the residual major category, allowing us to construct GPAs for these fields. However, there is a sizeable fraction of students who never take a business course, and these students are unlikely to be randomly selected.

[28] It is possible that a fraction of the random variable $\epsilon_{r,f^*,t}$ actually captures an idiosyncratic component of wages rather than strictly measurement error. This component would be an additional source of major-specific

our ability measures are critical for identifying the distribution of human capital.

3.4.2 Identification

In this section, we sketch our basic approach to identification. First, in order to simplify the discussion and eventually facilitate the estimation of the model, we substitute equations (1) and (2) when possible in order to formulate all equations in terms of H_0 and A_0. As a result some of the parameters are redefined. Equation (3) can be written as

$$\ln w_{r,f^*,t} = p_{r,f^*,c} + p_{r,f^*,m} H_{m,0} + p_{r,f^*,v} H_{v,0}, \tag{3'}$$

where we should note that the wage constant is now a function of both the chosen major and job type. Similarly, Equation (4) can be written as

$$R^* = \widetilde{\Upsilon}_t(f^*, H_{m,0}, H_{v,0}, Z_t) + u_t^{\Upsilon}. \tag{4'}$$

Finally, Equation (8) can be written as

$$GPA_{f,f^*} = \tilde{\eta}_{f,f^*} + \eta_{f,m} A_{m,0} + \eta_{f,v} A_{v,0} + u_f^{gpa}, \tag{8'}$$

where the constant in the above equation now includes the field specific shifters for human capital and scholastic ability illustrated in Equations (1) and (2).

We focus primarily on the identification of the wage parameters of Equation (3′), the measurement parameters of Equations (7) and (8′), and the standard deviations of the unobserved human capitals. Our strategy is based on an infinity argument as in Carneiro *et al.* (2003) and we refer to that paper for all the technical details. This assumption implies that we can move some element of the vector of exogenous variables Z in such a way that the resulting probability of selecting a certain combination of major choice and relatedness is equal to one. Clearly the infinity argument is a strong argument when compared to the patterns that we observe in the data. Our point of view is that this identification strategy indicates whether theoretically, the data can be rich enough to identify the model even without any restrictions implied by distributional assumptions. Of course, imposing a set of distributional assumptions will help in the actual implementation.

uncertainty that we do not consider in our model. For example, the empirical framework developed by Nielsen & Vissing-Jorgensen (2010) is aimed at capturing this component of uncertainty.

To show that the key parameters of the model are identified one simply needs to examine the correlations amongst our measures, conditional on being able to identify a set of individuals who always choose a particular major or job type. These correlations can be expressed as functions of the parameters of interest and are directly observed in the data. As an example, we use the covariances between SAT and GPA to identify the factor loadings in the SAT and GPA measurement equations and the variance of total ability. To disentangle human capital from scholastic ability we utilize the covariances between wages and SAT measures. A more formal presentation of identification is available in Appendix B.

Our approach yields identification of the "reduced form" equations in the sense that we have not identified the structural parameters that govern the evolution of the human capitals. As an example, we show identification of Equation ($3'$), where the constant in this equation incorporates the major specific returns to math and verbal human capital. For the purposes of this paper, the parameters we have identified are sufficient to describe the model.

4 Estimation and Model Fit

In this section, we first discuss our approach to model estimation. We then review some of the key parameter estimates from the model and provide some simulation results regarding model fit. Finally, we close this section by testing one of the key theoretical assumptions of the model, the premise that students are uncertain about their human capitals when making their major choice.

4.1 Estimation

We estimate our model by maximum likelihood. Let the data for an individual i be:

$$Y_i = \{f_i^*, SAT_i, GPA_i, G_{i,t}, r_{i,t}, w_{i,t}, Z_i\} \qquad \text{for } t \in \{1994, 1997, 2003\} \tag{9}$$

where $f_i^* = \{\text{Science(S), Business(B), Other(O)}\}$, $r_{it} = \{1 \text{ if Related, } 0 \text{ if Non-Related}\}$, SAT_i includes both math and verbal scores, and GPA_i includes grades across the three major fields. Z is the vector of exclusion restrictions while G is a dummy for graduate studies (which are assumed to be exogenous to the model). Ω is the vector of parameters that describe the model.

For expositional convenience, suppose for a moment that the econometrician is able to observe human capitals $H_i = \{H_{i,m,0}, H_{i,v,0}\}$, total abilities $A_i = \{A_{i,m,0}, A_{i,v,0}\}$, and skill in acquiring a related job, θ_i. Under this assumption, the individual contribution to the likelihood function $L(Y_i|Z_i, H_i, A_i, \theta_i; \Omega)$ can be written as follows:

$$\begin{aligned} L(Y|Z,H,A,\theta;\Omega) &= \Pr(f^*|Z,A,\theta,gpa_{f^*,f^*};\Omega) \\ &\times \prod_{j\in\{m,v\}} f_{u_j^{SAT}}(SAT_j|A;\Omega)) \times \prod_{f\in\{S,B,O\}} f_{u_j^{GPA}}(GPA_{f,f^*}|A,f^*;\Omega) \\ &\times \prod_{t\in\{1994,1997,2003\}} \Pr(r_t|Z,H,\theta,f^*,G_t;\Omega) \\ &\times \prod_{t\in\{1994,1997,2003\}} f_{u_t^{wage}}(w_t|Z,H,f^*,G_t,r_t;\Omega) \end{aligned} \tag{10}$$

where for ease of presentation we suppress the individual subscripts.[29] In order to calculate the terms in the above likelihood function we need to impose some parametric assumptions. As noted earlier, we assume that the utility for each major choice contains an idiosyncratic extreme-value shock, which yields a simple logit-type probability. Similarly, we assume that the probability of working in a related job takes the standard logit form. We also assume that the measurement errors in the GPAs, SATs, and wages are normally distributed. We let the dispersion of the idiosyncratic component of log wages vary by education and type of job.

Of course the econometrician cannot observe H_i, A_i, or θ_i, so we cannot evaluate the above likelihood function directly. The unobserved random variables must be integrated out of the likelihood function:

$$L(Y|Z;\Omega) = \int L(Y|Z,H,A,\theta;\Omega)dF_{H,A,\theta}(H,A,\theta;\Omega) \tag{11}$$

We assume that the unobserved human capitals, scholastic abilities, and skill at acquiring a related job are normally distributed, and take 10,000 draws from the vector of unobservables to evaluate the above integral.[30]

[29] The likelihood is actually slightly more complex since there is an additional selection equation for observing a business GPA. We model this probability as a simple probit where the only explanatory variables are total math and verbal ability.

[30] We investigate the sensitivity of our model to this assumption by allowing the human capitals and scholastic abilities to be drawn from a mixture of normals. While the estimates are not exactly the same, the qualitative predictions of the model are unchanged. In order to make a true comparison between the models we would need to increase the number of draws dramatically in the mixture model since there are now four additional

In order to calculate the probability of choosing a certain major, we need to calculate the expected human capitals conditional on the total abilities and the model parameters. Recall that when making the major choice, students only observe the sum of their scholastic ability and human capital. Under the assumption of normality, we can derive closed form solutions for these expectations. However, since there exists a one-to-one mapping between expected human capital and total ability, we just utilize total ability in our major choice equations. If the GPA is missing for a particular major, we integrate the probability over the distribution of its measurement error.

