The role of learning in returns to college major: evidence from 2.8 million reviews of 150,000 professors

Novik, Vitaliy

George Washington University, Washington, United States, United States Census Bureau, Siutland, United States

12 November 2022

The Role of Learning in Returns to College Major: Evidence from 2.8 million reviews of 150,000 Professors

Vitaliy Novik*

November 12, 2022

Abstract

Why do some college majors have much higher returns than others? I ask if differences in returns are due to differences in quality of education across majors. I use a novel dataset in which college students rate courses and professors on difficulty, the level of effort required to obtain an A. Major difficulty correlates positively with study hours, implying students respond to grading standards when choosing study time. Using the American Community Survey, I show that harder majors earn more, under a variety of specifications. To deal with selection concerns I use the College Scorecard to compare graduates from the same university-major but exposed to harder or easier professors due to being in different graduation year cohorts. Those exposed to harder professors earn more one year after graduation. I also use the NLSY97 panel to estimate an event study, finding that difficulty causes lower earnings while in college but higher earnings after graduation and provides access to higher skilled occupations. I estimate that one-third of the variance in returns to major can be explained by differences in learning as proxied by difficulty.

*George Washington University and U.S. Census Bureau: vnovik@gwu.edu. I'm grateful for comments from Burt Barnow, Kathryn Blanchard, Barry Chiswick, Jishnu Das, Alessandra Fenizia, Remi Jedwab, Elira Kuka, Don Parsons, Douglas Webber, and participants at the GWU SAGE Research Pitch, Virginia Association of Economists, and GWU Development Tea. I'm grateful for research assistance from Arthur Wu. Any opinions and conclusions expressed herein are those of the author and do not necessarily reflect the views of the U.S. Census Bureau.

1 Introduction

"O God, Thou sellest all good things to men at the price of effort."

- Leonardo da Vinci,

The difference in returns between low and high paying college majors are as large as the gap in earnings between high school and college graduates. Moreover, the literature consistently finds this gap to be causal [Altonji et al., 2016, Kirkeboen et al., 2016, Bleemer and Mehta, 2022]. Thus much of the difference in returns is likely unrelated to selection, leading to the question, why are there such large differences in returns under seemingly similar investments in education?

One explanation is differences in quality of education by major, or that investment in education differs by major. Some majors are more affected by grade inflation than others [Butcher et al., 2014, Johnson, 2006]. Moreover, I show (Figure 1) that weekly study times vary widely by major, with majors like chemical engineering putting in close to twice as many hours as physical fitness majors in 2017. Furthermore, Arum and Roksa [2011] find that nearly half of college students do not make statistically significant gains on a test designed to measure general critical thinking skills after their first two years of college. They find that students who took courses with more required coursework (reading, writing, assignments) made larger gains. Thus in this paper I ask if differences in learning across majors is an explanation for the large gaps in returns? And if so, how much of the gap can it explain?

As far as I know, I am the first in the literature to study this question, though others have observed that human capital accumulation likely differs by major [Altonji et al., 2016], but until now, data has been lacking.

I circumvent this issue by collecting data from an online site on which college students publicly and anonymously rate their college courses along various dimensions, but most saliently, for *difficulty* on a scale of 1 to 5. Difficulty measures the level of effort needed to obtain an A in a class taught by a certain professor. Intuitively, in courses with low difficulty, students can casually attend class, study for a few hours a week, and obtain a high GPA in the major (Arum and Roksa [2011] describe this type of behavior in their book). Thus, difficulty seems like a good proxy for learning. I collect extensive professor level data and average the professor difficulties within a department to estimate major difficulty at the national and university level. Henceforth, I refer to this value as the *major's difficulty* (See Table 1). My data is based on close to 3 million ratings of about 150,000 professors spanning the last 20 years.

I show that major difficulty correlates strongly with survey measured weekly study hours per week by major (See Figure 1). Besides validating the difficulty data, this shows that in majors with low study hours, students rated the professors as being easy. This implies that the average student closely follows academic expectations within the department. To the extent this is true, grade inflation will lower study time and learning. I also show that though difficulty and study time are related, difficulty includes more information than hours studied per week; this is consistent with difficulty measuring overall study intensity.

Overall STEM professors were rated as hardest at an average difficulty of 3.2. Interestingly, the average difficulty in the Health, Social Sciences, Liberal Arts, and Education fields was below the level of 3, meaning on average, students find these fields to require less effort relative to their core courses.

I begin my analysis by linking major difficulty to the American Community Survey (ACS), and show that a one unit increase in major difficulty is associated with 45% higher annual earnings, 4% higher likelihood of being employed, and 4% more hours worked annually. Moreover, there is a premium to difficulty within every field (e.g., hard liberal arts majors earn more than easier liberal majors). Respondents with more difficult majors also seem to make better entrepreneurs.

To find if learning, as proxied by difficulty, is responsible for differences in returns to major the OLS analysis of the ACS is insufficient. In the ideal random experiment students would be randomly assigned to majors and randomly assigned a difficulty within major. In this case OLS would estimate the average treatment effect among college-attending students.

In the observational data there is likely to be selection. Students who choose to major in something hard may be more motivated, harder working, learn easier, be more eager for financial gain, etc. In short, they may have traits that would have caused them to earn more anyway, making the observational OLS estimates biased.

To deal with selection concerns I rely on two strategies tailored to the strengths of the College Scorecard and National Longitudinal Survey of Youth 97.

The College Scorecard provides median earnings one year after graduation by university-major by graduation cohort-pair cells. This suggests comparing students within the same university-major but who graduated at different times, and as a result experienced different difficulties. Performing this analysis requires review-level data, and I collect this data for the Psychology, History, Economics, and Biology majors from large universities in California and New York. In this analysis, I find a 10% premium to a one unit increase in major difficulty. The effect is not explained through attrition, or reduction of size of the graduating cohort and is robust to using difficulty from upper level courses, where professor difficulty is arguably exogenous from the student's perspective.

Next I turn to the NLSY97, which is a panel tracking respondents over time. The structure of the data suggests a dynamic two-way fixed effects difference in differences strategy. Where under the key assumptions of parallel trends between units in the absence of treatment as well as lack of anticipation, the causal effect can be estimated.

I show that prior to treatment, trends are parallel, but that during the five years prior to college graduation, students in harder majors reduce their earnings and work time, due to the demands of their coursework. After graduation the returns to major difficulty grow rapidly, approaching a 30% premium per unit of difficulty 10 years after graduation. I show that this is mediated through students with harder majors working more hours as well as working in occupations that are higher skilled in all areas. This is consistent with harder majors building more human capital while in college.

I compare the event study results with the results obtained from an OLS analysis of the NLSY data and find similar magnitudes, though the OLS coefficient is smaller (in the literature on returns to education, the OLS estimate is often smaller than the causal estimate [Card, 1999, Bleemer and Mehta, 2022]). This suggests there may be little omitted variable bias in the basic OLS regression.

I then use the ACS data to estimate how much of the variance in major earnings premiums are explained by difficulty. I find that 1/3 of the variance in returns to major can be explained by major difficulty.

My results suggest a conceptual model in which college students are indifferent between a period in college with high leisure and lower future returns, and a period of intense study and much higher future returns. This conceptual model is shown in Figure 2. In exchange for a major with 6 fewer weekly study hours, these students will earn 45% less per year over the course of their careers.

Overall, I find evidence that systemic differences in learning by college major are driving much of the variance in major wage premiums. Evidence is given that grade inflation and its incentive to reduce study time is partially responsible.

My work speaks to the broad literature on returns to college major (see Altonji et al. [2016] for a summary). Works such as Bleemer and Mehta [2022], Kirkeboen et al. [2016], Webber [2014] have shown that choice of major has a causal effect on earnings. I build on the work of Hamermesh and Donald [2008], who show that within the same major, coursework in science and math enters positively into returns to major. My paper contributes by proposing and providing evidence for educational quality as a fundamental reason for the different returns – a new mechanism as far as I know.

My work also contributes to the literature on grade inflation and educational standards at the college level. Butcher et al. [2014] show that grade inflation affects what students major in, and that non-STEM and non-Economics majors were particularly affected by an anti-grade inflation policy at Wellesley College. Babcock and Marks [2011] show that study hours have declined significantly between 1961 and 2003, and the gap in study hours between engineering majors and others, has widened. This work, combined with my findings, suggest differential learning across majors over time, may help explain the rising wage gap between majors [Altonji et al., 2014, Gemici and Wiswall, 2014]. Moreover, some have suggested more grade inflation in STEM may be beneficial in increasing enrollment in these fields by women and minorities[1]. My work implies that an anti-grade inflation policy across the board for all majors as in Butcher et al. [2014] would be the better approach as it would increase study time for low paying majors while helping to shift students into STEM and other high paying majors, while simultaneously raising premiums in lower paying majors.

Finally, my work speaks to the literature on learning at the college level [Arum and Roksa, 2011]. This is the first work to my knowledge, to suggest that learning differs systemically by college major. While Insler et al. [2021] use ratemyprofessors difficulty as a measure of professor standards, I put it forth as stand-alone proxy for learning.

The paper is structured as followsL: Section 2 describes the data, Section 3 provides stylized facts, Section 4 describes the empirical strategy, Section 5 provides results, and Section 6 concludes.

2 Data

Professor Level Difficulty Data. — I collect data on college Professor difficulty from a public website where college students review their professors anonymously. The reviewer (student) rates at the professor-course level, and all reviews include ratings of quality and difficulty on a 1-5 scale, the date of the review, and the course number. Raters are also able to leave comments. The site's guidelines make clear that difficulty is a measure of how much effort the class required in order to achieve an A.

The site aggregates the data at the Professor-level into averages of quality and difficulty, and provides the total number of reviews and the department of the professor. I collect the data at the professor level for convenience and most of the analysis in this paper consists of professor level data. To exploit variation in difficulty within a university-major faced by different cohorts of students, I also collect limited course-level review data, due to the time intensity of its collection.

