Ex Ante Returns and Occupational Choice*

Peter Arcidiacono† V. Joseph Hotz‡ Arnaud Maurel§
Teresa Romano¶

First version: October 2014
This version: December 2016

Abstract

We show that data on subjective expectations, especially on outcomes from counterfactual choices and choice probabilities, are a powerful tool in recovering *ex ante* treatment effects as well the relationship between individual choices and expected gains to treatment. In this paper we focus on the choice of occupation, and use elicited beliefs from a sample of male undergraduates at Duke University. By asking individuals about potential earnings associated with counterfactual choices of college majors and occupations, we can recover the distribution of the *ex ante* returns to particular occupations, and how these returns vary across majors. We also examine how students update their beliefs over the course of college. We find large differences in expected earnings across occupations, and substantial heterogeneity across individuals in the corresponding *ex ante* returns. Our results also point to the existence of sizable complementarities between college major and occupations. Finally, we find clear evidence of sorting on expected gains, with the *ex ante* returns measured while the individuals were still in college being very informative about their actual occupational choices as well as about their future earnings.

*We are grateful to the editor and four anonymous referees for helpful comments. We also thank seminar participants at Arizona State University, Berkeley, Carnegie Mellon, CREST, London School of Economics, New York Fed, Stanford University, University College London, University of Chicago, University of Oslo, University of Tennessee, Uppsala University and Washington University, as well as participants at the 2015 AEA Meetings (Boston, MA), the NBER Labor Studies 2014 Spring Meeting, the Workshop on Subjective Expectations and Probabilities in Economics and Psychology (Essex, March 2014), the 2014 European Meeting of the Econometric Society (Toulouse), and the 2014 Empirical Micro Jamboree (Duke) for helpful comments and suggestions. Luis Candelaria Barrera, Ashley DeVore and Kate Maxwell provided excellent research assistance. This paper previously circulated under the title "Recovering Ex Ante Returns and Preferences for Occupations using Subjective Expectations Data."

†Duke University, NBER and IZA.
‡Duke University, NBER and IZA.
§Duke University, NBER and IZA.
¶Goucher College.

1 Introduction

Subjective expectations data are increasingly being used in economic research. While early work focused on the accuracy of individual's forecasts over objective events (Manski, 1993, 2004; Hurd and McGarry, 1995, 2002; Dominitz and Manski, 1996, 1997),[1] more recent articles have used elicited probabilities of taking particular courses of actions in the future, along with expectations about potential future outcomes corresponding to counterfactual choices (or treatments), to analyze how individuals are making their decisions under uncertainty (see, e.g., Arcidiacono et al., 2012; Zafar, 2013; Stinebrickner and Stinebrickner, 2014; Delavande and Zafar, 2014; Wiswall and Zafar, 2015, 2016a, 2016b).[2]

In this paper, we show that capturing future choice probabilities as well as expectations both on and off the individual's choice path can be a powerful tool in recovering treatments effects as well the relationship between individual choices and expected gains to treatment. While the proposed approach can be applied to a broad class of potential outcomes models, we consider the choice of occupations for different college majors and document the extent of sorting on *ex ante* monetary returns in this context. As recently emphasized in a series of papers on schooling decisions in the presence of heterogeneity and uncertainty (see, e.g., Carneiro et al., 2003; Cunha et al., 2005; Cunha and Heckman, 2007; and Cunha and Heckman, 2008), agents' decisions are based on *ex ante* monetary returns, as opposed to *ex post* ones. Complementing this literature that uses observational data, we use data that directly elicits agents' *ex ante* returns, thus allowing us to remain agnostic about how agents form their information sets.[3]

There is substantial heterogeneity in earnings across majors and occupations. For instance, data from the American Community Survey (2009-2010) reveal that those who majored in engineering earn as much as 77% more than those who majored in the humanities. To the extent that a sizable fraction of college graduates work in an occupation which does not match their major, those earnings differentials across majors mask the existence of sub-

[1]See Manski (2004) and Hurd (2009) for surveys of measuring and using subjective expectations in economics.

[2]Several important studies have also incorporated subjective expectations about objective events in the estimation of structural dynamic models (Delavande, 2008; van der Klaauw and Wolpin, 2008; van der Klaauw, 2012). Using agents' subjective expectations typically requires milder assumptions about how individuals form their beliefs about future outcomes than usually needed to estimate such forward-looking models. See also Pantano and Zheng (2013) who show how subjective expectations data about agents' future choices can be used to recover unobserved heterogeneity in dynamic structural models.

[3]Most of our analysis focuses on sorting across occupations based on expected, as opposed to *ex post*, gains. As such, our paper complements the literature using observational data to show that individuals sort on *ex post* gains. Notable recent examples in the schooling context include Heckman et al. (2016) and Kirkeboen et al. (2016).

stantial within-major dispersion.[4] For instance, Kinsler and Pavan (2015) estimate that there is a 30% premium for STEM college graduates who work in an occupation related to their major. While these earnings differentials are based on individuals who chose particular majors and occupations and, as such, are not causal, they clearly suggest that occupational choice is a key economic decision, even after conditioning on college major.

In this paper, we use beliefs that were elicited from male undergraduates at Duke University between February and April 2009 to recover the distribution of *ex ante* monetary returns to different occupations, and to quantify the importance of sorting across occupations on *ex ante* monetary returns. This unique dataset contains student expectations regarding the probability of working in different occupations as well as their expected income in each of the occupations where the period of reference is ten years after they graduate.[5] Importantly, these occupation probabilities and expected incomes were asked not only for the major the individual chose but also for counterfactual majors, making it possible to identify how the returns to different occupations vary across majors and to examine the importance of complementarities between majors and occupations.

The data allow us to identify both the *ex ante* treatment effects of particular occupations (relative to a reference occupation) on earnings, for any given college major, as well as the *ex ante* treatment effects of particular majors on the probabilities of working in any given occupation. In order to quantify the importance of (expected) sorting across occupations on expected gains, we also define and compute two types of weighted averages of the individual *ex ante* treatment effects. Taking the major as given, we compute the weighted average of the *ex ante* treatment effects for a given occupation k (relative to the baseline) using as weights the probabilities the individuals report they will work in occupation k (over the sum of declared probabilities of working in that occupation). This weighted average *ex ante* treatment effect, which coincides with the average *ex ante* treatment effect on the treated if individuals form rational expectations over their future occupational choices and in the absence of unanticipated aggregate shocks, will be larger than the average *ex ante* treatment effect of occupation k if individuals expect to sort positively across occupations based on expected gains. We proceed similarly for the *ex ante* treatment effect on the untreated, weighting instead by the declared probability that the individual will not work in occupation k.

[4]See Altonji et al. (2012) and Altonji et al. (2016) for recent reviews of the literature on college major and occupational choices.

[5]This dataset was previously used to examine the determinants of college major choice by Arcidiacono et al. (2012). Their paper treated occupations as lotteries, where the lotteries were affected by the choice of major. In this paper, we follow a more conventional route and treat occupations as choices, consistent with, e.g., Miller (1984), Siow (1984), Keane and Wolpin (1997), Antonovics and Golan (2012), van der Klaauw (2012) and Wiswall and Zafar (2016b).

Importantly, our data allows us to go beyond these average effects and investigate the heterogeneity across individuals by estimating the full distributions of the *ex ante* treatment effects of working in any given occupation k relative to education. We further estimate weighted distributions of *ex ante* treatment effects, using as weights the occupational choice probabilities. Comparing the weighted distributions of *ex ante* treatment effects with the unweighted ones allows us to investigate how sorting on expected gains varies throughout the distribution of *ex ante* returns.