Parameter estimates and standard errors from the model are reported in Appendix Tables A.1, A.2, and A.3.[31] Note that when we estimate the model we restrict the loading on the total math ability in the verbal SAT equation ($\eta_{v,m}$) to zero and utilize a residualized version of the verbal SAT score as a dedicated measure of total verbal ability.[32] The residual verbal SAT score is obtained as the residual from a regression of the verbal SAT score on the math SAT score. We found that the model fit the data better utilizing this restriction, since it essentially allows the original measurement errors of the SAT equations to be correlated.[33] Finally, we restrict the coefficients on the math human capital in the wage equation to be positive and the coefficients on verbal human capital in the GPA equations to be positive.[34]

There are a few results worth pointing out directly. First, Table A.1 indicates that the

unobservables. However, this would greatly increase the computational burden and for practical reasons we do not pursue this further. As an additional check on the model, we randomly split our sample in half and use one half to estimate the model and the other half for out of sample prediction. The accuracy of the predicted outcomes for the hold-out sample continues to be quite good. Additional details available upon request.

[31] We calculate standard errors by inverting a numerical approximation of the Hessian.

[32] If we estimate the model using the raw SAT verbal score and treat $\eta_{v,m}$ as a free parameter, the basic conclusions of the model are unchanged. We discuss this issue further in the results section.

[33] The SAT math and verbal tests are taken at essentially the same time, so it is likely that the components of the test scores unrelated to ability are correlated. Indeed the SAT math and verbal scores are much more strongly correlated than the GPA measures. Utilizing the residualized version of the SAT verbal score as a dedicated measure of verbal ability is consistent with our original model specification if we assume that the projection coefficient of total verbal ability on total math ability is equal to the projection coefficient of the verbal SAT measurement error on the math SAT measurement error. Note that it is not possible to identify the correlation in the error term and therefore our assumption cannot be tested.

[34] When calculating standard errors, we assume that the parameters that hit the non-negativity constraints are equal to zero.

standard deviations of the math and verbal human capitals are small relative to the standard deviations of the math and verbal scholastic abilities. This means that individuals can only extract limited information about their underlying human capitals when they observe only total ability. More precisely, $\frac{\sigma_{E[H_{m,0}]}}{\sigma_{H_{m,0}}} = 0.27$ and $\frac{\sigma_{E[H_{v,0}]}}{\sigma_{H_{v,0}}} = 0.22$. In addition, individuals appear to have little knowledge about their ability to obtain a related job. The coefficients on θ are positive and statistically significant in the major choice equation, and positive in the relatedness equation for science and other majors. Decomposing the variance of the relatedness shock, e_t^{Υ}, indicates that agents are able to observe 11% and 1.5% of the variability in this shock for the other and science majors at the time of choosing a major.

In Table A.2, the major choice coefficients show that individuals with high math total ability will likely choose to major in science, while those with high verbal ability are likely to choose the other major. Our exogenous major preference shifters, the share of students in an individual's home state choosing each major and whether the parents were foreign born, have a positive and significant impact on major choice. Consistent with the data, the constants from the relatedness equations indicate that individuals who major in business or science are more likely to be in related jobs. In addition, individuals with higher levels of math and verbal human capital are more likely to work in related jobs conditional on major. Lastly, the exogenous relatedness preference shifter, the share of workers in related jobs in that state, is positive and significant.

Table A.3 lists the parameters from the wage and measurement equations. The returns to math human capital are quite large for business majors, regardless of whether the individual is in a related job or not. For science majors, however, the return to math human capital is only large if the individual is working in a related job. This pattern is consistent with the descriptive evidence presented earlier indicating larger returns to relatedness in the science field. The GPA measurement equations indicate that math ability is productive in all fields, but is most important for science majors. Verbal ability is only relevant for GPA in the other major.

For additional evidence regarding the fit of our estimated model, we perform a set of simple validation exercises. Using the estimated model parameters we simulate major choices and labor market outcomes for a large number individuals (10 times the size of the original sample). We then compare the simulated data to the actual data, the results of which are shown in Tables 6 and 7.[35] In general, the model does well at matching key patterns in the data. Science majors tend to be highly selected on SAT math scores and have wages that are approximately 0.23 log points higher in both the data and the model. The model is also able to capture the heterogeneity across major in the returns to working in a related job. Wages for science majors in a related job are 0.29 log points higher than science majors in non-related jobs, both in the data and the model. The gap for the other (business) major is also consistent between the data and the model, 0.06 (0.10) log points. Finally, the model is also able to match the GPA patterns across the matrix of majors in the data, and the result that a one-point increase in GPA increases wages by about 6%.

Yet, the model is not able to capture all of the variation in the data, particularly as it pertains to SAT verbal. In the data and in the model, business majors have the lowest SAT verbal scores, but the model is not able to replicate the level of sorting on SAT verbal in the other major.[36] Additionally, the negative impact of SAT verbal on wages is not as strong in the model as it is in the data. The model has difficulty replicating the wage impact of our three measures of ability, SAT math, verbal, and GPA, with only two types of human capital. A third type of human capital that affects GPA only would help fit the data better, however, this adds to the computational burden and affects the identification strategy. The resulting patterns of SAT verbal seem to pay the highest price for this simplicity, as verbal skills are high correlated with college GPA but negatively correlated with wages, creating a distortion in the correlation structure that a model with only two skills is not quite able to replicate.

4.2 Testing the Information Set

Our baseline model assumes that individuals are uncertain about their true underlying human capitals when they make their major choice. Specifically, one of the arguments of our approximated choice function $E(\Psi)$ is $E(H_0|A)$, which implies that students can only extract some

[35] We generate standard errors for the simulated moments using a bootstrap procedure which entails repeatedly drawing from our original sample, estimating the model, and simulating data. We repeat this process 10 times.

[36] If we estimate the model using the raw SAT verbal scores we are able to capture the sorting into major based on this measure. However, in this version SAT verbal projects positively on wages in our validation exercise. Importantly, the estimated returns to major are not sensitive to our choice of verbal SAT or residualized verbal SAT.

information about their human capital from their total ability. Alternatively, students may be able to observe H_0 directly and therefore base their major decision on this rather than on expectations of H_0. Similar to Carneiro *et al.* (2003), Cunha *et al.* (2005), Navarro (2011), or Guvenen & Smith (2010), testing the information set of the agent is essentially equivalent to testing whether the individuals choices are a function of the human capitals or their expected values.