There were 1155 unique departments in the professor-level data I collected from the website, which I manually mapped to the CIP (Classification of Instructional Programs) and American Community Survey Field of Degree designations to create a crosswalk.

[1]See Ahn et al. [2019] and https://www.washingtonpost.com/education/2021/09/21/why-grade-inflation-is-useful/.

Though students majoring in a certain department take courses in variety of departments, most courses taken will be in their own department thus I use the difficulty of a department as a measure of major difficulty. Moreover, in the American college system, students of all majors tend to overlap on common required courses, making a relative measure of department difficulty advantageous to measure relative differences in study effort across majors.

I clean the dataset by retaining only Professors with at least 4 reviews, university-majors with at least 3 professors, and at least 3 universities in my data that have the major. After this, I am left with data from 218 Colleges and Universities, over 152,000 professors, and difficulties for 118 ACS and 144 CIP majors based on over 2.8 million reviews.

Table 1 shows a summary of the data by field, while Table 2 shows average difficulty by major. Consistent with the interpretation of difficulty as a measure relative to other common coursework, certain fields, such as STEM are rated above average in difficulty, and the overall average professor difficulty rating is close to 3. This table says that the average course in the Education field is not very difficult, requiring little effort in order to obtain an A, and vice versa for STEM.

Given the source of the data, several questions must be answered. First, are the students who voluntarily choose to rate their courses representative of the population of college students? Second, what if the data is contaminated by non-student reviewers, such as Professors seeking to boost their reputation? Third, is the data really measuring learning or is it measuring something else?

Regarding quality of the data, studies comparing ratemyprofessors.com reviews to formal University conducted student evaluations of teaching (SETs) conclude that the site's ratings are valid, correlating well on measures of teacher quality and difficulty. That is, SETs measures of workload and study time correlate strongly with the difficulty ratings on ratemyprofessors.com [Coladarci and Kornfield, 2007, Timmerman, 2008]. Timmerman [2008] notes that the ratings are valid even for those professors with small numbers of reviews.

Regarding difficulty as a measure of learning, Insler et al. [2021] use ratemyprofessors.com data to show that having an easy professor in previous courses has pernicious effects on learning in following sequential courses, with the effect of difficulty larger in magnitude than that of professor quality. Limited survey data on study hours per week of seniors by college major is made available by the National Survey of Student Engagement (NSSE). In Figure 1 I regress major difficulty on 2017 study hours per week. The relationship between difficulty and study hours is strong and positive statistically significant, with an R-sq of 53% (results for the 2011 Study Hours Data can be seen in Figure 10). Aside from validating the difficulty data, this result implies that students first accurately assess grading standards and then respond to these standards when choosing how much to study. As a result, we should expect grade inflation, or the liberal distribution of good grades, to reduce study time.

In addition to calculating the national major difficulty, I calculate the university specific major difficulties, for those majors with enough data. As in Figure 1, I regress national weekly study hours by major on major difficulty at the university-level by university. Figure 8 shows that the correlation coefficient is positive for nearly all universities, implying the data is of high quality even at the university level.

Review Level Difficulty Data. — As mentioned previously, I also collect limited review level data from certain universities and majors in the states of California and New York. I chose these based on coverage in the College Scorecard, which redacts earnings data for small cell sizes for privacy reasons. In all, I collected data for all Professors in the Biology, Psychology, Economics, and History departments (as listed on department faculty pages) from CUNY Hunter College, UC Irvine, San Diego State University, SUNY Albany, CUNY Staten Island, UC Berkeley, University at Buffalo, San Francisco State University, SUNY New Paltz, CUNY Medgar Evers College, Sonoma State Unversity, New York University, Syracuse University, and SUNY Plattsburgh. Using the course number and review date, I first collapse the data at the semester level, and then collapse semesters to estimate the difficulty experienced by each graduation year cohort (or cohort-pair). Since I'm aiming to construct the cohort difficulty for students in the major, I use ratings from courses in the 200 level or higher, and in a robustness check, courses numbered 300 or higher, since 100 level courses are often taken as general education or elective requirements by students outside the major. In Figure 4, I plot the average difficulty by graduation year for three selected universities by department.

In SETs, it's common to see a negative relationship between Professor quality and difficulty, with the distribution of professor quality highly skewed. Using the review-level data, I confirm these patterns in Figure 9, but show that course difficulty tends to be normally distributed within university. Since this is in line with our intuition of difficulty as a relative measure, this implies the difficulty rating, despite correlations with quality, may not suffer from selection bias as much as the quality ratings.

Thus while on the whole the difficulty data appears valid as a proxy of learning by major, there are still sources of measurement error to consider. First, since difficulty is a relative measure, students must be comparing their department's courses against other courses they are required to take. But these vary by major, and intuitively, harder majors tend to take more courses in other hard departments (e.g., engineering students take courses in physics, math, etc, while english major students, take more courses in other liberal arts). Second, harder majors on average enroll smarter and harder working students. Intuitively, a lower ability student is likely to find hard majors even more difficult. Both of these imply that our observed measure of difficulty is biased by regression towards the mean, and the true measure of difficulty if we were to randomly assign students to majors, is likely higher for hard majors and lower for easier majors. This may be the reason why I find the OLS estimates to be similar in magnitude to causal estimates.

The datasets I use to assess outcomes by college major are the American Community Survey (ACS), the College Scorecard, and the NLSY97 (National Longitudinal Survey of Youth 97).

American Community Survey. — The ACS is a representative survey of the US population conducted annually by the US Census Bureau and used extensively in this field[2]. Since 2009, respondents' field of degree has been coded from among 173 options, making it the most detailed dataset available for assessing outcomes by college major. After cleaning the data[3] I am left with 2.5 million cross sectional observations spanning 2009 to 2019.

For covariates, the ACS includes the respondents' sex, race, population of their city, their state and county of residence, their citizenship status, year of immigration, their residence in a metropolitan area, detailed place of birth, occupation, and industry. As outcome variables the ACS has wage income, income from business and farming, total income, employment status (employed, unemployer, not in labor force), as well as total hours worked in the year.

College Scorecard. — The College Scorecard data is created by the US Department of Education, and consists of median earnings by University and Major for pairs of graduation year cohorts. Currently, earnings data is available 1 year after graduation for the 2016 and 2017 pooled graduation year cohorts, and 1 and 2 years after graduation for the 2017 and 2018 pooled cohorts, thus I focus on earnings 1-year after graduation for my analysis.

Note that the earnings data are calculated only from the sample of graduates that received federal aid in the form of grants or loans and who were working[4]. Importantly, for the 2017-18 cohorts data release, the earnings are grouped by "highest credential", meaning, I can compare earnings of students who received bachelor's degrees only – those who received a masters or higher are excluded. This is not the case with the 2016-2017 data.

Majors in the Scorecard are categorized using CIP codes, which are more detailed than

[2]Accessed using Steven Ruggles and Sobek [2021]

[3]I retain only those respondents whose highest educational level is a 4-year degree and are not currently in school. To minimize bias, only those of ages 24-64 are retained in the sample, and only those with positive income from wages, and working more than 20 weeks per year and more than 15 hours per week are included in regressions of earnings. I estimate experience from age and assume entry into the workforce at age 22; I deflate the income from wages into 2010 dollars. For respondents with more than one major, I assign a major randomly.

[4]For the 2015-16 academic year, this amounts to 60% of undergraduates at public institutions and 64% at private institutions. See https://collegescorecard.ed.gov/assets/FieldOfStudyDataDocumentation.pdf – accessed 3/14/22. Foote [2022] compares earnings of those who received aid (Scorecard) with all graduates using the Census' PSEO (Post Secondary Employemnt Outcomes) and finds that short run earnings are very similar for the two samples, though long run earnings are different.

the ACS categories.

NLSY97. — The National Longitudinal Survey of Youth 97 (NLSY97) is conducted by the Bureau of Labor Statistics and is a nationally representative sample of 8,984 men and women, born between 1980 and 1984. The cohort was first interviewed in 1997 when respondents were aged 12-17, and has been interviewed 19 times since then.

Not only does the NLSY97 include extensive individual level information that help control for selection on observables (high school GPA, high school math level, high school percentile on the ASVAB test of verbal and mathematical reasoning, parental income and assets, parental education), its panel structure allows for two-way fixed effects difference in differences and event study estimation strategies.

I retain the respondents whose maximum completed education was a bachelor's degree. About 2500 respondents had obtained at least a bachelor's degree, and about 840 of these had obtained at least a master's degree, leaving about 1700 respondents for analysis.

The NLSY97 records respondents' major into 35 categories, roughly at the 2-digit 2010 CIP Code level.

3 Descriptive Results

This section presents a series of stylized facts using the ACS.

Figure 3 shows that there is a strong positive relationship between major difficulty and earnings for both men and women using the ACS. To describe the relationship between major difficulty and key outcomes more precisely, I use the following model:

$$LogInc_{iymsc} = \beta_0 + \beta_1 Difficulty_m + \chi_i \gamma + \epsilon_{iymsc} \tag{1}$$

Where the log of income for individual i in year y with major m in state s and year-of-birth cohort c is regressed on difficulty and various controls. Where χ_i includes the individual's race interacted with their sex, population of their city, their state and county of residence, their citizenship status, year of immigration, their residence in a metropolitan area, detailed place of birth, whether or not they have two degrees, state-year fixed effects, and year of birth cohort fixed effects. Standard errors are clustered at the state level. The coefficient of interest in this model is β_1, the percentage by which annual earnings increase for a one unit increase in major difficulty. Table 3 Model 1 shows that for a one unit increase in difficulty, annual wage increase by 46% annually. Using weekly study hours as an IV for difficulty to address measurement error concerns gives a similar value (Model 2). In Model 3, adding additional controls on county, citizenship, immigration, and place of birth does not alter the estimates. Models 4 and 5 show that the premium to difficulty has the same magnitude among men and women. In Model 6 I use the log wage rate as the dependent variable, and the magnitude is unchanged. In Model 7 I use the log of total income from all sources (includes wage, farm, business income, and more), and again the coefficient is unchanged.