The results reveal substantial differences in expected earnings across occupations. Treating the education occupation as the baseline, the average *ex ante* return range from 30% higher earnings (science) to as much as 122% higher earnings (business) ten years after graduation. The *ex ante* returns are higher for the treated than for the untreated, suggesting positive selection into occupations based on the monetary returns. We also document the existence of a large degree of heterogeneity in the *ex ante* returns for each occupation across college majors, consistent with the accumulation of occupation-specific human capital within each major. For example, natural sciences majors anticipate a premium for a health career (relative to education) that is more than five times larger than the premium that public policy majors anticipate for the same occupation.

We then examine how beliefs vary over time by comparing the distributions of expected incomes between lower-classmen and upper-classmen. Consistent with students learning about the average incomes associated with each occupation as they progress through college, income beliefs about the average Duke student are generally more heterogeneous across lower-classmen than upper-classmen. We further show how changes in beliefs about the average student can be combined with changes in beliefs about own incomes to investigate whether students are learning about their own ability over the course of college. Our results indicate that upper-classmen do appear to have a tighter prior than lower-classmen about their abilities in the vast majority of occupation-major pairs, consistent with some of the uncertainty being resolved over the course of college.

Next, we investigate the relationship between the *ex ante* returns computed using our subjective expectations data and the actual choices of occupations as well as realized earnings. Specifically, using data from the social network LinkedIn completed with the Duke Alumni Database, we were able to collect for the vast majority of the individuals in our original sample the occupations they were working in as of July 2015. In order to collect additional data on ex-post outcomes for the individuals in our sample, we administered a follow-up online survey between February and April of 2016. The respondents were contacted via email, LinkedIn message and/or text message. 117 individuals (about 68% of the initial sample) replied to the follow-up survey. This survey collected information on their past and current occupations, as well as their current earnings.

Using these two additional sources of data, we find a positive and statistically, as well as economically significant association between the *ex ante* returns and the actual choice of occupation. This provides evidence that the beliefs that college students hold about future (choice-specific) labor market outcomes are predictive of their actual behavior in the labor market. We also find that beliefs about earnings are predictive of actual future earnings seven years later. The elasticity of realized earnings with respect to expected earnings remains non-negligible even after controlling for major and occupation, suggesting that earnings beliefs at the time of college matter above and beyond their effect on the choice of occupation. Finally, while expected earnings are important predictors of the choice of occupation, we show that non-monetary components also play a significant role. In particular, we find that a large fraction of individuals expect to give up a sizable share of their maximum potential earnings when choosing their occupation.

Overall, these results highlight the value of eliciting beliefs about potential outcomes, and using those beliefs to estimate the *ex ante* treatment effects. Interestingly, we also show that the occupational choice probabilities that were elicited while the individuals were in college are in fact very informative about actual sorting across occupations on expected gains. This indicates that, at least in this context, *ex ante* data on choice probabilities as well as potential outcomes are a valuable source of information about actual sorting behavior.

The rest of the paper proceeds as follows. In Section 2, we discuss the initial survey and the two follow-up data sources used in the paper. Section 3 shows how to obtain *ex ante* treatment effects given the data, and then discuss the estimated treatment effects. We then compare the distributions of expected incomes for lower-classmen with those of upper-classmen, and outline a set of assumptions under which this comparison can be used to infer how students update their beliefs in Section 4. In Section 5, we use data on the actual occupational choices made by the individuals from our sample and realized earnings, and examine the extent to which the *ex ante* returns that were measured while the individuals were still in college are predictive of their actual occupational choices and earnings. In Section 6 we then use our subjective expectations data to investigate the role played by non-monetary factors in the choice of occupation. Finally, we offer some concluding comments in Section 7. Additional details on the data and supplementary estimation results are collected in the Appendix.

2 Data

2.1 Phase 1 data

Our primary data source is the Duke College Major and Expectations Survey (DuCMES) that was collected on a sample of male undergraduate students at Duke University between February and April 2009.[6] Gender was the only restriction on sample recruitment; male students from any major or year in school were eligible to participate in the survey. Sample members were recruited by posting flyers around the Duke campus. Surveys were administered on computers in a designated room in Duke's Student Union. All 173 students who completed the survey were paid $20.

The DuCMES collected information on students' background characteristics and their current or intended major. Due to the large number of majors offered at Duke University, we divided majors into six broad groups: natural science, humanities, engineering, social sciences, economics, and public policy.[7] Table 1 presents a descriptive overview of our sample. The composition of our sample corresponds fairly closely to the Duke male undergraduate student body. The sample includes slightly more Asians and fewer Hispanics and Blacks than in the Duke male student body, and it over-represents students in natural science majors while under-representing students in public policy. It also appears that a higher percentage of the sample receives some financial aid than is the case in the Duke student body, although the 22.0% figure for the student body is based on aid provided by Duke, whereas the higher percentage of students receiving financial aid (40.5%) is likely due in part to the fact that our survey asked about receipt of financial aid, regardless of source. Finally, we note that the sample is slightly tilted towards upper-classmen.

2.2 Expected choice probabilities and earnings

The DuCMES elicited from the students their expectations about their likelihood of choosing future careers, and how much they expected to earn in them. Namely, for each of the six majors groups displayed in the Table 1, we asked students the probability that they would enter a particular career and the earnings they would expect to receive in that

[6] Arcidiacono et al. (2012) also use the DuCMES data employed in this paper. We refer the reader to that paper for a more comprehensive overview of the data.

[7] In most of the paper we refer for simplicity to the current or intended major as the chosen major. The mapping of students' actual college majors into the major groups displayed in Table 1 is reported in the Appendix. A copy of the questionnaire used in the survey can be found at `http://www.econ.duke.edu/~vjh3/working_papers/college_major_questionnaire.pdf`www.econ.duke.edu/∼vjh3/working_papers/college_major_questionnaire.pdf and is discussed further in Kang (2009).

Table 1: Sample Descriptive Statistics

	Sample	Duke Male Student Body
Current/Intended Major:		
Sciences	17.9%	14.8%
Humanities	9.3%	9.4%
Engineering	19.1%	20.7%
Social Sciences	17.9%	18.8%
Economics	19.7%	18.0%
Public Policy	16.2%	18.0%
Class/Year at Duke:		
Freshman	20.8%	
Sophomore	20.2%	
Junior	27.2%	
Senior	31.8%	
Characteristics of Students:		
White	66.5%	66.0%
Asian	20.2%	16.6%
Hispanic	4.6%	8.3%
Black	4.0%	5.9%
Other	4.6%	3.0%
U.S. Citizen	94.8%	94.1%
Receives Financial Aid	40.5%	22.0%
Sample Size	173	

Data Sources: DuCMES for the Sample characteristics and Campus Life and Learning (CLL) Project at Duke University for Duke Male Student Body. See Arcidiacono and Miller (2011) for a detailed description of the CLL dataset. Current/Intended Major: Respondents were asked to choose one of the six choices (sciences, humanities, engineering, social science, economics, public policy) in response to the questions What is your current field of study? If you have not declared your major, what is your intended field of study?'.

career 10 years after graduation. We used the following six broad sectors to characterize possible careers: Science/Technology, Health, Business, Government/Non-Profit, Education and Law.[8] It is important to note that, for all students in the sample, these probabilities and expected earnings were elicited for all possible occupation-major combinations, i.e. both for the chosen (or intended) majors and the counterfactual majors.