Rather than assuming that individuals either know or do not know their human capitals, we specify a more general setting in which agents base their decision on the value of:

$$(1-\alpha)\,E[H_{j,0}|A;\Omega]+\alpha H_{j,0}\quad\text{for } j\in\{m,v\},$$

where $0\leq\alpha\leq 1$.[37] When $\alpha=0$ we obtain the main specification of our model and when $\alpha=1$ we obtain an alternative specification with perfect knowledge about human capital. When we estimate this model with α as a free parameter, we maximize the likelihood function at $\hat{\alpha}=0.012$. Using the results of Chant (1974) and Self & Liang (1987) for testing a parameter at its boundary, we test the null hypothesis that $\alpha=0$ versus the alternative that $\alpha>0$. Under the null, the $\mathrm{P}(\hat{\alpha}<0.019)$ is equal to 0.95, and thus we fail to reject the null hypothesis that $\alpha=0$.

While we are able to reject that individuals know their human capital with certainty when choosing a major, it is interesting to consider how well a model assuming perfect knowledge performs. Rather than present the parameter estimates and simulations from the perfect knowledge model, we simply compare the model fit for the perfect certainty model to our baseline case. If we implement a naive χ^2 test using the results from Table 6, where we erroneously assume that the covariance matrix is diagonal, we find that the joint test of equality between the data and the baseline model is 79.4. The 5% critical value for the χ^2_{17} distribution is 27.6, which means that we reject the baseline model. However, the analogous statistic for the model assuming perfect information is 144.4, indicating that it yields a worse fit of the data. If we compute a proper covariance matrix for the statistics in Table 6, the χ^2 test statistic for the baseline and perfect certainty models is 26.5 and 42.7 respectively. In this case our baseline model is not rejected while the perfect certainty model is rejected.[38] Finally, the Holm-Bonferroni test can also help us understand what the source of the model

[37] A similar expression is included for the quadratic terms related to human capital.

[38] In order to construct the full covariance matrix for the data moments we had to decrease the number of

failure is for the baseline and perfect certainty cases.[39] When comparing the baseline model to the data, two of the test statistics are rejected, the math and verbal SAT scores for the science graduates. For the model with perfect information, six statistics are rejected, again suggesting a decidedly worse fit.

We also compare the two models using the results from Table 7. Here we perform an F-test on the most complete wage specification and obtain for our baseline model a statistic of 1.9 with an implied a p-value of 0.03. The alternative model with perfect information delivers an F-statistic of 8.2 with an implied p-value of 0. Again we find that the null hypothesis is rejected for both models, but the model with imperfect information continues to outperform the perfect certainty model.

5 Counterfactuals

While the parameter estimates and model fit are informative, the primary advantage of the structural approach is that we can utilize our model to estimate the returns to college major and quantify the specificity of human capital when individuals differ in their unobserved abilities. In the following sections we estimate the returns to major and relatedness, quantify the role of varying skill prices, and explore how major choices would change if individuals had perfect information about their human capital at the time of the major decision. In all of the following regressions we use simulated data from the baseline model with human capital uncertainty as described in Section 4.

5.1 The Returns to Major Choice and Relatedness

Here we provide our estimates of the average return to each major if we forced students into each major (Average Treatment Effect - ATE), the average return to each major for those who selected that major (Treatment of the Treated - TT), and finally the average return for those that selected a different major (Treatment of the Untreated - TUT). In Table 8,

---

observations utilized since there are a number of missing business GPA observations. As a result, the estimated variances of the statistics increase, making it more difficult to reject the model.

[39]The Holm-Bonferroni test is a multiple test procedure wherein hypotheses are rejected one at a time until no further rejections can be done. The cutoff point is determined according to $P(k) > \frac{\alpha}{m+1-k}$, where $P(k)$ are the ordered p-values of the m test statistics of Table 6. For further detail see Holm (1979).

we report the results jointly with the OLS estimates from Table 7. Across the population, the average return to a business degree declines to 0.145 log points from an OLS estimate of 0.204. The estimated return to obtaining a science degree drops from 0.229 log points in the OLS specification to 0.184 in the structural model. In contrast, controlling for observable measures of ability using SAT and major specific GPA reduces the OLS returns by only one or two log points, highlighting the importance of accounting properly for the presence of measurement error in the available ability measures. Yet, despite the declines, the average return to obtaining a business or science degree continues to be quite large.

In the remaining columns of Table 8, we report the average returns for those who selected a business, science, or residual major, respectively. Regardless of the chosen major, the average return to a science degree is always large relative to the other major and does not vary much across the population. The average return to a business major is actually smallest for those who choose business. The main conclusion to draw from Table 8 is that there are real monetary returns to majoring in business or science. However, individuals continue to major in the residual category, reflecting the fact that individuals do not choose majors based only on expected earnings.

The descriptive evidence presented in Section 2 indicates that not only do individuals majoring in science related fields earn a significant wage premium, but also that this premium varies greatly according to the type of job the individual ultimately chooses. The evidence from Table 8 indicates that the average return to science is smaller than the reduced form evidence indicates. Table 9 examines whether the wage returns associated with working in a related job are robust to sorting based on unobserved skill.

The first column of Table 9 presents the OLS estimates of the returns to working in a related job across all fields estimated on the simulated data.[40] Note that the numbers in Table 9 are not the coefficient estimates themselves, but rather the total returns which require adding the business and science coefficients to the coefficient for the residual major group. The results are quite similar to those from the first column of Table 4, however, the OLS estimates are likely biased by both sorting into major and relatedness. One method for minimizing this bias is to incorporate individual fixed effects, a strategy we pursued in the descriptive analysis of Section 2. The second column of Table 9 lists the relatedness returns estimated on the

[40] In each regression we also control for major, year, and graduate degree effects.

simulated data when individual fixed effects are incorporated.[41] The results for business and science are quite similar to the fixed-effect estimates presented in Table 4, an indication that our model is able to replicate the returns to working in a related job for those workers who are on the margin between working in a related and non-related job.

As noted in Section 2, the problem with the fixed effect approach is that the estimated returns for the marginal worker likely understate the returns for the average worker. This is largely borne out by the results in column 3 of Table 9, which presents the average return to working in a related job when there is no selection into major. In other words, these numbers reflect the population returns to working in a related job for each major. For science and other majors, the average return to working in a related job is higher than the corresponding fixed-effect estimate. Moreover, the returns to relatedness for those choosing to work in a related job is much higher then the corresponding fixed effect estimate. In particular, the return to a related job for science majors choosing to work in related job is 0.341 as compared to the fixed effect estimate of 0.159. In contrast to science and other majors, the returns to working in a related job for business majors is relatively flat across marginal, average, and treated workers.

The final three columns of Table 9 present the average returns to relatedness conditional on major choice. Overall, the patterns in the returns to relatedness are quite similar to those presented in columns 3 through 5. However, the relatedness return for science majors do increase, a result of the fact that high math human capital students tend to sort into the science field and high math human capital individuals benefit the most from working in a related job. The sizeable returns to working in a related job indicates that the specificity of human capital within major is a distinct feature of the data. In the next section we investigate further the specificity of schooling human capital.

[41] In our simulations we estimate the fixed effect model using first differences as it was simpler to incorporate into our bootstrap procedure. The two estimators are asymptotically equivalent under our assumptions, but yield marginally different estimates.