In Table 4 I examine several other relationships of interest. Model 1 shows that interacting difficulty with work experience implies that more difficult majors see steeper gains to work experience. In Model 2, I control for study hours, and the coefficient on difficulty grows and remains significant (the coefficient on study hours is significant as well). Model 2 implies that major difficulty is measuring more than just study hours. One explanation is that difficulty is measuring the intensity of studying which usually is unobserved.

One concern is that with the publication of professor level ratings on sites such as ratemyprofessors.com, selection effects may have grown stronger over time, implying more recent cohort returns will be higher. Model 4 tests this, by running the regression separately for respondents who graduated before the year 2000 (based on age), and finds that the premium to difficulty is larger than with the whole sample. This suggests selection into major has not grown stronger over time. One explanation is that while platforms like ratemyprofessors have made individual professor level ratings public, the data has not been aggregated to construct major difficulties uuntil now.

Model 4 adds occupation and industry fixed effects and finds that the premium to difficulty decreases by 3/4. Thus, while much of the premium to more difficult majors is driven through occupation and industry, 1/4 of the premium is unrelated to industry or occupation.

In Model 5, I use the log of business income as the dependent variable, the positive coefficient implies that respondents with difficult majors make better entrepreneurs. In Model 6 I set the dependent variable to employment status and find that a one unit increase in difficulty corresponds to 4% higher chance of being employed. Model 6 shows a unit increase in difficulty corresponds to working 4% more hours annually as well (excluding those unemployed or out of the labor force).

In Table 5 I show that when running the regression separately by field, all fields have a positive premium to difficulty, though the magnitude varies from a high of close to 50% within the social sciences to a low of 5% in the liberal arts. In Tables 6 and 7 I include the nonworking by regressing on the log of income+1, and find that the previously discussed results all hold, and even grow stronger once taking into account the decision to exit the labor force as well as the unemployed.

4 Empirical Strategy

To estimate the effect of difficulty on learning I rely on the following strategies, each tailored to the respective strengths of the College Scorecard and the NLSY97.

College Scorecard. — The College Scorecard provides earnings by university-major for pooled graduation year cohorts over time. This implies an identification strategy of comparing cohorts who graduated from the same university-major but for exogenous reasons were treated with different levels of difficulty, i.e, a two-way fixed effects difference in differences specification, where the unit is the university-major:

$$LogMedInc_{umy} = \beta_0 + \beta_1 Difficulty_{umy} + \alpha_{um} + \lambda_y + \epsilon_{umy} \tag{2}$$

In Equation 2, the log median income one year after graduation at university u in major m in years pair y is regressed on the cohort's average experienced difficulty (as described in the review level data section), university-major fixed effects and year fixed effects.

For β_1 to have a causal interpretation, the cohorts entering each university-major must be similar in unobservables over time, where the only difference is exogenous differences in treatment of difficulty, i.e, some are assigned harder professors and some are assigned easier professors. If difficulty is correlated with unobservables that also correlate with earnings, the coefficient will be biased.

Arguably, students entering a university-major over time are similar, and in the event that students rate professors as more difficult, it is unlikely that they are of higher ability, since our prior is higher ability students would find coursework to be easier. However, it remains possible that perhaps, professors teach to the class, and noticing students of higher ability increase the rigor of coursework. However, this seems unlikely.

By using the difficulty ratings of 200-400 level courses or even 300-400 level courses, I argue the difficulty assignment becomes exogenous. As students progress in their coursework, the option set of courses shrinks until only required courses remain, leaving little room for selection into or out of difficult professors.

Finally, since I am using median earnings, another way difficulty could increase earnings apart from learning, is though attrition of lower ability students. Since the scorecard provides the number of graduates by cohort, this is something I can check for.

NLSY97. — The NLSY97 is structured as a panel, suggesting an event study specification:

$$LogInc_{it} = \alpha + \sum_{k=-n}^{-1} \beta_k (lead\, k) difficulty_{m(i)} + \sum_{j=0}^{m} \beta_j (lag\, j) difficulty_{m(i)} + \gamma_i + \delta_t + X_{it} + \epsilon_{it} \tag{3}$$

Where the Log of Income of individual i at year t is regressed on leads and lags of the difficulty of the major individual i attains $m(i)$) along with individual fixed effects γ_i and year fixed effects δ_t along with time varying individual controls X_{it} (experience and its square, high school and associates degree completion status, and age fixed effects).

For comparison, I also run a two-way fixed effects difference in differences specification.

$$LogInc_{it} = \gamma_i + \delta_t + \beta difficulty_{m(i,t)} + \epsilon_{it} \tag{4}$$

Where log of income is regressed on individual and year fixed effects, as well as the difficulty of the individual's major.

In difference-in-differences models, the identification assumptions are that apart from treatment, individuals' outcomes would have followed parallel trends, and that the treatment does not have a causal impact before its implementation.

It usually takes 4 to 5 years to obtain a bachelors degree, and the effects are likely to be present pre-graduation depending on difficulty of major (such as difficult majors having less time available to work). Hence, to satisfy the no anticipation assumption, I use six years prior to the receipt of the degree as the baseline period instead of the usual one year prior.

The parallel trends assumption can be partially tested by examining the pre-trends. Prior to treatment, there should be no difference between the treated and untreated (later treated) units.

5 Results

In this section I discuss the results from the College Scorecard and NLSY97.

College Scorecard. — In Model 1 of Table 8, I start by regressing log income on the university specific major difficulty (but not specific to the cohort) as well as year fixed effects, and find a premium of 26% to each unit of major difficulty. In Model 2, I add University fixed effects, growing the premium. In Model 3, I turn to the specification in Equation 2 and add university-major fixed effects and use the difficulty specific to the cohort (a sample of this variation can be seen in Figure 4). The coefficient falls to 9.5% but remains significant. As discussed earlier, it's possible for the effects to be coming from attrition of lower ability students, not learning. To test this, I regress the graduation class size on cohort specific major difficulty and do not find a negative effect. Thus, there is no evidence that the results are generated through attrition.

The earlier results calculated cohort specific major difficulty using ratings on courses from the 200-400 level. As a robustness check, I rerun the results using cohort difficulty from 300-400 level courses. As students progress later into their studies, professor difficulty is more plausibly exogenous. Table 9 shows that the results are robust to this, and once again the results are not driven by attrition.

NLSY97. — In Table 10 Model 1, I first regress log income on the basic set of controls similar to those contained in the ACS: indicators for possession of high school and associates degrees, age fixed effects, experience and its square, year fixed effects, and sex by race fixed effects. The coefficient on difficulty is about 18.7%. Note, that this coefficient is likely smaller than that of the ACS because of large measurement error in difficulty, due to the NLSY's having just 35 categories of majors.

In Model 2 I add personal controls, including ASVAB test percentile, high school GPA, and citizenship status. Since some respondents don't have these, I set missing values to zero and fully interact each control with an indicator for missingness. The addition of these controls does not alter the coefficient on difficulty, but does lower standard error. In Model 3, I add parental controls for family income while in high school, household assets, fathers and mothers highest grade completed, dealing with missing values as before. The coefficient on difficulty is unchanged. In Model 4, I eliminate all individual specific and time-invariant controls and instead use an individual fixed effect as in Equation 4. The coefficient increases to 22%. Implementing the De Chaisemartin and D'Haultfoeuille [2022] correction for heterogeneous and dynamic effects in Model 5, further increases the coefficient. Together, these results imply there is virtually no omitted variable bias in a simple OLS specification as in Model 1. Though this may not be true for all cohorts or samples.

I next turn to the dynamic specification as in Equation 3. Panel A of Figure 6 shows that upon entering college, those enrolled in more difficult majors experience a steeper decline in income, falling by almost 10% per unit increase in difficulty. Figure 7 indicates that this is mediated through people in harder majors working fewer hours while in college. Beyond studying more and with greater intensity, people in harder majors forego more income while still in college due to the demands of their coursework, thus investing more in their college education on multiple dimensions.

Note that the parallel trends assumption seems valid, since the placebo coefficients prior to entering college show no effect and appear trendless.

Upon graduation Figure 6 shows that earnings rise rapidly, with a relative income premium of 18% one year after graduation. The earnings premium per unit of difficulty continues to rise. At the ten year point, the premium to a unit of difficulty is 30% or about $12,000 higher earnings.

Figure 7 indicates that this is mediated through 5% more hours worked per year per unit of difficulty, about the same as estimated in the ACS.

To gain more insight into how difficulty affects skill, I match the occupations in the NLSY to the O*NETs 2019 skills file, in which experts and workers rate occupations on the skill level from 1 to 7, for various key skills. This match provides a panel of skill for each NLSY97 respondent, based on their occupations over time.

In Figure 5 I show the effect of difficulty on eight key skills: Mathematics, Science, Critical Thinking, Complex Problem Solving, Learning Strategies, Writing, Active Learning, and Speaking. Prior to treatment (entering college), the placebo checks show no difference in occupational skill levels. After obtaining their degree, difficult majors see gains in all areas of skill. But especially in Mathematics, Critical Thinking, and Learning Strategies.

These results are consistent with difficult majors developing more all around skills while in college, and thus working in higher skilled occupations after graduation.

To summarize, the NLSY97 data shows that difficulty increases earnings, through opening doors to more skilled occupations, and through more hours worked.

5.1 Counterfactual Exercises

Thus far in this paper I have shown that differences in difficulty across majors are responsible for much of the observed differences in returns to major. Now, I seek to estimate how much of the variance in major fixed effects can be explained by difficulty.

To do this analysis, I will assume that the point estimates generated by the ACS have a causal interpretation (as the NLSY97 analysis indicates no omitted variable bias). First, I regress log total income on the full set of controls from the ACS as before, leaving out difficulty but adding major fixed effects. Second, I regress the values of the major fixed effects on the major's difficulty, in order to estimate the share of variance in fixed effects explained by difficulty. This analysis abstracts from general equilibrium concerns and assumes that the premium on difficulty remains constant.