Specifically, to elicit career probabilities, students were asked:

> "Suppose you majored in each of the following academic fields [Sciences, Humanities, Engineering, Social Sciences, Economics, Public Policy]. What are the probabilities that you will pursue the following career field [Science, Health, Business, Government/Non-Profit, Education, Law] AFTER majoring in this academic field?".

To elicit expected earnings associated with different careers and majors, students were asked:

> "For the following questions regarding future income, please answer them in pre-tax, per-year, US dollar term, ignoring the inflation effect. Suppose you majored in the following academic field. How much do you think you will make working in the following career 10 years after graduation?".

Finally, we also asked each student to provide us with their assessments of what the "average" Duke [male] undergraduate would earn in different major-career combinations 10 years after graduation.[9]

Table 2 reports the mean expected incomes for the various major-occupation combinations.[10] Note that each cell contains averages of the responses by each of the 173 students. Expected incomes exhibit sizable variation both across majors and occupations. For instance, majoring in the natural sciences or engineering is perceived to lead to higher earnings in science and health careers, while expected earnings in business are on average higher for economics majors. Differences across occupations are even starker. In particular, average

[8] In most of the paper, we simply refer to these six career groups as occupations.

[9] Students were asked the following question: "Suppose an average Duke student majored in [Sciences, Humanities, Engineering, Social Sciences, Economics, Public Policy]. How much do you think he will make working in the following careers [Science, Health, Business, Government, Education, Law] 10 years after graduation?"

[10] In our sample, only 1.6% of the expected earnings are missing. For these cases, expected earnings, for each major and occupation, are set equal to the predicted earnings computed from a linear regression of log-earnings on major and occupation indicators, interaction between major and occupation, individual-specific average log-earnings across all occupations and majors and an indicator for whether the subjective probability of working in this occupation is equal to zero. One individual in our sample declared that he expected to earn $1,000 for some occupation-major combinations. We assume that this individual declared monthly rather than yearly incomes, and rescale his expected income accordingly.

Table 2: Mean expected incomes for different major/occupation combinations (Annual Incomes, in dollars)

	Occupation:					
Major:	Science	Health	Business	Government	Education	Law
Natural Sciences	109,335	162,636	139,527	95,628	73,597	145,846
Humanities	82,897	126,891	131,254	92,024	71,925	149,058
Engineering	119,601	153,935	154,274	98,738	76,229	167,650
Social Sciences	86,686	126,614	145,856	96,632	71,996	151,323
Economics	96,004	131,822	198,665	103,085	79,303	160,526
Public Policy	90,319	126,521	157,341	110,517	72,928	166,211

Note: Major can either be the chosen major or a counterfactual major so each cell contains the average of 173 observations.

expected incomes are lowest for a career in education and generally highest for a career in law, with the exception of natural sciences and economics majors for which expected incomes are highest for health and business occupations, respectively.

Turning to the choice of occupation, Table 3 presents the average subjective probabilities of working in each occupation that were elicited from students who were asked to condition on having majored in each of the various subject areas. The subjective probabilities of entering each occupation vary substantially across majors. It is worth noting that none of the majors are concentrated into one, or even two, occupations. For any given major, the average subjective probabilities are larger than 10% for at least three occupations. Even for majors which appear to be more tied to a specific occupation, such as business for economics majors, the corresponding subjective probabilities exhibit a fairly large dispersion across individuals (see Figure 1). Overall, the likelihood of working in the various occupations appear to be selectively different across individuals, even after conditioning on a college major.[11]

Finally, Table 4 reports the prevalence of zero probability reported by students, for each major-occupation combination.[12] While some combinations display a large share of zero subjective probabilities, the shares stay well away from one, suggesting that particular majors do not rule out certain occupations for all individuals.

[11]Results for other combinations of occupations and majors are not reported here to save space, but are available from the authors upon request.

[12]The survey design was such that the default values of the subjective probabilities were set equal to zero for all occupation-major combinations. As a result, it might be that some of the zero probabilities observed in the data reflect missing probabilities rather than true zeros. However, in the former case, it seems likely that the latent (unobserved) probabilities are typically close to zero, so that aggregating these two types of zero probabilities should not be too much of a concern.

Table 3: Mean elicited probabilities of choosing alternative occupations, conditional on majoring in alternative fields

	Occupation:					
Major:	Science	Health	Business	Government	Education	Law
Natural Sciences	0.352	0.319	0.120	0.070	0.068	0.070
Humanities	0.067	0.122	0.235	0.145	0.230	0.200
Engineering	0.411	0.194	0.190	0.072	0.065	0.068
Social Sciences	0.091	0.139	0.246	0.193	0.128	0.204
Economics	0.067	0.076	0.515	0.154	0.062	0.125
Public Policy	0.054	0.113	0.228	0.317	0.075	0.214

Note: Major can either be the chosen major or a counterfactual major so each cell contains the average of 173 observations.

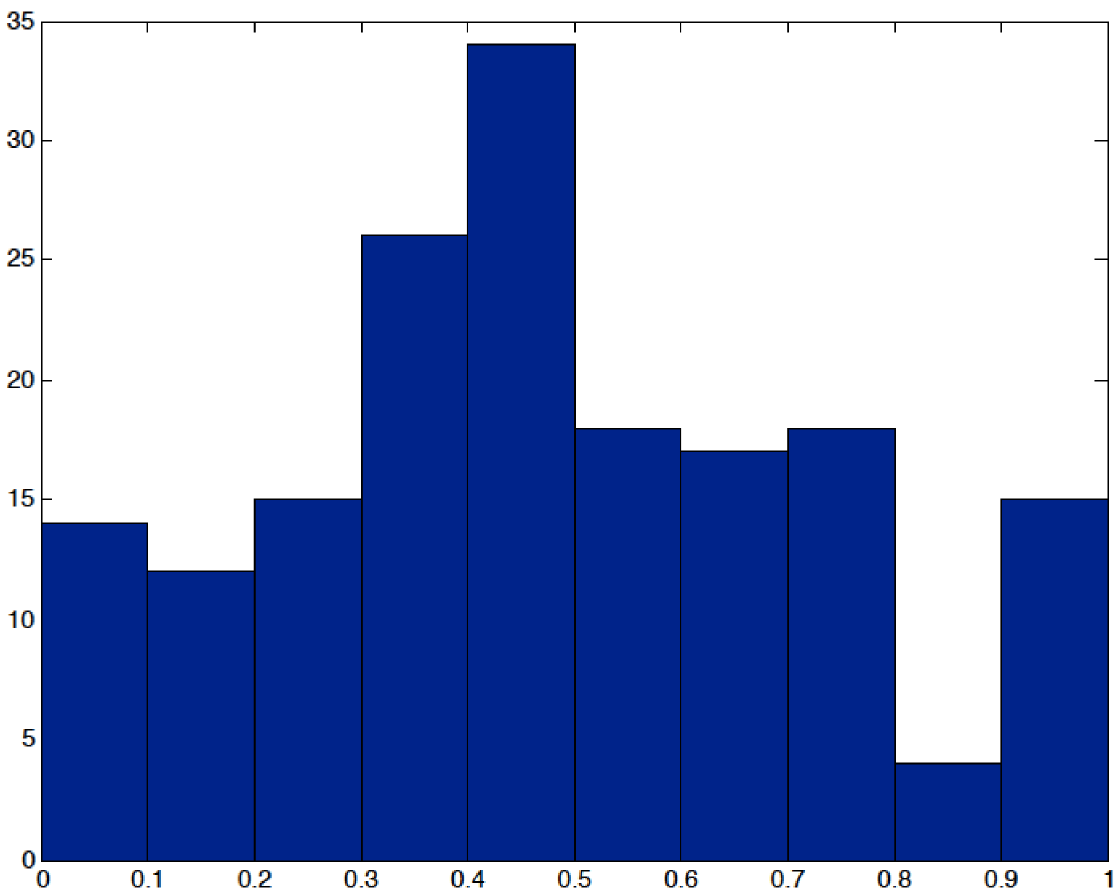

Figure 1: Frequency distribution of subjective probabilities (economics major, business occupation)