5.2 The Importance of the Specificity of Human Capital

In this exercise, we decompose the *within-major* variation in wages related to human capital into general and specific components.[42] Recall that the wage equation is given by

$$\ln w_{r,f^*,t} = p_{r,f^*,c} + p_{r,f^*,m} H_{m,0} + p_{r,f^*,v} H_{v,0}.$$

We view the wage variation that arises strictly through $H_{m,0}$ and $H_{v,0}$ as variation in the general skills that individuals bring to the labor market. In contrast, the variation in wages due to p is related to how math and verbal general skills apply to a specific type of job.[43]

The variation in wages related to general human capital can be pinned down by fixing the p's for each major using a weighted average of the related and non-related coefficients. We use the predicted proportion of individuals within each major working in related jobs as our weight. Combining these coefficients with the within-major variation in human capital we can easily generate the overall variation in wages related to general human capital. Table 10 provides the decomposition of wages into initial human capital and other components, followed by a decomposition of the impact of initial human capital into its general and specific components. We find that initial human capital plays the largest role in the science field, explaining 32% of the overall variation in wages. Moreover, we find that for science majors variation in math and verbal human capital alone can explain only 81% of the wage variation related to human capital, meaning that 19% of the variation is driven by skill prices. In contrast, math and verbal human capital are responsible for 96% and 91% of the overall variation in wages related to human capital for business and other majors. Thus, not only does human capital

[42] The initial human capitals and their coefficients are responsible for approximately a third of the overall wage variation. The remaining variation arises through the graduate degree and year dummies, as well as the measurement error. Obtaining a graduate degree and/or accumulating additional work experience certainly boosts human capital. Thus, the wage variation arising from human capital is most certainly larger than what is presented here.

[43] Notice that we only explore the within major variation in wages when decomposing wage variability into its general and specific components. However, the between major variation could potentially be decomposed into general and specific components given that the price of human capital differs by major. We are unable to do this since we cannot decompose the cross-major variation in the wage constants into specific and general components. This is a consequence of estimating the reduced-form parameters of the model rather than the structural parameters governing the accumulation of human capital. Therefore, the results of this section can be seen as a lower bound of the importance of specific human capital.

play the largest role in the science field, but the specificity of its application is also large relative to business or other majors.[44] The results in Table 10 also complement the evidence on the returns to relatedness presented in Table 9. Science majors have the highest returns to working in a related job and also the greatest dispersion in wages induced by the presence of specificity. In contrast, the specificity of human capital plays the smallest role in both returns and wage dispersion for business majors.

The specificity of human capital is an additional source of risk in the sense that even within major the returns to human capital can vary considerably. In this exercise, we find evidence that in addition to taste, there is room for risk to help explain why so few students choose to major in science related fields. The fact that relatively few students also tend to choose a business degree is unlikely to be explained by risk, since as noted above the specificity of human capital plays the smallest role in the wage dispersion for business majors.

Finally, we take our model quite literally and simulate the counterfactual major field of study that students would have selected had they known their human capitals precisely, rather than just observing their total abilities. Although we estimated the major choice equation as a function of total ability, A, we can transform the estimated coefficients as if the equation were directly a function of $E(H|A)$. The counterfactual then requires replacing the expectation with the true H. This is a courageous exercise since our model assumes that total ability affects the choice of major because it affects students' expectations over their human capital. We cannot allow math and verbal total ability to enter the utility for major directly because we would be unable to distinguish this channel from the human capital impact. However, our model does allow preferences for major to be linked to academic performance through the potential GPA, which is itself a function of an individual's total abilities. Papers that have tried to understand how important are labor market expectations for major choice have usually found that, although they are significant and economically important, they have a smaller impact than preferences (for example, Arcidiacono (2004), Beffy *et al.* (2012), and

[44] It should be noted that the residual major group has the greatest overall variability in wages. Therefore, there is still much to learn about what generates such variability. The residual group is certainly the most heterogenous in terms of the underlying group of majors and we cannot rule out that a large fraction of what we impute as the effect of measurement error is in reality a dimension of heterogeneity that we do not account for.

Wiswall & Zafar (2012)). For this reason we think that the counterfactual exercise discussed below should be seen as a upper bound on the actual impact that a richer information set would have on student choices and as a qualitative indication on the direction of the effect.

In the counterfactual world where individuals are assumed to know their human capitals, we observe that 30.0% of students would choose a science major, 43.8% would choose a business major and only 25.8% would choose the residual major. Although we observe that students in all majors tend to change their choice given the new information, the vast majority of students who change major end up in either a science or business major. The increase in the share of science majors is consistent our earlier hypothesis. Interestingly we also observe that the fraction of business majors increases. It should be noted that the estimation of the parameters of the major choice equation is not directly linked to labor market outcomes, as it would be in a complete dynamic programming model. Therefore, it is difficult to speculate on which are the driving forces for this result.

6 Conclusion

It is well documented that the returns to obtaining a college degree vary significantly across fields of study, with business and science majors earning a significant wage premium relative to all other fields. In this paper we illustrate that there also exists significant variation in wages within major according to the type of career an individual pursues. The question is whether the observed differences in wages are driven by selection, or if there truly are differential returns both within and across field.

We estimate a structural human capital model that allows for sorting into major and job type based on observed and unobserved characteristics to determine whether the observed wage gaps are driven primarily by selection. Our findings indicate that selection plays a role in generating the observed wage gaps across major, particularly for business. However, the large wage gaps across job type are instead explained by true differences in returns. Even when we account for selection into major, the returns to business and science majors are still economically significant. Thus we are still left with the question of why do individuals pursue less remunerative majors?

One reason individuals may not pursue a science degree is a lack of knowledge about the true returns. Future success in the labor market depends on the skills a worker accumulates and the type of job pursued. If individuals do not know precisely their skills when making their major choice, then they face two sources of uncertainty. First, there is a risk that the human capital students accumulate will be de-valued if they do not obtain a job related to their field of study. Second, upon obtaining a related job, individuals face an uncertain wage since they do not know their skill level exactly. This story is consistent with our finding that both dimensions of uncertainty may make science and math related majors unattractive relative to business or other majors. Human capital plays the largest role in wages for science majors, and the specificity of human capital is also most severe for graduates with a science degree. The impact of skill uncertainty on major choice has thus far not been considered when designing policies to increase the number of students majoring in science and technology fields. However, reducing uncertainty by incrementing information might be more valuable than utilizing monetary incentives, such as those provided by the SMART grant.

In our counterfactual exercises we indeed find that, were students aware of their human capitals they would choose a science major more frequently. These results should just be seen as qualitative evidence since in our estimation framework we do not separately estimate monetary and non-monetary preferences over major and working in a related job. This makes it difficult to convincingly estimate the curvature of the utility function, which is crucial for understanding the importance of risk, and most likely tends to overstate the impact of a change in the information set. Future research should seek to estimate the effect of risk on major choice more directly, while also incorporating additional sources of risk not considered here, such as risk related to college dropout, the business cycle, and wage variability over the life cycle.