I find that 1/3 of the variance in major fixed effects in the ACS can be explained by major difficulty.

5.2 What's Driving Differences in Difficulty

Why some majors are more rigorous than others, and how this came to be, or if it was always the case, is outside the scope of this paper. However, this is a valuable area for future research, as the literature indicates academic standards in college have declined over time, and differentially by field [Babcock and Marks, 2011].

Given the definition of difficulty as a rating of how hard it is to obtain an A, intuition suggests majors with more grade inflation are likely to be easier. Or, that grade inflation will result in lower major difficulty. In Figure 11, I show that the relationship between major GPA and major difficulty is negative and significant at the 10% level for UC Berkeley[5]. This implies grade inflation reduces future earnings via a reduction in learning.

Likewise, Ahn et al. [2019] find a negative relationship between the average grade for a course and study time using detailed data from the University of Kentucky. They interpret this result as suggestive that grade inflation has a negative effect on learning.

6 Conclusion

The question of why some majors have much higher returns than others has been unanswered in the literature on higher education. In particular, the role of disparate learning, or educational quality by major, though hinted at, has not been investigated due to the lack of data.

I identify and use a novel data source: an online site on which college students publicly rate their professor-course pairs along various dimensions, but most saliently, for difficulty on a scale of 1-5, where difficulty is defined as the study effort required to get a good grade in the class. This extensive data spanning 2000 through 2022, allows me to measure difficulty at the detailed major level, as well as over time at the university-major level.

I show that difficulty is positively correlated with weekly study hours by college major, validating its use as a proxy for learning, and implying students respond strongly to grading standards when choosing study time.

Using the American Community survey I show that respondents with more difficult majors earn more, work more hours every year, and are less likely to be unemployed. Moreover, there is a premium to difficulty within every field (Business, Education, STEM, Liberal Arts, Social Sciences, and Health). There is also evidence that they make better entrepreneurs.

For identification I turn to the College Scorecard and the NLSY97. In the Scorecard, I compare graduates from the same university-major but who graduated at different times and thus were exposed to more or less difficult professors. Those exposed to more difficult professors earn 10% more per unit of difficulty one year after graduation. These results are robust to using course difficulty from 300 and 400 level courses, where most courses tend to be required and assignment of difficulty is plausibly exogenous.

[5] See https://pages.github.berkeley.edu/OPA/our-berkeley/gpa-by-major.html – accessed 3/23/22.

Next I turn to the NLSY97, where the panel structure of the data invites an event study specification. I find that prior to entering college, treated and later treated groups are similar but upon entering college, respondents enrolled in difficult majors earn and work less while in college, and earn 30% more per unit of difficulty 10 years after graduation. These estimates are similar in magnitude to the OLS estimate, implying little omitted variable bias.

This paper helps explain why the returns to major are so large, identifying differences in learning across majors as responsible for around 1/3 of the variance in returns to major. This is the first paper to consider this channel in the literature.

Recently, the literature has noted the positive effects of grade inflation in increasing college completion rates [Denning et al., 2022]. My work implies that grade inflation has large negative effects on future earnings. Intuitively, this suggests the importance of measuring quality of graduates, not only quantity.

7 Tables

Table 1: Summary of Difficulty Data

Discipline	Mean	Median	CV	Universities	Majors	Professors	Ratings
All	2.94	3	0.27	218	118	152275	2814972
STEM	3.19	3.2	0.25	192	51	42327	859038
Business	3.04	3.1	0.27	171	9	13160	255761
Health	2.88	2.9	0.31	144	7	3986	50153
Social Sciences	2.86	2.9	0.27	175	16	28670	591492
Liberal Arts	2.81	2.8	0.28	199	28	58115	983136
Education	2.6	2.5	0.33	156	7	6017	75392

Note: Summary of the difficulty data collected for this paper using the ACS major classification.

Table 2: Average Difficulty by Major (CIP Classification)

Dept	Difficulty	Dept	Difficulty
Molecular Biosciences	3.62	Literature	2.88
Chemical Engineering	3.48	Graphic Arts	2.87
Botany	3.46	Management	2.87
Biomedical	3.46	Aviation	2.86
Materials Science	3.42	Natural Resources	2.85
Genetics	3.42	Sports Management	2.85
Aerospace Engineering	3.42	Environmental Studies	2.85
Chemistry	3.42	Agriculture	2.84
Accounting	3.39	Design & Merchandising	2.84
Physics	3.36	Japanese	2.84
Finance	3.35	Engineering Technology	2.84
Electrical Engineering	3.34	Geography	2.84
Bioengineering	3.34	Urban Studies	2.82
Visual Communication Design	3.31	Anthropology	2.82
Neurological Sciences	3.31	Textiles & Clothing	2.81
Biology	3.31	Arabic	2.81
Physiology	3.31	Chicano Studies	2.81
Anatomy Physiology	3.30	Cultural Studies	2.81
Architecture	3.28	Nutrition Food Science	2.81
Microbiology	3.28	Sign Language	2.81
Nuclear Engineering	3.27	English	2.80
Wildlife	3.26	Psychology	2.80
Mechanical Engineering	3.26	Classics	2.80
Athletic Training	3.24	Nutrition	2.79
Physical Sciences	3.23	Industrial Technology	2.79
Materials Engineering	3.21	Theology	2.78
Law	3.19	Languages	2.78
Engineering	3.19	Culinary Arts	2.78
Physics & Astronomy	3.18	Computer Information Systems	2.76
Civil Construction Environmental Engineering	3.18	Russian	2.75
Ecology & Evolutionary Biology	3.17	Conflict Analysis	2.75
Economics	3.17	Communication	2.75
Biomedical Engineering	3.17	Social Work	2.75
Computer Engineering	3.15	Humanities	2.74
Nursing	3.14	Public Health	2.73
Computer Science	3.12	Tourism Management	2.72
Civil Engineering	3.12	Human Dev & Family Sciences	2.72
Statistics	3.11	Public Relations	2.71
Mathematics	3.10	Portuguese	2.71
Counseling	3.09	Religion	2.69
Design	3.09	Family Studies	2.68
Industrial Engineering	3.09	Kinesiology	2.68
Management Information Systems	3.06	Criminal Justice	2.68
Animal Science	3.06	Sociology	2.68
Paralegal	3.04	Education	2.67
History	3.03	Ethnic Studies	2.67
Rehabilitative Studies	3.01	Music	2.66
Forestry	2.99	Criminology	2.66
Political Science	2.99	Health Science	2.66
Construction	2.98	Human Resources	2.65
Medicine	2.98	Horticulture	2.65
Human Services	2.98	Health Care Administration	2.65
Geology	2.98	Nutrition Diebetics	2.65
Speech & Hearing Sciences	2.97	Social Science	2.64
Information Science	2.96	Marine Sciences	2.63
Entomology	2.95	Recreation	2.63
Latin	2.95	German	2.63
Business	2.94	Human Development	2.61
Journalism	2.94	Dance	2.61
Library Science	2.93	Creative Writing	2.57
Meteorology	2.92	Allied Health	2.56
Philosophy	2.92	Child Development	2.56
International Studies	2.92	Hospitality	2.55
Writing	2.92	Interdisciplinary Studies	2.55
Food Science	2.91	Elementary Education	2.54
Public Policy	2.91	Telecommunications	2.53
Fine Arts	2.91	Educational Psychology	2.52
Information Technology	2.90	Theater	2.52
Visual Arts	2.90	Behavioral Sciences	2.50
Urban Planning	2.89	Special Education	2.49
Mathematics & Statistics	2.89	English As A Second Language	2.45
Marketing	2.88	Physical Ed	2.31

Table 3: ACS: LogInc Regressed on Average Major Difficulty table

	(1)	(2) HrsIV	(3)	(4) Men	(5) Women	(6)	(7)
Difficulty	0.456***	0.401***	0.451***	0.468***	0.431***	0.409***	0.440***
	(0.011)	(0.024)	(0.007)	(0.008)	(0.009)	(0.008)	(0.007)
Observations	2463325	1902458	2463324	1245729	1217595	2456242	2572542
Dependent Var	WageInc	WageInc	WageInc	WageInc	WageInc	HrlyWage	TotInc
R-Sq	0.16	0.16	0.20	0.19	0.13	0.20	0.20
Cty/Cit/Immig FE			X	X	X	X	X

SE Clustered at State Level in parantheses, State-Year Fixed Effects in all regressions, Controls for experience, experience-sq, race x sex, year of birth cohort, and indicator for more than one degree in all regressions. Those working more than 15 hours per week and 20 weeks per year. All dependent variables in logs.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: ACS: LogInc Regressed on Average Major Difficulty - Additional Checks

	(1)	(2) StudyHrs	(3) pre2000BA	(4) OccIndFE	(5)	(6)	(7)
Difficulty	0.345***	0.530***	0.467***	0.122***	0.264***	0.044***	0.039***
	(0.020)	(0.011)	(0.009)	(0.004)	(0.025)	(0.002)	(0.004)
Difficulty X Exp	0.010***						
	(0.002)						
Difficulty X Exp2	-0.000***						
	(0.000)						
Observations	2463324	1902457	1538393	2657657	226182	3240246	2574519
Dependent Var	WageInc	WageInc	WageInc	WageInc	BusInc	Employed	HrsWrkd
R-Sq	0.20	0.20	0.16	0.34	0.07	0.09	0.06

SE Clustered at State Level in parentheses. Controls for State-Year Fixed Effects, experience, experience-sq, race-by-sex, year of birth cohort, indicator for more than one degree, detailed place of birth (US States and Countries), city population, citizenship, and year of immigration in all regressions. Those working more than 15 hours per week and 20 weeks per year (except Employed model - Model 6). All dependent variables are in logs except for Model 6, and have a percentage change per unit increase in difficulty interpretation.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: ACS: LogInc Regressed on Average Major Difficulty By Field