Table 4: Incidence of elicited zero probabilities of choosing occupations, conditional on majoring in alternative fields

	Occupation:					
Major:	Science	Health	Business	Government	Education	Law
Natural Sciences	4.62%	9.25%	30.06%	37.57%	41.04%	44.51%
Humanities	50.29%	35.84%	15.61%	20.81%	19.08%	17.92%
Engineering	8.09%	24.28%	22.54%	46.82%	48.55%	51.45%
Social Sciences	46.82%	32.95%	12.14%	15.03%	27.17%	18.50%
Economics	53.76%	50.87%	3.47%	19.65%	46.82%	30.64%
Public Policy	56.65%	38.15%	15.03%	5.78%	40.46%	12.72%

Note: Major can either be the chosen major or a counterfactual major so each cell contains the average of 173 observations.

2.3 Phase 2 and phase 3 data

Are beliefs about future labor market outcomes predictive of the actual choices made by the individuals after graduating from college? In this paper we address this important question by collecting additional data on the actual occupational choices made by the individuals from our sample.[13] For the vast majority of the individuals in our sample we were able to collect those occupations, as of July 2015, using data from the social network LinkedIn. In order to construct a match between our survey data with LinkedIn data, we utilized data from the Duke Alumni database. The Duke Alumni Database is maintained by the Duke Alumni Association and contains graduation year and major information for all Duke graduates. Duke alumni can also update their profile in the database to include past and current job titles and companies, graduate degrees, as well as demographic and contact information. Using information on individual's name, major, and graduation year from the Duke Alumni Database we are able to find the occupations of 143 out of the 173 individuals from our original sample on LinkedIn. For another 18 individuals, occupations were obtained from an internet search where we matched on at least two pieces of information from our initial survey and/or the Duke Alumni Database to ensure an accurate match. Finally, occupations were gathered directly from up to date information in the Alumni Directory for 5 more respondents. In the following we refer to this data as Phase 2 data.[14]

The occupation data from these sources was then mapped into each of the six broad occupation classifications: science, health, business, government, education and law. For example, engineers and software developers were mapped into science careers; doctors, resi-

[13] While occupational choices were collected in the follow-up survey, we were able to obtain this information from publicly available sources for a larger percentage of our sample.

[14] Profiles were updated in the last 48 months and in 3 of the cases they are currently enrolled in graduate school at Duke.

dents and medical students into health; teachers, instructors, and school administrators into education; law clerks and lawyers into law; and lieutenants and policy analysts at government organizations into government. The business classification contained the largest variety of reported occupations including associate, account executive, analyst, manager, and CEO. In each case, both the current job title as well as the employer were considered in constructing the mapping from reported occupation to the six broad occupational classifications.

Finally, in order to collect additional data on ex-post outcomes for the individuals in our sample, we administered a follow-up survey between February and April of 2016 (Phase 3 data). The respondents were contacted via email, LinkedIn message and/or text message. 117 individuals (about 68% of the initial sample of 173 individuals) replied to the follow-up survey. This survey collected information on their past and current occupations as well as their current earnings.[15] Respondents were also asked to update their expectations about what they expect their earnings and occupations to be 10 years after college graduation. All individuals who completed the survey were emailed a coupon code for one 2015 Duke Basketball Championship T- shirt that could be redeemed through the Duke University Bookstore's website.

2.4 Subjective versus actual choice probabilities

We now explore the relationship between our survey data (both Phase 1 and Phase 3) and the choices of occupations (Phase 2). The first two columns of Table 5 show beliefs at Phase 1 as well as the chosen occupation at Phase 2. A much greater share were working in business than what the self-reports would suggest. More individuals are also pursuing a career in health relative to what would be predicted from the subjective probabilities. Correspondingly, smaller shares are seen in several occupations, namely government, law and to a lesser extent in education than the shares in the self-reports.

Although the beliefs are on average off for some of the occupations, the fourth and fifth columns of Table 5 show that the self-reports do have informational content. Column (4) shows the average self-reported probability of working in a career conditional on working in that career. For example, among those who actually chose a science career, the average self-reported probability of choosing science was about 35 percent. Column (5) show the average self-reported probability of working in a career conditional on *not* working in that career. Hence those who did not end up working in science on average thought there is a

[15] 19 individuals declared an occupation that did not match the occupation imputed using the information from LinkedIn, Duke Alumni database and internet search. For those cases we used the declared occupation. Using the follow-up survey we were also able to find the occupation of 2 additional individuals. Overall, we end up with a sample of 167 respondents with non missing actual occupational choices.

14.5 percent chance they would working in the sciences. That the shares are so much higher in column (4) than in column (5)–over twice as high with the exception of education–points to students giving credible answers in the self-reports.

While in principle the ex post and self-reported shares should match under rational expectations, there are at least two reasons why they would not. First, we are not measuring their occupations ten years after graduation, but rather between 3 and 6 years after graduation. We may be picking up students who are working in careers in business now with the intention of transitioning into government or law later. Second, aggregate shocks to the labor market could also lead to differences between ex post shares and their corresponding self reports. For example, LSAT takers peaked in 2009-10 but fell by 45% between then and 2013-14 (Benjamin H. Barton, The Glass Half Full: The Decline and Rebirth of the Legal Profession, Oxford University Press, 2015 page 160).[16]

The third column of Table 5 shows Phase 3 beliefs. Here we see that the reason for the discrepancy between Phase 1 beliefs and occupational shares cannot be because individuals have not yet moved into their preferred occupation. The patterns for Phase 3 beliefs are similar to the actual choices in Phase 2. Namely, individuals in Phase 3 on average perceive a much higher probability of working in business and correspondingly lower probabilities of working in law or government.

The last two columns of Table 5 show the expected probability of working in each career conditional on currently working (or not working) in that career. In all cases, the average perceived probability of working in their current careers in three to six years is over fifty percent which is significantly higher than the correspondingly probabilities for Phase 1 (column (4). This suggests much of the uncertainty regarding occupational choices has been resolved. The discrepancy between the conditional means in column (6) and (7) is particularly large for occupations such as Health (89.3% conditional on working in health, versus 1.4% conditional on not working in that occupation) or Law (76.1% versus 1.8%), consistent with the existence of large switching costs for those occupations. Nevertheless, the probabilities are all significantly lower than one, suggesting that, while uncertainty has been reduced, many individuals perceive a significant chance of moving to another career in the near future.