References

Altonji, Joseph G. 1993. The Demand for and Return to Education When Education Outcomes are Uncertain. *Journal of Labor Economics*, **11**(1), pp. 48–83.

Arcidiacono, Peter. 2004. Ability sorting and the returns to college major. *Journal of Econometrics*, **121**(1-2), 343–375.

Avery, Christopher, & Turner, Sarah. 2012. Student Loans: Do College Students Borrow too Much - or Not Enough? *Journal of Economic Perspectives*, **26**(1), 165–192.

Becker, Gary S. 1962. Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy*, **70**(5), pp. 9–49.

Beffy, Magali, Fougre, Denis, & Maurel, Arnaud. 2012. Choosing the Field of Study in Post-Secondary Education: Do Expected Earnings Matter? *Review of Economics and Statistics*, **94**(1).

Carneiro, Pedro, Hansen, Karsten T., & Heckman, James J. 2003. 2001 Lawrence R. Klein Lecture Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice. *International Economic Review*, **44**(2), 361–422.

Chant, D. 1974. On Asymptotic Tests of Composite Hypotheses in Nonstandard Conditions. *Biometrika*, **61**(2), 291–298.

Cunha, Flavio, Heckman, James J., & Navarro, Salvador. 2005. Separating Heterogeneity from Uncertainty in Life Cycle Earnings. *Oxford Economic Papers*, **57**(2), 191–261.

Gathmann, Christina, & Schnberg, Uta. 2010. How General Is Human Capital? A Task-Based Approach. *Journal of Labor Economics*, **28**(1), 1–49.

Guvenen, Fatih, & Smith, Tony. 2010. *Inferring Labor Income Risk from Economic Choices: An Indirect Inference Approach.* Working Paper.

Heckman, James J., & Honore, Bo E. 1990. The Empirical Content of the Roy Model. *Econometrica*, **58**(5), 1121–1149.

Holm, Sture. 1979. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, **6**(2), 65–70.

Keane, Michael P, & Wolpin, Kenneth I. 1997. The Career Decisions of Young Men. *Journal of Political Economy*, **105**(3), 473–522.

Kennan, John, & Walker, James R. 2011. The Effect of Expected Income on Individual Migration Decisions. *Econometrica*, **79**(1), 211–251.

McCall, Brian P. 1990. Occupational Matching: A Test of Sorts. *Journal of Political Economy*, **98**(1), pp. 45–69.

Montmarquette, Claude, Cannings, Kathy, & Mahseredjian, Sophie. 2002. How do young people choose college majors? *Economics of Education Review*, **21**(6), 543–556.

Murphy, Kevin M., Shleifer, Andrei, & Vishny, Robert W. 1991. The Allocation of Talent: Implications for Growth. *The Quarterly Journal of Economics*, **106**(2), pp. 503–530.

Navarro, Salvador. 2011. *Using Observed Choices to Infer Agent's Information: Reconsidering the Importance of Borrowing Constraints, Uncertainty and Preferences in College Attendance.* Working Paper 2011-8 CIBC Working Paper Series.

Neal, Derek. 1995. Industry-Specific Human Capital: Evidence from Displaced Workers. *Journal of Labor Economics*, **13**(4), pp. 653–677.

Nielsen, Helena Skyt, & Vissing-Jorgensen, Annette. 2010. *The Impact of Labor Income Risk on Educational Choices: Estimates and Implied Risk Aversion.* Working Paper.

Oi, Walter Y. 1962. Labor as a Quasi-Fixed Factor. *Journal of Political Economy*, **70**(6), pp. 538–555.

Parent, Daniel. 2000. Industry Specific Capital and the Wage Profile: Evidence from the National Longitudinal Survey of Youth and the Panel Study of Income Dynamics. *Journal of Labor Economics*, **18**(2), pp. 306–323.

Pavan, Ronni. 2011. Career Choice and Wage Growth. *Journal of Labor Economics*, **29**(3), pp. 549–587.

Poletaev, Maxim, & Robinson, Chris. 2008. Human Capital Specificity: Evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys, 1984-2000. *Journal of Labor Economics*, **26**(3), 387–420.

Roy, A.D. 1951. Some Thoughts on the Distribution of Earnings. *Oxford Economics Papers*, **3**(2), 135–146.

Self, Steven G., & Liang, Kung-Yee. 1987. Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. *Journal of the American Statistical Association*, **82**(398), 605–610.

Silos, Pedro, & Smith, Eric. 2012. *Human Capital Portfolios*. Working Paper.

Sullivan, Paul. 2010. A Dynamic Analysis Of Educational Attainment, Occupational Choices, And Job Search. *International Economic Review*, **51**(1), 289–317.

Wiswall, Matthew, & Zafar, Basit. 2012. *Determinants of College Major Choice: Identification using an Information Experiment*. Working Paper.

Yamaguchi, Shintaro. 2012. Tasks and Heterogeneous Human Capital. *Journal of Labor Economics*, **30**(1), 1–53.

Table 1: Schooling Statistics

Overall		**by Major**			
			Business	Science	Other
Age in 1993	23.31	SAT Math	525	596	525
	(1.48)		(113)	(106)	(111)
SAT Math	542	SAT Verbal	450	496	474
	(116)		(98)	(105)	(102)
SAT Verbal	473	Business GPA	3.14	3.07	2.71
	(103)		(0.44)	(0.75)	(0.78)
		Business Credits	37.04	2.91	5.70
% Business Major	26.12		(18.98)	(7.55)	(12.16)
% Math/Science Major	24.66	Science GPA	2.75	3.23	2.62
			(0.65)	(0.47)	(0.7)
% Other Major	49.22	Science Credits	20.01	78.43	17.93
			(12.61)	(29.34)	(17.24)
% Graduate Degree	27.34				
		Other GPA	2.89	3.14	3.23
Both Parents Foreign Born	0.08		(0.51)	(0.53)	(0.45)
		Other Credits	43.00	29.34	62.69
			(25.51)	(19.88)	(33.47)
		Both Parents Foreign Born	0.08	0.13	0.04

Statistics are based on data from the Baccalaureate and Beyond Longitudinal Study. The sample includes males who are below 30 at the time of graduation who have valid SAT scores and parental information. There are total 2,476 individuals in the sample. The Business major includes fields such as economics, accounting, and general business. The Science major includes engineering, the physical and natural sciences, and computer science. All remaining fields, such as social sciences, education, psychology, humanities, etc., fall into the Other major category. Further sample details can be found in Section 2.1.