	(1) STEM	(2) Busin	(3) Educ	(4) Health	(5) Lib Arts	(6) Social Sci	(7) Engineer
Difficulty	0.203***	0.230***	0.106***	0.280***	0.046*	0.487***	0.347***
	(0.021)	(0.013)	(0.023)	(0.024)	(0.027)	(0.022)	(0.044)
Observations	545744	678804	197430	176537	492904	371905	199808
Dependent Var	WageInc	WageInc	WageInc	WageInc	WageInc	WageInc	WageInc

SE Clustered at State Level in parentheses. Controls for State-Year Fixed Effects, experience, experience-sq, race-by-sex, year of birth cohort, indicator for more than one degree, detailed place of birth (US States and Countries), city population, citizenship, and year of immigration in all regressions. Those working more than 15 hours per week and 20 weeks per year. Engineers in Model 7 are a subset of STEM majors.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: ACS: LogInc Regressed on Average Major Difficulty For All (working and non-working) Respondents

	(1)	(2) Men	(3) Women	(4)
Difficulty	0.812***	0.703***	0.872***	0.626***
	(0.030)	(0.026)	(0.042)	(0.019)
Observations	3240246	1507071	1733175	3235818
Dependent Var	WageInc	WageInc	WageInc	TotInc
R-Sq	0.09	0.08	0.08	0.10

SE Clustered at State Level in parentheses. Controls for State-Year Fixed Effects, experience, experience-sq, race-by-sex, year of birth cohort, indicator for more than one degree, detailed place of birth (US States and Countries), city population, citizenship, and year of immigration in all regressions. Dependent variable is log(income + 1) to include non-working respondents.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: ACS: LogTotalInc Regressed on Average Major Difficulty For All (working and non-working) Respondents

	(1) STEM	(2) Busin	(3) Educ	(4) Health	(5) Lib Arts	(6) Social Sci	(7) Engineer
Difficulty	0.101***	0.272***	0.328***	0.323***	0.138***	0.473***	0.280***
	(0.019)	(0.009)	(0.020)	(0.024)	(0.034)	(0.021)	(0.038)
Observations	641818	811558	261105	207699	626960	462502	228989

SE Clustered at State Level in parentheses. Controls for State-Year Fixed Effects, experience, experience-sq, race-by-sex, year of birth cohort, indicator for more than one degree, detailed place of birth (US States and Countries), city population, citizenship, and year of immigration in all regressions. Those working more than 15 hours per week and 20 weeks per year. Dependent variable is log(totalincome + 1) to include non-working respondents.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Pooled College Scorecard Log Median Earnings Regressed on Difficulty

	Median Log Earnings 1 Year After Graduation			LogGrads
	All	Uni FE	UniMaj FE	UniMaj FE
	(1)	(2)	(3)	(4)
Difficulty	0.257***	0.261***	0.095**	0.075
	(0.071)	(0.073)	(0.042)	(0.083)
Observations	6,077	6,077	66	66
R^2	0.099	0.230	0.959	0.992
Adjusted R^2	0.099	0.208	0.911	0.983

Note: Standard errors clustered at the 4-digit college major (CIP CODE) level in parentheses. Year-Pair fixed effects in all regressions. University-Major difficulty is calculated using difficulty ratings from sophomore through senior level undergraduate courses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Pooled College Scorecard Log Median Earnings Regressed on Difficulty

	Median Log Earnings 1 Year After Graduation	LogGrads
	UniMaj FE	UniMaj FE
	(1)	(2)
Difficulty	0.059***	0.037
	(0.018)	(0.064)
Observations	65	65
R^2	0.959	0.992
Adjusted R^2	0.910	0.983

Note: Standard errors clustered at the 4-digit college major (CIP CODE) level in parentheses. Year-Pair fixed effects in all regressions. University-Major difficulty is calculated using difficulty ratings from junior through senior level undergraduate courses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: NLSY: LogInc Regressed on Average Major Difficulty table

	(1)	(2)	(3)	(4) TWFEDD	(5) didmultiplegt
Difficulty	0.187***	0.187***	0.189***	0.222***	0.263***
	(0.018)	(0.017)	(0.016)	(0.014)	(0.020)
N	20344	20344	20344	20344	29904
R-Sq	0.71	0.71	0.72	0.79	–
Personal		X	X		
Parental			X		
Individual FE				X	X

SE Clustered at field of degree level in parentheses. All regressions control for experience and its square, high school completion, aa degree completion, age fixed effects, and year fixed effects. didmultiplegt computes estimates that are unbiased under heterogenous and dynamic effects as described in De Chaisemartin and D'Haultfoeuille [2022]

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

8 Figures

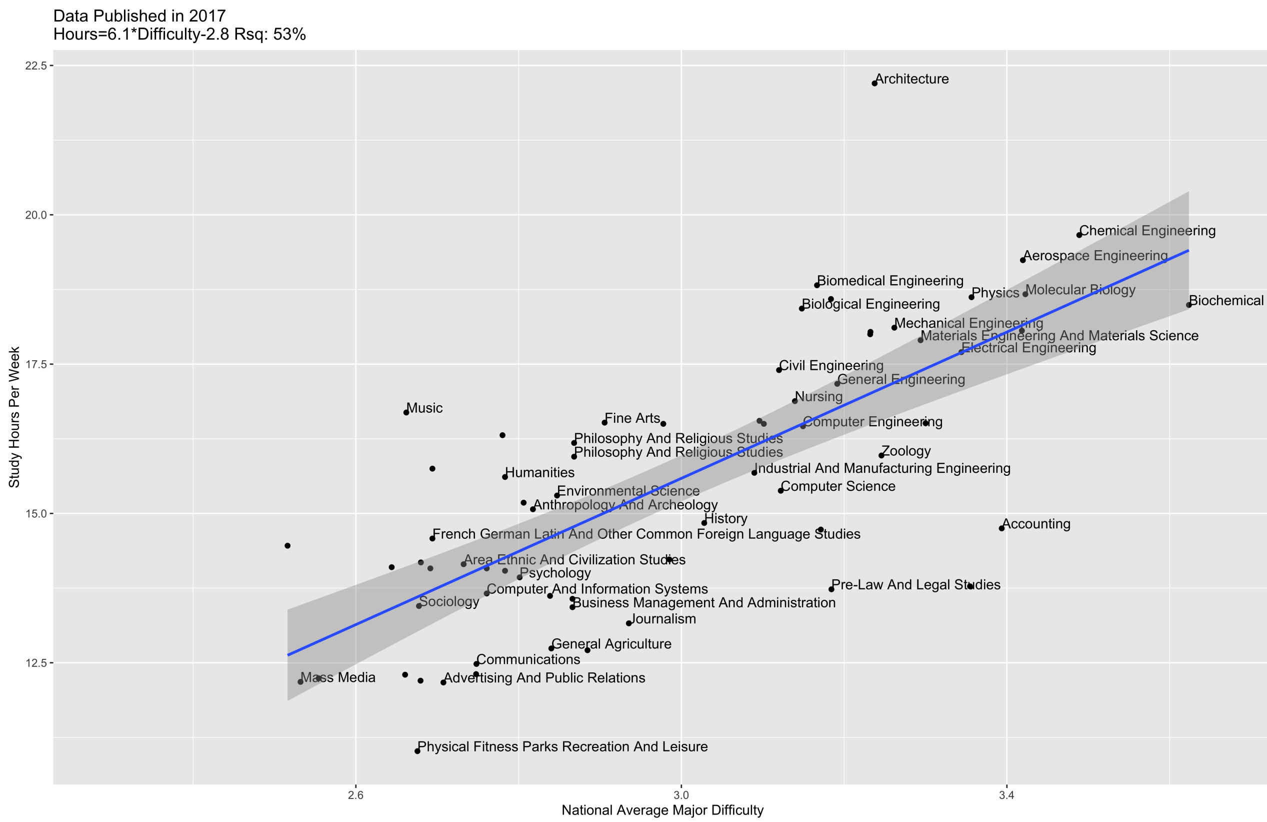


Figure 1: Study Hours Per Week by Major is plotted against the Average national major difficulty. The study hours data is collected by the National Survey for Student Engagement (NSSE) of Indiana University. Average national difficulty data was compiled and calculated by the author. Source: https://thetab.com/us/2017/02/06/ranked-majors-work-hardest-59673.

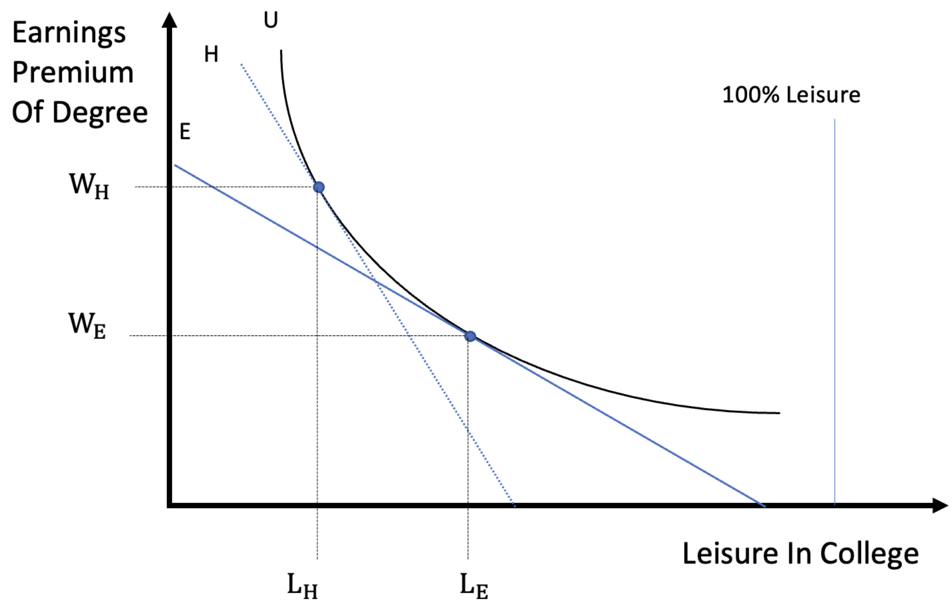


Figure 2: A conceptual model, with a hard (H) and easy (E) major. The budget lines E and H cross the X-axis at a level below the 100% leisure point, since all programs require some effort to stay enrolled. In this model, the easy major takes a high level of leisure (L_E) while in college and receives a low wage premium W_E after graduation. To be indifferent between the easy and hard major, the the hard major must have a higher wage premium W_H after graduation.