2.4.1 Decision to work in business

While the previous section showed that the subjective probabilities in Phase 1 have informational content, a natural question is whether they have informational content beyond the chosen major. The link between all majors and all occupations can not be examined due to

[16] The decline continued in 2014-15. See Keith Lee "The Continuing Decline in LSAT Takers" Aug. 20th 2015, http://abovethelaw.com/2015/08/the-continuing-declining-in-lsat-takers/

Table 5: Beliefs and Actual Occupations

	Phase 1 Beliefs	Phase 2 Actual	Phase 3 Beliefs	Phase 1 Cond. Phase 2 Occ=k	Phase 1 Cond. Phase 2 Occ$\neq k$	Phase 3 Cond. Phase 2 Occ=k	Phase 3 Cond. Phase 2 Occ$\neq k$
Science	0.177	0.156	0.170	0.350	0.137	0.662	0.082
Health	0.165	0.210	0.226	0.424	0.098	0.893	0.014
Business	0.261	0.437	0.414	0.374	0.186	0.791	0.120
Government	0.143	0.054	0.062	0.301	0.134	0.536	0.039
Education	0.086	0.054	0.051	0.122	0.087	0.690	0.021
Law	0.169	0.090	0.078	0.391	0.148	0.761	0.018

Columns 1-2, and 4-5 based on 167 matched individuals between Phase 1 and Phase 2. Columns 3 and 6-7 based on 112 individuals from the follow-up survey.

the small size of our sample. However, we can examine the decision to work in business. All majors have at least one individual who chose business. Table 6 shows estimates of a linear probability model of choosing business. Column (1) controls for the self-reported probability of choosing business conditional on the student's actual major as well as an intercept. Controlling for this one variable results in an R-squared close to 0.16. The coefficient itself (0.840) is not statistically different from one, which is what would be expected with rational expectations. Column (2) controls instead for major dummies, with economics majors being 48.5% more likely to choose a business occupation. However, the R-squared (0.125) is actually lower here than when only the self-reported probability was used. Column (3) adds the self-reported probability of choosing business to the specification in Column (2). While the coefficient on the self-reported probability declines relative to column (1), the difference is not significant and the coefficient is still large in magnitude. Interestingly, the coefficient on being an economics major falls substantially (from 0.485 to 0.143) and is no longer statistically significant. These results provide additional evidence that the subjective probabilities are quite informative about future career decisions.

It is possible, however, that the results from the full sample are driven by seniors who already have jobs lined up. In Columns (4)-(6) of Table 6 we perform the same analysis as in Columns (1)-(3) but remove seniors from the sample. The same patterns emerge: the self-reported business probability has more explanatory power than major dummies and its inclusions renders the coefficient on being an economics major insignificant. The coefficients associated with the subjective probability of choosing business, while smaller than with the full sample, remain statistically indistinguishable from 1 at any standard level. As with the full sample, the results show that the subjective probabilities are very informative about future career choices.

Table 6: Linear Probability Model of Choosing Business

	Full Sample			Excluding Seniors		
	(1)	(2)	(3)	(4)	(5)	(6)
Subjective probability	0.840		0.733	0.659		0.569
of choosing business	(0.152)		(0.198)	(0.196)		(0.248)
Major:						
Engineering		0.017	-0.079		0.126	0.020
		(0.121)	(0.119)		(0.160)	(0.164)
Humanities		0.305	0.255		0.318	0.281
		(0.158)	(0.153)		(0.183)	(0.180)
Social Science		0.211	0.100		0.299	0.200
		(0.126)	(0.125)		(0.155)	(0.158)
Economics		0.485	0.143		0.423	0.172
		(0.121)	(0.149)		(0.153)	(0.186)
Public Policy		0.252	0.123		0.273	0.148
		(0.120)	(0.121)		(0.147)	(0.154)
R2	0.158	0.125	0.194	0.094	0.082	0.127

Standard errors in parentheses.
Full sample includes 167 individuals, excluding seniors includes 113 individuals.
All specifications include a constant term.
Subjective probability of choosing business is conditional on their chosen major.

3 *Ex ante* treatment effects

In this section we outline how the different types of *ex ante* treatment effects we are interested in can be measured, and show the corresponding effects in our data. We begin by considering standard treatment effect measures such as the average treatment effect, the average treatment on the treated, and the average treatment on the untreated. We then show how to calculate the full distribution of the various treatment effects and report examples from certain occupations. Finally, we consider treatment effects conditional on different choices of major. All of these estimates are obtained using beliefs about earnings that are collected in our initial (Phase 1) survey. We will examine the evolution of individual beliefs between Phase 1 and Phase 3 surveys in the following section.

3.1 Average *ex ante* treatment effects

We define the *ex ante* treatment effects (or *ex ante* returns) of working in particular occupations on earnings relative to pursuing a career in education, which serves as our baseline occupation and is labeled as occupation $k = 1$.[17] For any given individual i, the *ex ante* treatment effect of occupation $k \in \{2, 3, 4, 5, 6\}$, conditional on majoring (or intending to major) in field j, is simply given by $w_{ijk} - w_{ij1}$ where w_{ijk} is individual i's expected earnings in occupation k given major j (measured in 2009 when the individuals were enrolled in college).[18]

These *ex ante* treatment effects are directly observed in our data. The average *ex ante* treatment effect of occupation k, denoted by $ATE(k)$, is then defined by:

$$ATE(k) := E\left(\sum_j d_{1ij}\left[w_{ijk} - w_{ij1}\right]\right) \tag{3.1}$$

where d_{1ij} is an indicator for whether i chose (or intends to choose) major j. This population parameter is estimated using its sample analog:

$$\widehat{ATE(k)} = \frac{\sum_i \sum_j d_{1ij}\left[w_{ijk} - w_{ij1}\right]}{N}, \tag{3.2}$$

where N is the sample size. Note that our measure of *ex ante* returns does not incorporate any (expected) differences in direct and opportunity costs across occupations, which may be significant here since some careers such as Law typically require an advanced degree. As such, it should not be understood as an *ex ante* internal rate of return, but rather as the expected effect of working in a particular occupation (relative to education) on the earnings ten years out.

To quantify the importance of (expected) sorting across occupations on expected gains, we also compute two types of weighted averages of the individual *ex ante* treatment effects. First, we compute the weighted average of the *ex ante* treatment effects for occupation k (relative to the baseline occupation), using as weights the probabilities the individuals report they will work in occupation k 10 years after graduation (over the sum of declared probabilities of working in occupation k). Namely:

$$\widehat{TT(k)} := \frac{\sum_i \sum_j \widehat{\omega_{ijk}^{TT}} d_{1ij}\left[w_{ijk} - w_{ij1}\right]}{N}, \tag{3.3}$$

[17]We choose to use education as a baseline because the earnings in this occupation do not vary much across college majors (see Table 2), thus making it easier to interpret the heterogeneity across majors in the *ex ante* treatment effects.

[18]In this paper we define and estimate the *ex ante* treatment effects of working in particular occupations on future earnings. Recent work by Wiswall and Zafar (2016a) applies a similar methodology to estimate the expected effect of college major choice on future earnings as well as other outcomes, including future labor supply and spousal earnings.

with $\widehat{\omega_{ijk}^{TT}} = \frac{p_{ijk}}{\sum_i \sum_j d_{1ij} p_{ijk}/N}$, where p_{ijk} is the elicited probability from individual i that he would choose occupation k given major j. $\widehat{TT(k)}$ consistently estimates the following population parameter:

$$TT(k) = E\left(\sum_j \omega_{ijk}^{TT} d_{1ij}\left[w_{ijk} - w_{ij1}\right]\right) \tag{3.4}$$

where $\omega_{ijk}^{TT} = \frac{p_{ijk}}{E\left(\sum_j d_{1ij} p_{ijk}\right)}$. $TT(k)$ is a weighted average *ex ante* treatment effect of occupation k, which oversamples the *ex ante* treatment effects for the individuals with a high subjective probability of choosing occupation k. This parameter will be larger than the average *ex ante* treatment effect of occupation k, $ATE(k)$, if individuals expect on average to sort positively across occupations based on expected gains.