Table 2: Labor Market Statistics

	Overall	by Major		
		Business	Science	Other
Job is Related to Major Field of Study	70.46	78.35	79.81	61.01
Log(Annualized Salary)	10.52	10.59	10.65	10.41
	(0.62)	(0.60)	(0.57)	(0.65)
Log(Annualized Salary) if Related	10.57	10.59	10.70	10.45
	(0.59)	(0.58)	(0.52)	(0.62)
Log(Annualized Salary) if NOT Related	10.40	10.56	10.42	10.34
	(0.69)	(0.66)	(0.69)	(0.69)

Statistics are based on data from the Baccalaureate and Beyond Longitudinal Study. Sample selection and major definitions are provided in the notes to Table 1. Sampled individuals are observed working in 1994, 1997, and 2003. For the 1994 and 1997 surveys, a job is defined as being related to an individual's field of study if the respondent reported that their job was either closely or somewhat related to their field of study. In 2003 a respondent is considered to be working in a job related to their field of study if their undergraduate education is very important in their current job or if their graduate education is very important in their current job and the respondent reports obtaining a graduate degree. Further details can be found in Section 2.2.

Table 3: Returns to College Major

Dep. Var.	Log(Inc)	Log(Inc)	Log(Inc)	Log(Wage)
Business	0.191*	0.181*	0.185*	0.162*
	(0.024)	(0.024)	(0.023)	(0.021)
Science	0.234*	0.210*	0.215*	0.198*
	(0.021)	(0.022)	(0.022)	(0.021)
SAT Math/100		0.045*	0.042*	0.044*
		(0.011)	(0.011)	(0.010)
SAT Verbal/100		-0.040*	-0.045*	-0.039*
		(0.013)	(0.014)	(0.013)
Major Specific GPA			0.070*	0.065*
			(0.020)	(0.019)
Year/Graduate Degree Effects	Y	Y	Y	Y
N	4,927	4,927	4,927	4,927

*,** Indicate a coefficient is statistically significant at a 5 and 10% significance level. Heteroskedastic-robust standard errors reported in parentheses. Estimates are based on data from the Baccalaureate and Beyond Longitudinal Study. Sample selection and major definitions are provided in the notes to Table 1. Unit of observation is an individual-year combination.

Table 4: Returns to College Major Accounting for Utilization

Dep. Var.	Log(Inc)	Log(Inc)	Log(Inc)	Log(Inc)
Business	0.143*			0.116*
	(0.051)			(0.047)
Science	0.019			0.001
	(0.047)			(0.045)
Job is Related to Studies	0.063**	0.019	0.048	0.080*
	(0.035)	(0.030)	(0.030)	(0.037)
Business*Job is Related to Studies	0.039	0.096**	0.003	-0.057
	(0.058)	(0.055)	(0.048)	(0.058)
Science*Job is Related to Studies	0.229*	0.227*	0.118*	0.133*
	(0.051)	(0.049)	(0.054)	(0.053)
SAT Math/100	0.041*	0.031*		0.021*
	(0.011)	(0.011)		(0.010)
SAT Verbal/100	-0.042*	-0.033*		-0.038*
	(0.013)	(0.012)		(0.013)
Major Specific GPA	0.064*	0.085*		0.076*
	(0.020)	(0.019)		(0.019)
Year/Graduate Degree Effects	Y	Y	Y	Y
Worker Fixed Effects	N	N	Y	N
Detailed Major Effects	N	Y	N	N
Detailed Occupation Effects	N	N	N	Y
N	4,927	4,927	4,927	4,927

*,** Indicate a coefficient is statistically significant at a 5 and 10% significance level. Heteroskedastic-robust standard errors reported in parentheses. Estimates are based on data from the Baccalaureate and Beyond Longitudinal Study. Sample selection and major definitions are provided in the notes to Table 1. The definition of a related job is provided in the notes to Table 2. Unit of observation is an individual-year combination.

Table 5: Occupation Distribution by Major and Relatedness

	Business Majors		Science Majors		Other Majors	
Occupation Category	Rel=0	Rel=1	Rel=0	Rel=1	Rel=0	Rel=1
Administrative / Clerical	6.99	3.53	7.2	0.75	10.91	4.39
Mechanics / Laborers	7.86	1.7	10.09	2.26	11.45	3.14
Service industries	22.27	19.95	17	3.01	20.83	10.5
Human / Protective Service	2.62	0.61	4.9	0.83	4.69	10.4
Finance Professionals	15.72	36.13	5.48	3.01	11.54	6.55
Business and Management	20.52	18.86	15.85	8.65	16.32	11.75
Legal Professionals	0.44	2.07	0.58	0.53	0.81	2.82
Medical Professionals	0	0.61	2.59	5.49	0.81	6.77
Educators	4.37	3.28	5.76	6.54	5.05	17.81
Engineering / Architecture	1.31	1.22	7.2	38.87	1.08	3.74
Research / Scientists / Technical	2.18	2.68	5.48	15.86	3.52	7.58
Computer Science	9.61	5.35	10.66	12.56	5.77	2.6
Editors / Writers / Performers	3.06	2.43	1.73	0.23	4.33	10.23
Other	3.06	1.58	5.48	1.43	2.89	1.73
N	229	822	347	1330	1,109	1,847

Statistics are based on data from the Baccalaureate and Beyond Longitudinal Study. Sample selection and major definitions are provided in the notes to Table 1. The definition of a related job is provided in the notes to Table 2. Unit of observation is an individual-year combination.

Table 6: Model Validation: Summary Statistics

	Business Majors		Science Majors		Other Majors	
	Data	Model	Data	Model	Data	Model
% Selected	26.1	26.7	24.7	24.6	49.2	48.7
	(0.9)	(0.7)	(0.9)	(0.7)	(1.0)	(1.0)
SAT Math	529	539	596	573	525	533
	(3.0)	(8.5)	(2.3)	(3.9)	(1.8)	(4.2)
Residualized SAT Verbal	462	459	467	479	485	478
	(2.2)	(4.4)	(1.8)	(2.4)	(1.2)	(2.2)
Business GPA	3.15	3.17	3.11	3.09	2.78	2.82
	(0.01)	(0.02)	(0.03)	(0.05)	(0.02)	(0.04)
Science GPA	2.76	2.79	3.23	3.21	2.64	2.67
	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)	(0.02)
Other GPA	2.90	2.93	3.16	3.13	3.23	3.24
	(0.01)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)

The numbers in parenthesis are the standard deviation of the averages. For example, we report the standard deviation of the average GPA, not the standard deviation of the individual GPAs. The standard deviations for the model are obtained by bootstrapping the original data set ten times and estimating the parameter vector for each bootstrapped data set. Each one of these parameter vectors is then utilized to simulate data ten times the size of the original data. The standard deviations arise from the dispersion in the calculations of the table across these ten simulated datasets. Details on the model and estimation can be found in Sections 3 and 4.