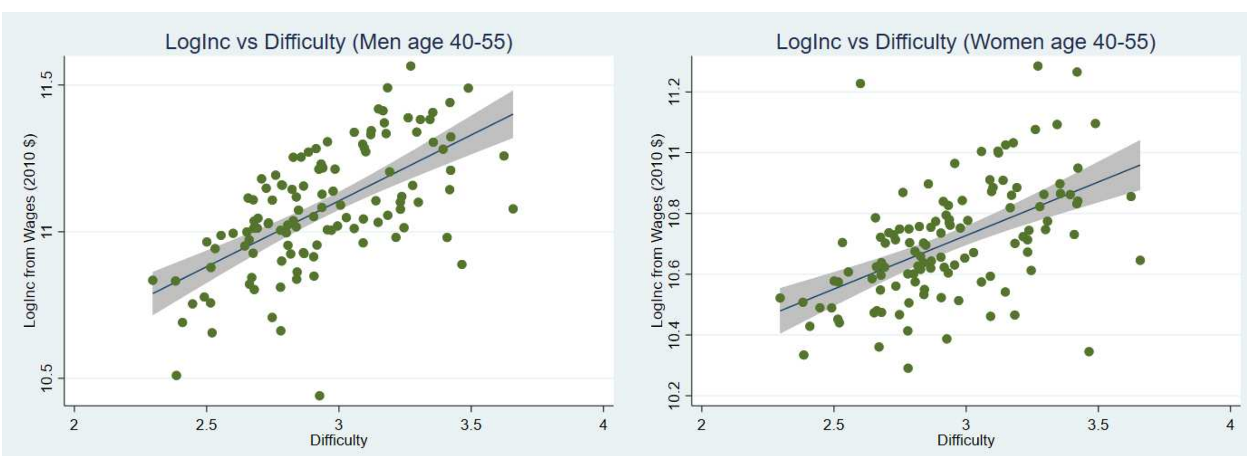


Figure 3: Collapsed Log Wage Earnings for employed men and women age 40 to 55 by major difficulty using ACS data.

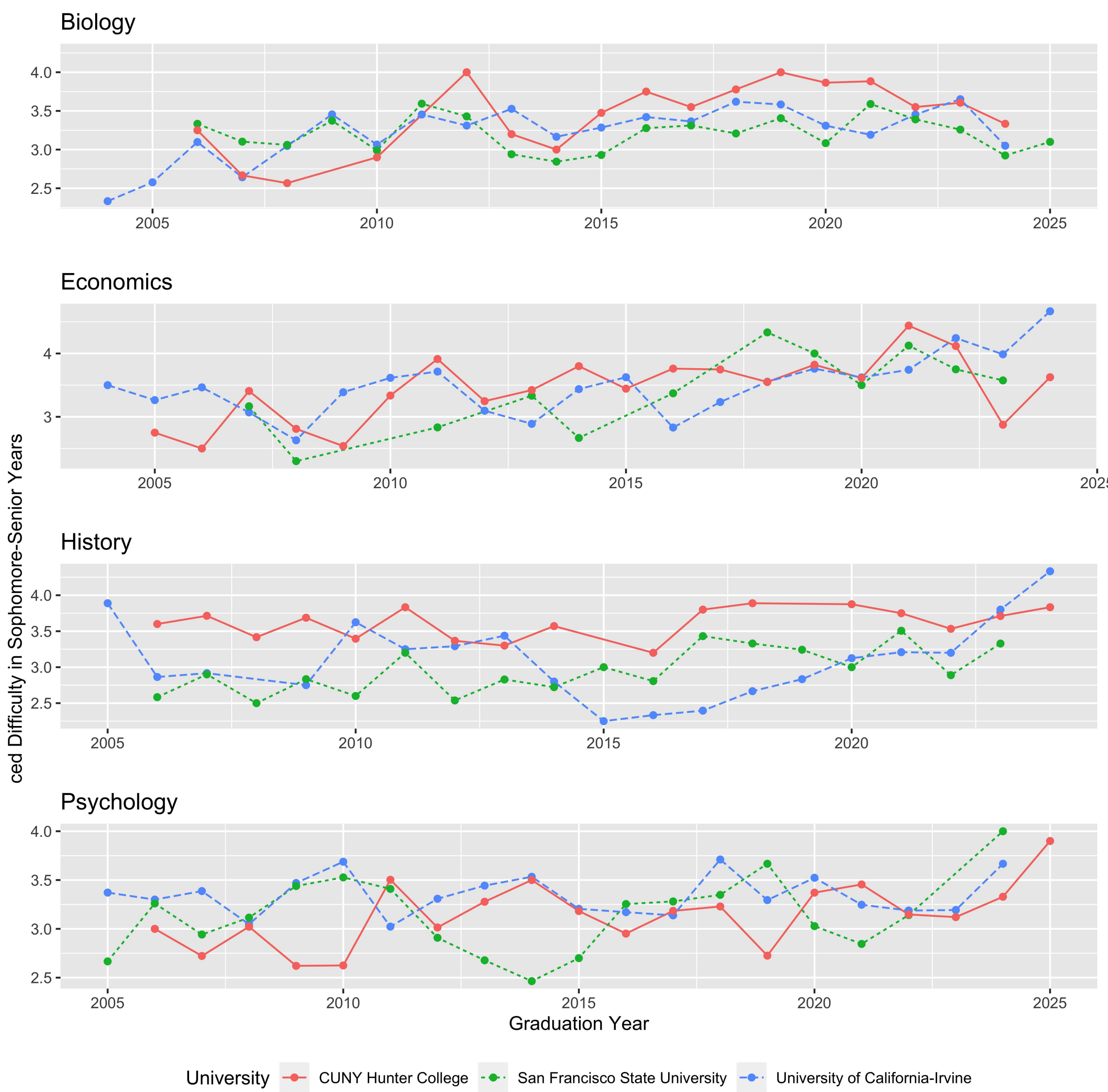


Figure 4: This figure plots difficulty by graduation year for selected universities, using review-level data.