If individuals form rational expectations over their future occupational choices, and in the absence of unanticipated aggregate shocks, it is easy to show that the weights can be rewritten as $\omega_{ijk}^{TT} = \frac{d_{2ijk}}{E\left(\sum_j d_{1ij} d_{2ijk}\right)}$ (where d_{2ijk} is an indicator for whether i works in occupation k 10 years after graduating from major j). It follows that $TT(k)$ coincides with the average *ex ante* treatment effect of occupation k on the treated:

$$TT(k) = E\left(\sum_j d_{1ij}\left[w_{ijk} - w_{ij1}\right] \mid \sum_j d_{1ij} d_{2ijk} = 1\right) \tag{3.5}$$

More generally, Equality 3.5 still holds in the presence of unanticipated aggregate shocks (denoted by e^{α_k}) affecting the share of individuals working in occupation k, such that, for any major j, $E\left(d_{2ijk}\right) = e^{\alpha_k} E\left(p_{ijk}\right)$.[19]

Finally, we compute the weighted average of the *ex ante* treatment effects for occupation k (relative to the baseline occupation), using as weights the probabilities of *not* working in occupation k 10 years after graduation (over the sum of probabilities of not working in occupation k):

$$\widehat{TUT(k)} := \frac{\sum_i \sum_j \widehat{\omega_{ijk}^{TUT}} d_{1ij}\left[w_{ijk} - w_{ij1}\right]}{N}, \tag{3.6}$$

with $\widehat{\omega_{ijk}^{TUT}} = \frac{1-p_{ijk}}{\sum_i \sum_j d_{1ij}(1-p_{ijk})/N}$. The previous derivations for the treatment on the treated can be directly transposed to the untreated case, after replacing p_{ijk} by $1 - p_{ijk}$ and d_{2ijk} by $1 - d_{2ijk}$. In particular, $\widehat{TUT(k)}$ consistently estimates the *ex ante* treatment effect of occupation k on the untreated if students form rational expectations over future choices

[19]In Section 5.1 we provide sufficient conditions under which unanticipated shocks on the occupation-specific earnings lead to these types of shocks.

Table 7: *Ex Ante* Treatment Effects of Occupations (Annual Earnings, in dollars)

Occupation	TT	TUT	ATE	ATE: share of Education earnings
Science	29,820	20,674	22,320	30.0%
	(4,786)	(3,246)	(3,121)	
Health	117,700	57,808	68,065	91.6%
	(18,802)	(6,879)	(8,575)	
Business	104,224	84,201	89,533	120.5%
	(14,664)	(8,052)	(8,480)	
Government	26,733	25,753	25,875	34.8%
	(7,162)	(3,918)	(3,970)	
Law	110,423	84,343	88,750	119.4%
	(20,033)	(10,595)	(11,280)	

Note: Standard errors are reported in parentheses.

and in the presence of unanticipated aggregate shocks that affect the occupation shares multiplicatively.

The estimated effects, $\widehat{ATE(k)}$, $\widehat{TT(k)}$ and $\widehat{TUT(k)}$, defined in (3.2), (3.3), and (3.6), respectively, are not based on actual occupational choices, since these students have not yet chosen an occupation. Rather, we use students' elicited probabilities of choosing the various occupations to characterize these choices.

Table 7 presents estimates of the three *ex ante* treatment effects of working in particular occupations on earnings 10 years after graduation which correspond to the estimators defined earlier in (3.2)-(3.4). Relative to education, the average *ex ante* treatment effects range from \$22,320 for science (30.0% of the mean expected earnings in education) to as much as \$89,533 in business (120.5% of the mean expected earnings in education). Health, business and law careers all have very large earnings premia of over 91%, while those working in a science or government occupation expect a much smaller premium of 30.0% to 34.8% ten years after graduation.[20] Consistent with positive sorting on expected gains across occupations, our estimates show that, for each occupation, the untreated anticipate lower premia than the treated. The difference is particularly large for health occupations, where the expected premium is more than two times smaller for the untreated. These sorting effects turn out to be much weaker for science careers, where the untreated anticipate to earn 69% as much as the treated, and are negligibly small for government careers.

[20]Table 19 in the appendix reports the estimated average *ex ante* treatment effects separately for lower-classmen and upper-classmen. While the estimates are larger for all occupations for upper-classmen than for lower-classmen, none of them are significantly different at standard statistical levels.

3.2 Full distribution of treatment effects

Importantly, our data allows us to generate not only average effects, but also estimate the distributions of the *ex ante* treatment effects of working in any given occupation k, relative to the baseline occupation. We first discuss the estimation of the unconditional distribution of the *ex ante* treatment effects, before turning to the distribution of the *ex ante* treatment effects on the treated and untreated subpopulations.[21]

First, the density of the distribution of the *ex ante* treatment effects on the overall population can be simply estimated with a kernel density estimator, using the fact that we have direct measures of the *ex ante* treatment effects for each occupation k, $k = 2, ..., 6$, for each student in our sample. We denote the resulting density by $f_{TE,k}(\cdot)$ and its estimator by $\widehat{f_{TE,k}(\cdot)}$.

Second, we define a weighted version of the estimator $\widehat{f_{TE,k}(\cdot)}$ (denoted by $f_{TE,k}^{Treated}(\cdot)$) that puts more weight on the parts of the *ex ante* treatment effects distribution for which the declared probabilities of choosing occupation k are higher. This extends to the full distribution of *ex ante* treatment effects the weighting scheme that was proposed in the previous section for the average treatment effects. Specifically, for any scalar u:

$$\widehat{f_{TE,k}^{Treated}}(u) \quad := \quad (\widehat{\omega_{ijk}^{TT}} * \widehat{f_{TE,k}})(u) \tag{3.7}$$

where the weights are given by $\widehat{\omega_{ijk}^{TT}}(u) = \frac{\hat{g(u)}}{\sum_i \sum_j d_{1ij} p_{ijk}/N}$, $g(u) = E(\sum_j d_{1ij} p_{ijk} | w_{ijk} - w_{ij1} = u)$ and $\hat{g}(u)$ denotes a consistent estimator of $g(u)$ (e.g. Nadaraya-Watson estimator). $\widehat{f_{TE,k}^{Treated}}(.)$ consistently estimates the following density:

$$f_{TE,k}^{Treated}(u) \quad = \quad (\omega_{ijk}^{TT} * f_{TE,k})(u) \tag{3.8}$$

where $\omega_{ijk}^{TT}(u) = \frac{g(u)}{E(\sum_j d_{1ij} p_{ijk})}$. If individuals form rational expectations over their future occupational choices, and assuming multiplicative aggregate shocks affecting the shares of workers in each occupation, it follows from Bayes' rule that $f_{TE,k}^{Treated}(.)$ coincides with the density of the distribution of the *ex ante* treatment effects on the treated subpopulation. Finally, the distribution of the *ex ante* treatment effects on the untreated can be estimated in a similar fashion by replacing p_{ijk} with $(1 - p_{ijk})$ in (3.7).

Figures 2, 3, and 4 plot the densities of the *ex ante* treatment on the treated and treatment on the untreated for government, health, and business occupations, respectively.[22] Each of

[21] As in the previous section, all of the *ex ante* treatment effects are computed for students' chosen (as opposed to counterfactual) college majors.

[22] All densities were estimated using 100 grid points over the support, and a Gaussian kernel with optimal default bandwidth returned by the procedure `ksdensity` in Matlab.