Table 7: Model Validation: Wage Regressions

	Data	Model	Data	Model	Data	Model
Business	0.191	0.204	0.185	0.206	0.144	0.163
	(0.017)	(0.031)	(0.017)	(0.031)	(0.033)	(0.081)
Science	0.234	0.229	0.214	0.224	0.018	0.052
	(0.018)	(0.032)	(0.019)	(0.032)	(0.037)	(0.044)
Job is Related					0.064	0.064
					(0.022)	(0.085)
Related x Bus.					0.038	0.040
					(0.039)	(0.098)
Related x Sc.					0.228	0.203
					(0.029)	(0.051)
SAT Math/100			0.019	0.017	0.020	0.016
			(0.007)	(0.002)	(0.007)	(0.002)
SAT Residual Verbal/100			-0.043	-0.008	-0.040	-0.008
			(0.009)	(0.003)	(0.009)	(0.003)
Major GPA			0.069	0.062	0.062	0.054
			(0.017)	(0.011)	(0.017)	(0.012)
Year/Grad	Y	Y	Y	Y	Y	Y

The standard deviations for the model are obtained by bootstrapping the original data set ten times and estimating a new parameter vector for each bootstrapped data set. Each one of these parameter vectors is then utilized to simulate data ten times the size of the original data. The standard deviations arise from the dispersion in the calculations of the table across these ten simulated datasets. Details on the model and estimation can be found in Sections 3 and 4.

Table 8: Model Generated Returns to College Major

		Sample Selected by Chosen Major			
	OLS	All	Business	Science	Other
Business	0.204	0.145	0.090	0.177	0.160
	(0.031)	(0.035)	(0.037)	(0.044)	(0.037)
Science	0.229	0.184	0.165	0.200	0.187
	(0.032)	(0.029)	(0.034)	(0.032)	(0.029)
Year/Grad controls	Y	Y	Y	Y	Y

Standard deviations obtained through the bootstrap procedure described in notes to Table 7. The OLS column mimics the wage regression from Table 3 and assumes that all that is observed is the wage for the chosen major. The All column instead includes all the potential wages each individual could have realized had they chosen different majors. The final three columns condition the sample on the chosen major, but include all potential wages.

Table 9: Model Generated Returns to Working in a Related Job

			No Selection into Major			Sorting into Major		
	OLS	FD	All	Rel=0	Rel=1	All	Rel=0	Rel=1
Other	0.070	0.024	0.080	-0.097	0.172	0.083	-0.084	0.186
	(0.044)	(0.027)	(0.029)	(0.036)	(0.049)	(0.027)	(0.040)	(0.046)
Business	0.108	0.063	0.046	0.069	0.042	0.041	0.054	0.038
	(0.060)	(0.051)	(0.122)	(0.052)	(0.145)	(0.080)	(0.069)	(0.107)
Science	0.273	0.159	0.272	0.082	0.341	0.310	0.106	0.363
	(0.030)	(0.023)	(0.057)	(0.048)	(0.083)	(0.071)	(0.038)	(0.093)
Major/Year/Grad	Y	Y	Y	Y	Y	Y	Y	Y

Standard deviations obtained through the bootstrap procedure described in notes to Table 7. The OLS and first differences (FD) column mimics the wage regressions from Table 4 and assumes that all that is observed is the wage for the chosen major and job relatedness. The remaining columns estimate the average returns to a related job, first without conditioning on the chosen major, and then conditioning on the chosen major. The All column includes all the potential wages each individual could have realized had they chosen different relatedness outcomes. The Rel=0 and Rel=1 conditions the sample based on the observed relatedness outcomes.

Table 10: Log-Wage Decomposition

	Business	Science	Other
Wage Variability	0.355	0.318	0.418
	(0.020)	(0.015)	(0.033)
Fraction Related to Initial Human Capital	27.3%	32.3%	26.1%
	(2.3%)	(3.0%)	(1.2%)
General Skill	96.2%	81.0%	91.3%
	(3.9%)	(4.5%)	(2.2%)
Specificity	3.8%	19.0%	8.7%

Standard deviations obtained through the bootstrap procedure described in notes to Table 7. Total wage variation within a major is the result of variation in wages due to human capital, skill prices, year effects, graduate degree effects, and measurement error. Wage variation that arises strictly through human capital and skill prices is the variability related to initial human capital. Variation in wages related only to human capitals is deemed general skill, while variation in wages due to skill prices captures specific skill. Wage variation in wages related to general human capital can be pinned down by fixing the skill prices for each major using a weighted average of the related and non-related prices. We use the predicted proportion of individuals within each major working in related jobs as our weight. Further details on the decomposition can be found in Section 5.2.

A Parameter Estimates - Uncertainty

Table A.1: Skill Distributions

Standard Deviation Math Human Capital, σ_{H_m}	0.142	(0.018)
Standard Deviation Verbal Human Capital, σ_{H_v}	0.027	(0.006)
Standard Deviation Math Scholastic Ability, σ_{ν_m}	0.514	(0.023)
Standard Deviation Verbal Scholastic Ability, σ_{ν_v}	0.115	(0.015)

Table A.2: Major and Relatedness Choice Parameters

Major Choice			Relatedness Outcomes		
Constant, Business	-150.009	(20.755)	Constant, Business	2.146	(0.142)
$A_{m,0}$, Business	-11.963	(7.725)	$H_{m,0}$, Business	4.647	(1.853)
$A^2_{m,0}$, Business	1.039	(0.733)	$H_{v,0}$, Business	52.663	(13.203)
$A_{v,0}$, Business	201.437	(1.151)	Graduate, Business	1.785	(0.315)
$A^2_{v,0}$, Business	-34.271	(0.332)	Persistence, Business	0	-
Parents Foreign, Business	4.102	(1.374)			
Persistence, Business	3.106	(1.359)	Constant, Science	1.520	(0.277)
			$H_{m,0}$, Science	12.583	(2.126)
Constant, Science	-206.722	(24.657)	$H_{v,0}$, Science	22.253	(14.824)
$A_{m,0}$, Science	-11.906	(6.845)	Graduate, Science	1.185	(0.293)
$A^2_{m,0}$, Science	1.569	(0.661)	Persistence, Science	0.223	(0.263)
$A_{v,0}$, Science	83.586	(2.323)			
$A^2_{v,0}$, Science	-7.605	(0.703)	Constant, Other	1.051	(0.212)
Parents Foreign, Science	7.290	(1.730)	$H_{m,0}$, Other	17.076	(2.580)
Persistence, Science	9.716	(1.936)	$H_{v,0}$, Other	22.243	(9.346)
			Graduate, Other	2.179	(0.207)
Relative Potential GPA	64.827	(9.947)	Persistence, Other	0.638	(0.386)
Relative Potential GPA^2	-10.431	(1.605)			
			1997 Dummy	0.226	(0.086)
Own Share	14.614	(5.152)	2003 Dummy	-0.430	(0.087)
			Share Related	1.662	(0.485)