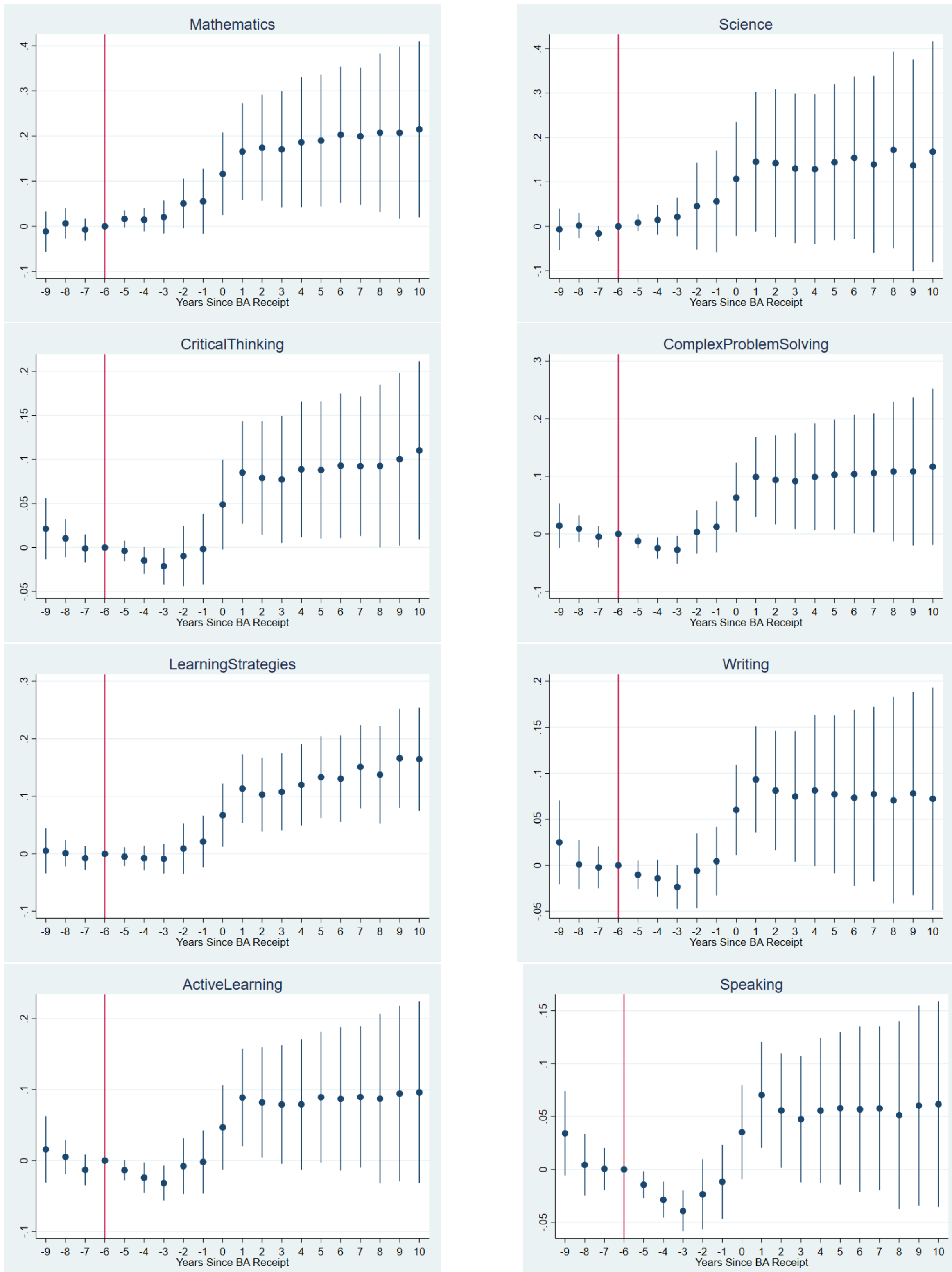
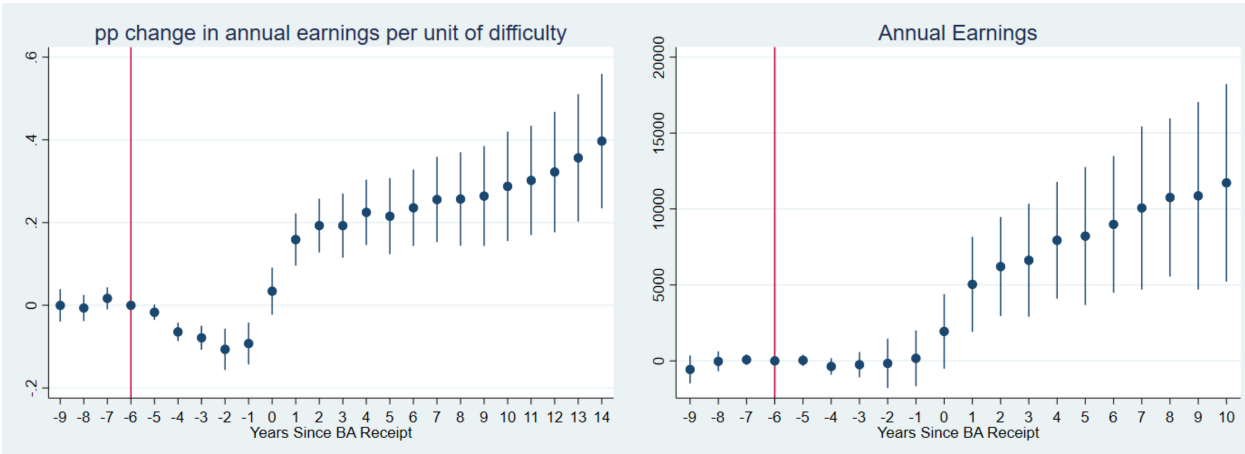
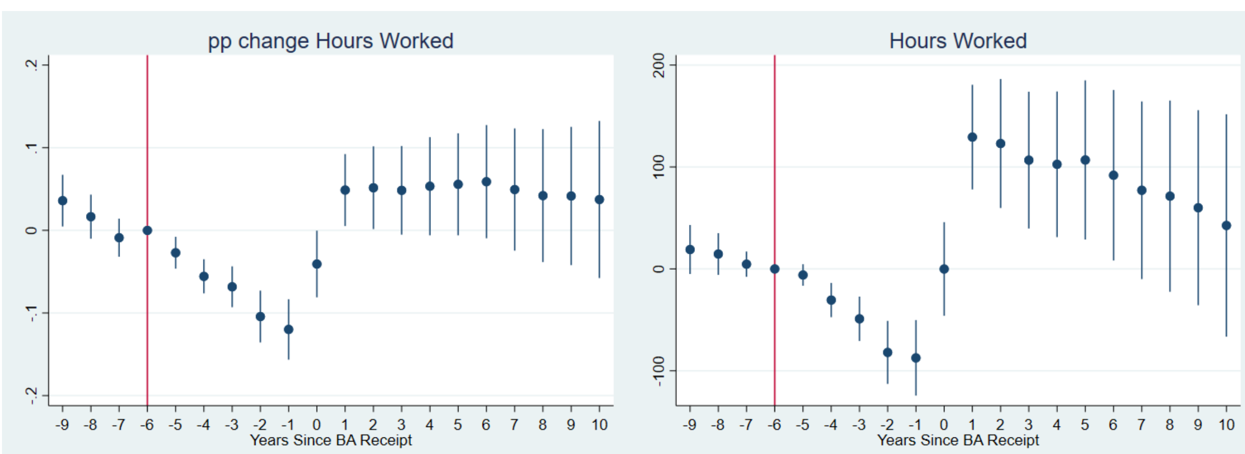


Figure 5: Event Study Specification as in Equation 3 of NLSY97 respondents, but with Skill as dependent variable. Respondent occupation is matched to O*NET occupational skill level for 2019.



(a) (b)

Figure 6: Event Study Specification as in Equation 3 of NLSY97 respondents.



(a) (b)

Figure 7: Event Study Specification as in Equation 3 of NLSY97 respondents.

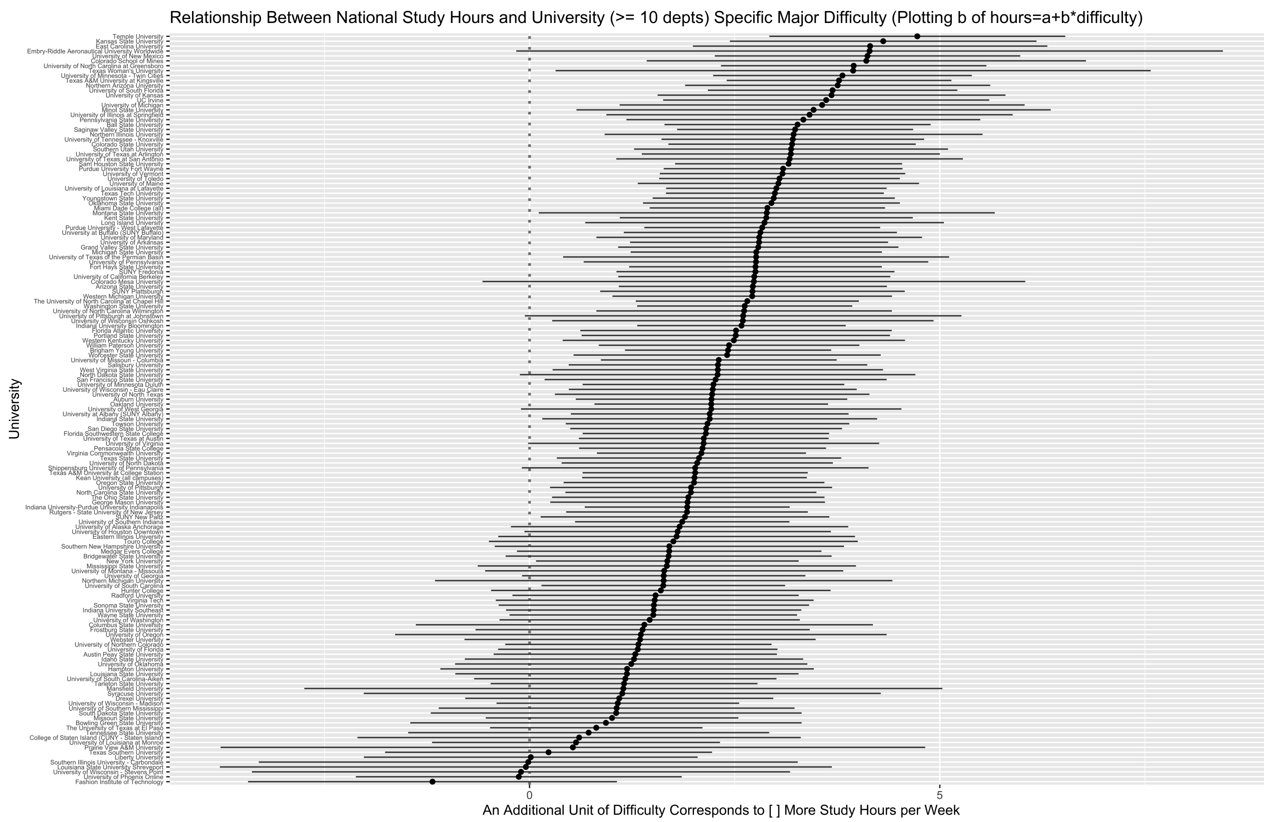


Figure 8: Correlation coefficient between (national) weekly study hours and university specifc major difficulty run separately by university.