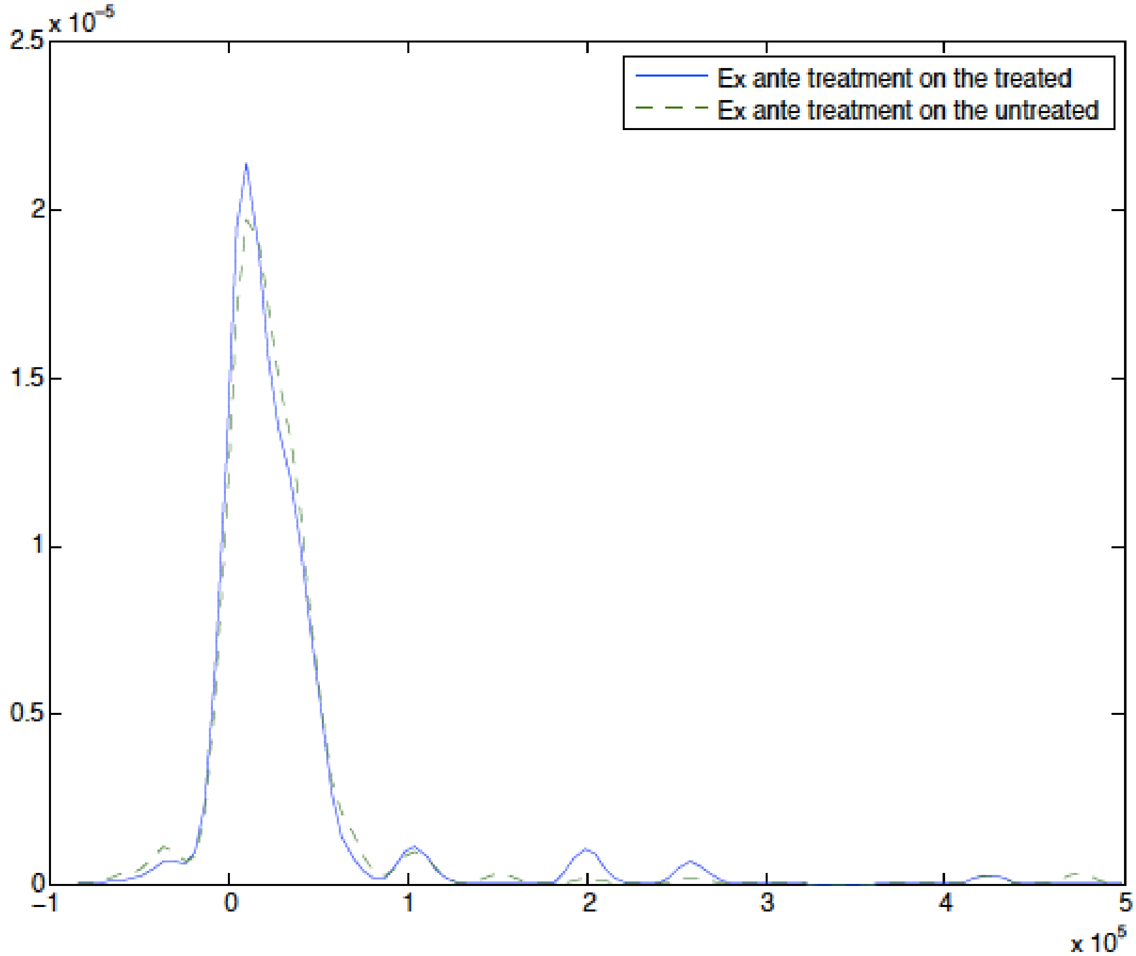

Figure 2: Densities of *Ex Ante* Treatment Effects: Government

the figures shows a different pattern of selection. For government, the distributions for the treated and the untreated are essentially the same: there is little role for selection into government jobs, at least relative to education. For health, the treated distribution is to the right of the untreated distribution, suggesting substantial positive selection. For business careers, while there appears to be significant selection at the bottom end of the distribution, the discrepancy between the two distributions is attenuated in the top end. This latter pattern suggests that there is a significant group of individuals who would do quite well in business—essentially as well as the highest returns individuals from the treated group—but whose preferences, or expected earnings in other occupations, lead them away from business. Overall, these results suggest that there is much more to the distributions of *ex ante* treatment effects than just their means.