Table A.3: Wage and Measurement Parameters

Wage Coefficients			Measurement Coefficients		
Constant, Business	10.151	(0.054)	SAT Math Constant	5.442	(0.022)
Related, Business	0.048	(0.050)	SAT Verbal Constant	4.744	(0.011)
$H_{m,0}$, Non-Related, Business	2.704	(0.424)			
$H_{m,0}$, Related, Business	1.880	(0.302)	Constant, Business	3.213	(0.027)
$H_{v,0}$, Non-Related, Business	-4.633	(2.758)	$A_{m,0}$, Business	0.750	(0.044)
$H_{v,0}$, Related, Business	-3.518	(2.051)	$A_{v,0}$, Business	0	-
Graduate, Business	0.156	(0.025)	$\eta_{O,B}$	-0.306	(0.044)
			$\eta_{S,B}$	-0.343	(0.058)
Constant, Science	9.991	(0.054)			
Related, Science	0.272	(0.050)	Constant, Science	2.950	(0.038)
$H_{m,0}$, Non-Related, Science	0.361	(0.577)	$A_{m,0}$, Science	0.945	(0.049)
$H_{m,0}$, Related, Science	1.532	(0.384)	$A_{v,0}$, Science	0	-
$H_{v,0}$, Non-Related, Science	-14.110	(3.496)	$\eta_{O,S}$	-0.175	(0.050)
$H_{v,0}$, Related, Science	-6.906	(2.440)	$\eta_{B,S}$	-0.125	(0.051)
Graduate, Science	0.052	(0.030)			
Constant, Other	9.937	(0.035)	Constant, Other	3.316	(0.018)
Related, Other	0.078	(0.037)	$A_{m,0}$, Other	0.658	(0.033)
$H_{m,0}$, Non-Related, Other	0	-	$A_{v,0}$, Other	0.679	(0.168)
$H_{m,0}$, Related, Other	1.292	(0.283)	$\eta_{B,O}$	-0.277	(0.042)
$H_{v,0}$, Non-Related, Other	-14.585	(3.182)	$\eta_{S,O}$	-0.457	(0.039)
$H_{v,0}$, Related, Other	-9.105	(2.281)			
Graduate, Other	0.101	(0.031)	σ_{u_m}	0.986	(0.015)
			σ_{u_v}	0.784	(0.009)
1997 Dummy	0.348	(0.014)	σ_{u_B}	0.339	(0.007)
2003 Dummy	0.462	(0.014)	σ_{u_S}	0.409	(0.010)
			σ_{u_O}	0.384	(0.012)

B Identification Proof

In our empirical implementation we assume that agents can choose among 3 majors: business (B), science (S) and other (O). We normalize the factor loadings of the math and verbal total abilities in the SAT math and verbal measurement equations $(\eta_{m,m}, \eta_{v,v})$ to one. Additionally, we normalize all random variables other than the scholastic abilities to have mean zero. We set the mean of the scholastic abilities equal to the mean of the SAT math and verbal scores. This allows for easier interpretability of the coefficients in the major choice equations. These are done without loss of generality.

We start by considering the identification of the parameters of the GPA measurement equations, the variances of the total abilities $\left(\sigma^2_{A_m}, \sigma^2_{A_v}\right)$, and the parameter $\eta_{v,m}$ that governs the correlation between the math and verbal SAT measures. In order to prove the identification of these parameters consider the following covariances

$$\begin{aligned}
A &= cov\left(SAT_m, SAT_v\right) = \eta_{v,m}\sigma^2_{A_m} \\
B_f &= cov(GPA_f, SAT_m) = \eta_{f,m}\sigma^2_{A_m} \\
C_f &= cov(GPA_f, SAT_v) = \eta_{v,m}\eta_{f,m}\sigma^2_{A_m} + \eta_{f,v}\sigma^2_{A_v} \\
D_{f,f'} &= cov(GPA_{f'}, GPA_f) = \eta_{f,m}\eta_{f',m}\sigma^2_{A_m} + \eta_{f,v}\eta_{f',v}\sigma^2_{A_v}
\end{aligned}$$

calculated for fields f, f' and f'' and all their combinations. Working with these 10 equations it is possible to show that:

$$\eta_{v,m} = \frac{D_{f,f''}C_{f'} - C_{f''}D_{f,f'}}{B_f\frac{B_{f''}C_{f'} - C_{f''}B_{f'}}{A} + D_{f,f''}B_{f'} - B_{f''}D_{f,f'}}$$

Once $\eta_{v,m}$ is identified we can also recover $\left(\sigma^2_{A_m}, \sigma^2_{A_v}\right)$ and $(\eta_{f,m}, \eta_{f,v})$ for all f. The measurement constants can then be identified by looking at the average SAT scores and the average GPAs across major for individuals that have a probability one of selecting each of the three majors.[45]

Treating the measurement parameters and variances of the total abilities as known, we can now study the identification of the labor market parameters $(p_{r,f^*,m}, p_{r,f^*,v})$ and the human

[45] Our identification proof relies on $\eta_{v,m}$ being different from zero. If $\eta_{v,m}$ is equal to zero the identification is more straightforward and utilizes the same measurement equations.

capital variances $\left(\sigma^2_{H_m}, \sigma^2_{H_v}\right)$. Consider first the set of agents that choose with probability one major f or major f' and work in a job with relatedness status r, where $r = 0$ or $r = 1$.[46] Calculating the covariances of the wages with the SAT measures and the auto-covariances of the wages we obtain:

$$
\begin{aligned}
a_f &= cov(\ln w_{r,f,t}, SAT_m) = p_{r,f,m}\sigma^2_{H_m} \\
b_f &= cov(\ln w_{r,f,t}, SAT_v) = \eta_{v,m} p_{r,f,m}\sigma^2_{H_m} + p_{r,f,v}\sigma^2_{H_v} \\
c_f &= cov(\ln w_{r,f,t}, \ln w_{r,f,t-1}) = (p_{r,f,m})^2\,\sigma^2_{H_m} + (p_{r,f,v})^2\,\sigma^2_{H_v}.
\end{aligned}
$$

Working with this set of 6 equations we can solve for the value of $p_{r,f,m}$:

$$
p_{r,f,m} = \frac{a_f c_f \left(b_{f'} - \alpha a_{f'}\right)^2 - a_f c_{f'} \left(b_f - \alpha a_f\right)^2}{a_f^2 \left(b_{f'} - \alpha a_{f'}\right)^2 - a_{f'}^2 \left(b_f - \alpha a_f\right)^2} \tag{12}
$$

and subsequently all other parameters. Now that we have the variance of the human capitals we can also calculate the variance of the scholastic abilities by subtracting the former from the variance of total ability. Once the identification of these parameters is achieved we can apply Kotlarski's theorem (1967) as in Carneiro *et al.* (2003) to show the non-parametric identification of the distribution of the unobserved variables.

What we have not shown here but it is also crucial for our estimation is the identification of the parameters describing the functions $\widetilde{\Upsilon}(f^*, H_{m,0}, H_{v,0}, Z_t)$, $\Psi_f(H_0, \theta, gpa_{f,f}, Z_f)$, the parameters $\alpha^{\Upsilon}_{f^*}$, and the distribution of θ. In this case we approximate the functions with quadratic specifications and apply the standard normalizations (for example $\Psi_f = 0$ for $f =$ Other). By looking at the correlations between the choices and the measurements we can identify all the relevant parameters.

[46] Using the same infinity argument we can also identify the wage constants by looking at the average wages. Similarly, to identify the variances of all measurement errors we can utilize the variances of the associated observable measures.