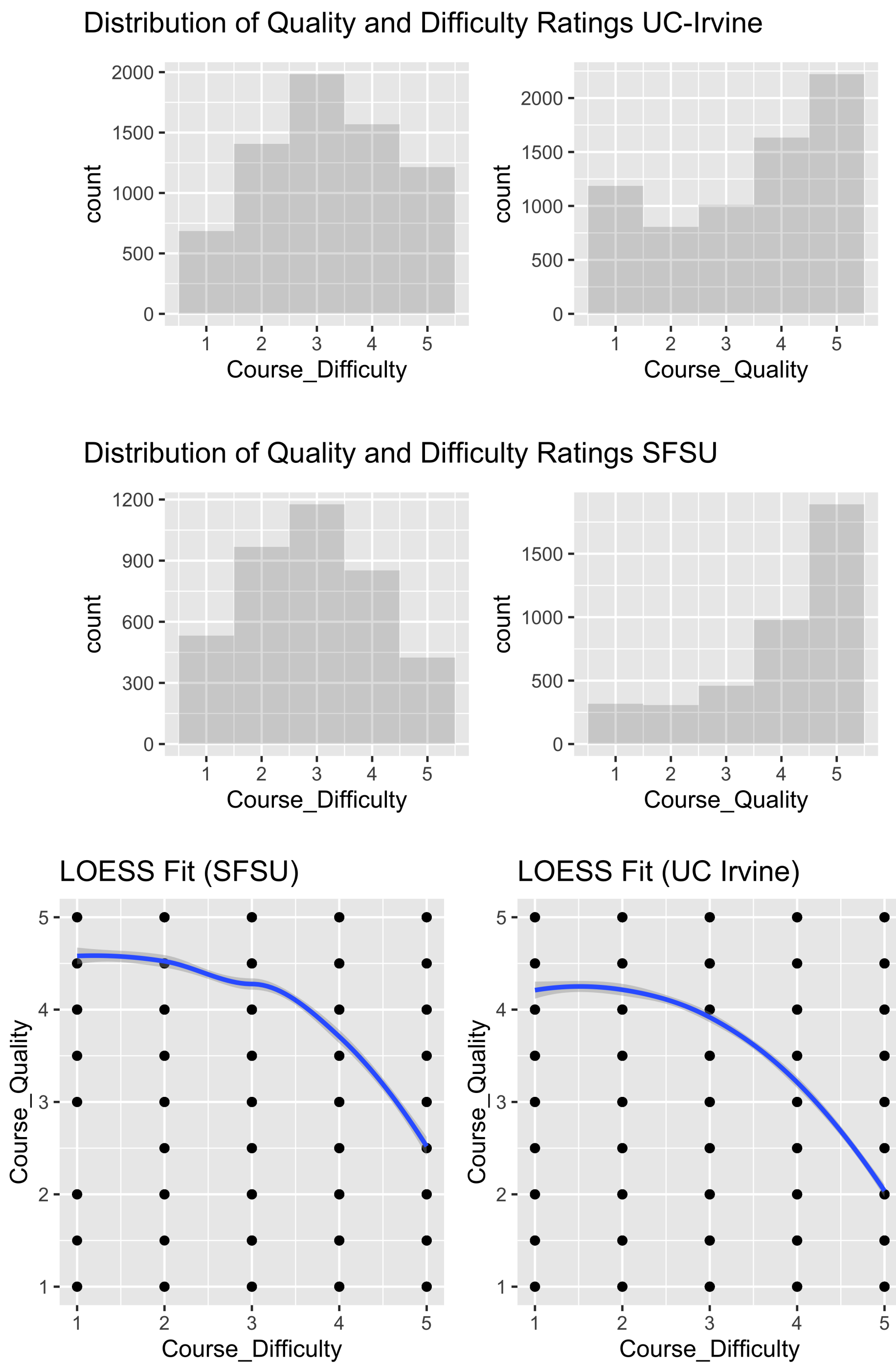


Figure 9: Description of the review level data for selected universities (San Francisco State Universirt and University of California Irvine). The distribution of quality is skewed, while the distribution of difficulty is approximately normal. The lower panel shows a negative relationship between quality and difficulty ratings.

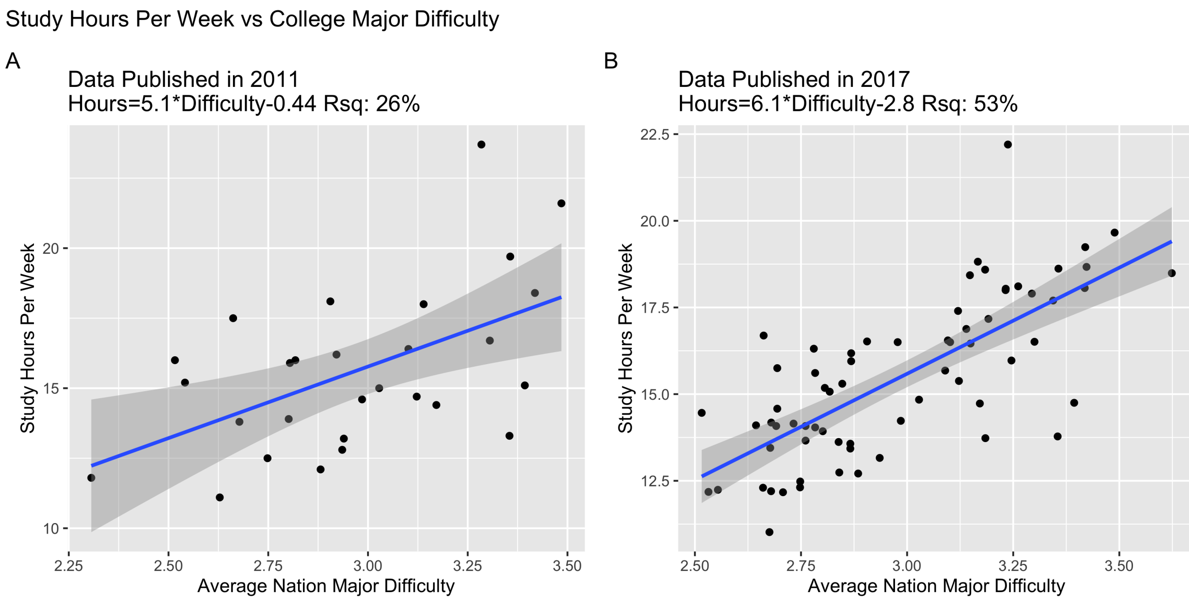


Figure 10: Study Hours Per Week by Major is plotted against the Average national major difficulty. The study hours data is collected by the National Survey for Student Engagement (NSSE) of Indiana University. Average national difficulty data was compiled and calculated by the author. Sources: Panel A: https://www-statista-com.proxygw.wrlc.org/statistics/226433/college-student-study-hours-by-major-2011/; Panel B: https://thetab.com/us/2017/02/06/ranked-majors-work-hardest-59673.

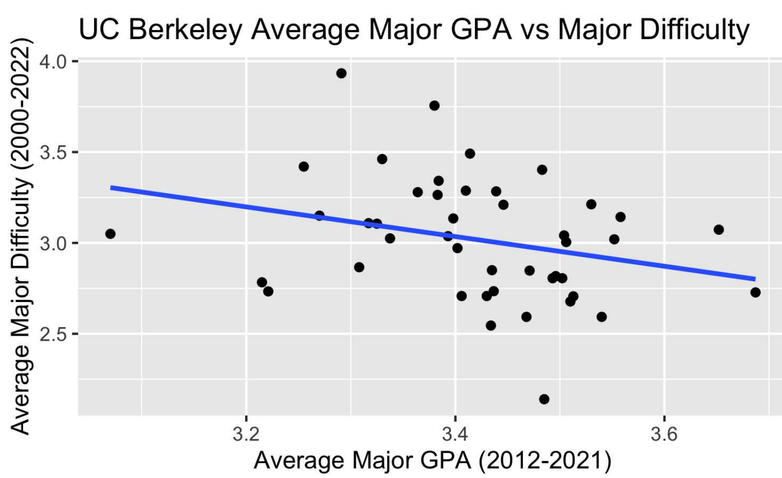


Figure 11: The average GPA by college major from UC Berkeley, for cohorts graduating between 2012 to 2021 is plotted against the average major difficulty for UC Berkeley, for the years 2000-2022, along with the best fit linear model. Source: compiled by the author using collected difficulty data and average GPA data provided by UC Berkeley – see Footnote 5.

References

Thomas Ahn, Peter Arcidiacono, Amy Hopson, and James R Thomas. Equilibrium grade inflation with implications for female interest in stem majors. Technical report, National Bureau of Economic Research, 2019.

Joseph G Altonji, Lisa B Kahn, and Jamin D Speer. Trends in earnings differentials across college majors and the changing task composition of jobs. *American Economic Review*, 104(5):387–93, 2014.

Joseph G Altonji, Peter Arcidiacono, and Arnaud Maurel. The analysis of field choice in college and graduate school: Determinants and wage effects. In *Handbook of the Economics of Education*, volume 5, pages 305–396. Elsevier, 2016.

Richard Arum and Josipa Roksa. Limited learning on college campuses. *Society*, 48(3): 203–207, 2011.

Philip Babcock and Mindy Marks. The falling time cost of college: Evidence from half a century of time use data. *Review of Economics and Statistics*, 93(2):468–478, 2011.

Zachary Bleemer and Aashish Mehta. Will studying economics make you rich? a regression discontinuity analysis of the returns to college major. *American Economic Journal: Applied Economics*, 14(2):1–22, 2022.

Kristin F Butcher, Patrick J McEwan, and Akila Weerapana. The effects of an anti-grade-inflation policy at wellesley college. *Journal of Economic Perspectives*, 28(3):189–204, 2014.

David Card. The causal effect of education on earnings. *Handbook of labor economics*, 3: 1801–1863, 1999.

Theodore Coladarci and Irv Kornfield. Ratemyprofessors com versus formal in-class student evaluations of teaching. *Practical Assessment, Research, and Evaluation*, 12(1):6, 2007.

Clément De Chaisemartin and Xavier D'Haultfoeuille. Difference-in-differences estimators of intertemporal treatment effects. Technical report, National Bureau of Economic Research, 2022.

Jeffrey T Denning, Eric R Eide, Kevin J Mumford, Richard W Patterson, and Merrill Warnick. Why have college completion rates increased? *American Economic Journal: Applied Economics*, 14(3):1–29, 2022.

Andrew Foote. Comparing earnings outcome differences between all graduates and title iv graduates. *Economics of Education Review*, 89:102266, 2022.

Ahu Gemici and Matthew Wiswall. Evolution of gender differences in post-secondary human capital investments: College majors. *International Economic Review*, 55(1):23–56, 2014.

Daniel S Hamermesh and Stephen G Donald. The effect of college curriculum on earnings: An affinity identifier for non-ignorable non-response bias. *Journal of Econometrics*, 144 (2):479–491, 2008.

Mike Insler, Alexander F McQuoid, Ahmed Rahman, and Katherine A Smith. Fear and loathing in the classroom: Why does teacher quality matter? *Available at SSRN 3767273*, 2021.

Valen E Johnson. *Grade inflation: A crisis in college education*. Springer Science & Business Media, 2006.

Lars J Kirkeboen, Edwin Leuven, and Magne Mogstad. Field of study, earnings, and self-selection. *The Quarterly Journal of Economics*, 131(3):1057–1111, 2016.

Sophia Foster Ronald Goeken Jose Pacas Megan Schouweiler Steven Ruggles, Sarah Flood and Matthew Sobek. *IPUMS USA Version*, 11.0, 2021.

Thomas Timmerman. On the validity of ratemyprofessors.com. *Journal of Education for Business*, 84(1):55–61, 2008.

Douglas A Webber. The lifetime earnings premia of different majors: Correcting for selection based on cognitive, noncognitive, and unobserved factors. *Labour economics*, 28:14–23, 2014.