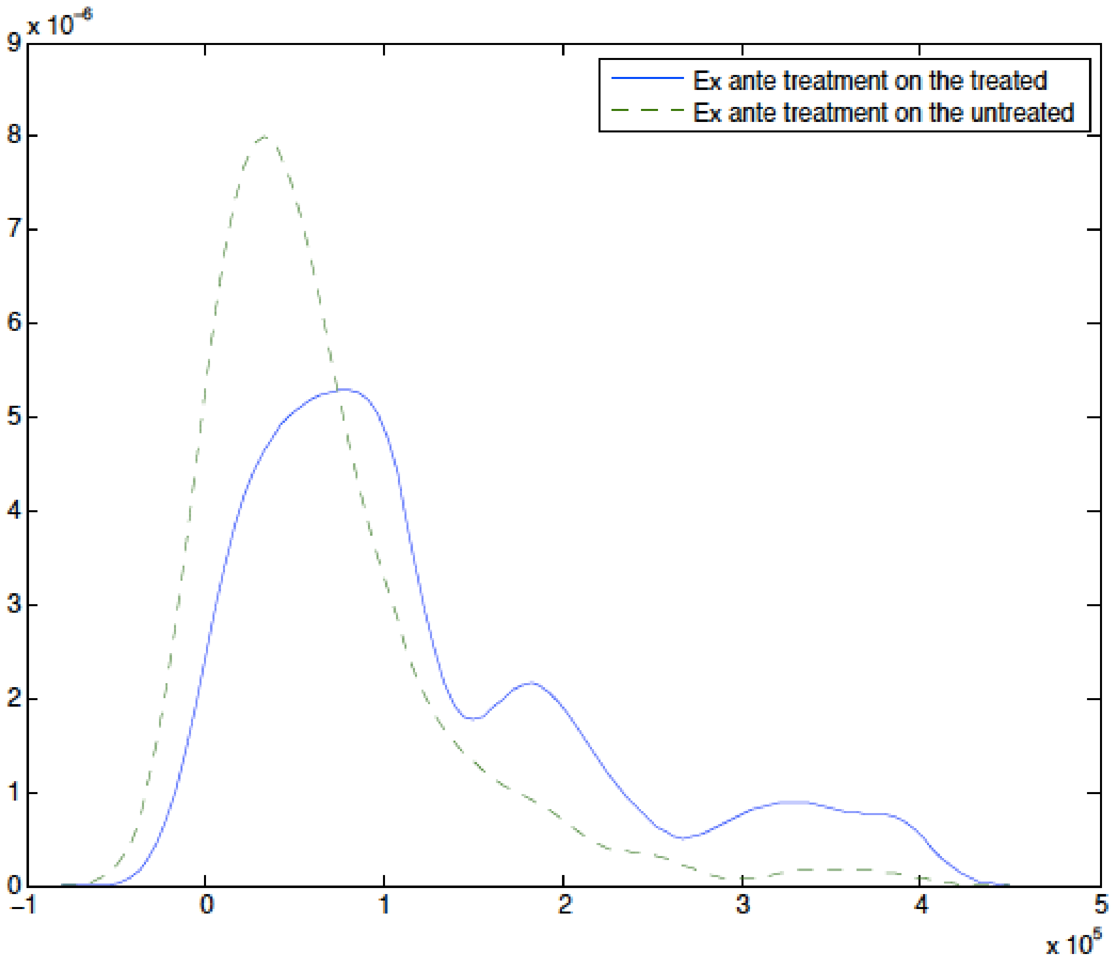

Figure 3: Densities of *Ex Ante* Treatment Effects: Health

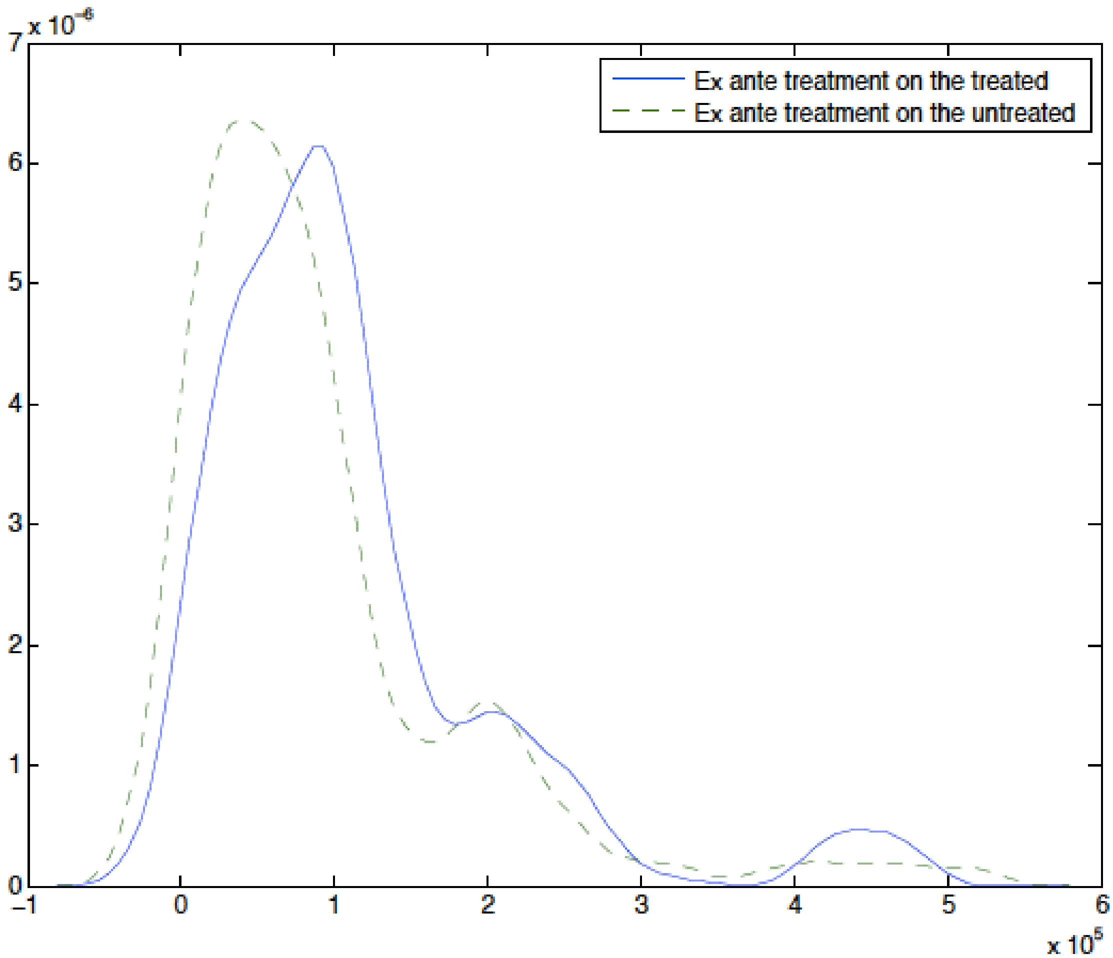

Figure 4: Densities of *Ex Ante* Treatment Effects: Business

3.3 Heterogeneity by major

While $\widehat{TT(k)}$, $\widehat{TUT(k)}$ and $\widehat{ATE(k)}$ are obtained by averaging over different choices of college major, we also can estimate the *ex ante* treatment effects of occupations conditional on each of the particular majors. Namely, we estimate the average *ex ante* treatment effect, *ex ante* treatment on the treated and treatment on the untreated for each chosen major j by:

$$\widehat{ATE(k|j)} := \frac{\sum_i d_{1ij} \left[w_{ijk} - w_{ij1}\right]}{\sum_i d_{1ij}}. \tag{3.9}$$

$$\widehat{TT(k|j)} := \frac{\sum_i d_{1ij} p_{ijk} \left[w_{ijk} - w_{ij1}\right]}{\sum_i d_{1ij} p_{ijk}}, \tag{3.10}$$

$$\widehat{TUT(k|j)} := \frac{\sum_i d_{1ij}(1 - p_{ijk}) \left[w_{ijk} - w_{ij1}\right]}{\sum_i d_{1ij}(1 - p_{ijk})}, \tag{3.11}$$

Given that we also elicit the subjective expectations for counterfactual majors, we can estimate the *ex ante* treatment effects for those who did *not* choose major j by replacing d_{1ij} with $1 - d_{1ij}$.

In Table 8, we present the treatment effect parameters conditional on students' chosen majors. There is a substantial amount of heterogeneity in the expected earnings premium for a given occupation across majors. Notably, natural science majors expect on average a $136,452 premium for a health career relative to education, which is more than six times larger than the $22,146 premium expected by public policy majors who anticipate to enter this type of occupation. Examining some of the other average *ex ante* returns, economics majors have the highest premium for business occupations, while engineering and natural science majors have the highest premia for science careers. Overall, these patterns provide evidence of complementarities between majors and occupations. In particular, the major-occupation pairs that are typically thought of as being closely related to one another – such as economics and business, natural science and health, as well as engineering or natural science and science occupations – do have the highest premia. While these results are consistent with the accumulation of occupation-specific human capital within each major, they are also consistent with a form of selectivity in choice of major, whereby individuals who expect to be more productive in health are more likely to choose a natural science major.

As can be seen in Table 8, *ex ante* treatment effects on the untreated by student's major generally are lower than the treatment effects on the treated, similar to the results obtained without conditioning on the major (Table 7). There are, however, a couple of exceptions. For instance, *ex ante* returns to science careers are higher for the untreated in social science majors, while *ex ante* returns to government careers are higher for the untreated in the humanities and social sciences. The differences between the *ex ante* treatment effects on the treated and the *ex ante* treatment effects on the untreated effects provide, for each major, a measure of the importance of selection on the expected returns to each occupation. For a majority of occupation-major pairs, this difference is positive, consistent with positive sorting on expected earnings in different occupations, but the differences tend to be quantitatively small. Notable exceptions include legal careers for social sciences majors, where selection explains about 45% of the expected premium among the treated, as well as government careers for science majors, where selection accounts for around half of the expected premium.

Finally, Table 9 provides estimates of the three *ex ante* treatment effects by counterfactual major. The treatment effects on the treated are again generally larger than the treatment effects on the untreated. It is worth noting that these *ex ante* treatment effects also exhibit a substantial degree of heterogeneity across majors. Notably, expected premia for business careers are higher for economics majors, while returns to science careers are higher for engineering and natural science majors. The fact that these types of complementarities between majors and occupations still hold when focusing on the majors which were *not* chosen by the individuals points to the accumulation of occupation-specific human capital within majors.[23]

3.4 Average *ex ante* treatment effects conditional on actual treatment status

Using our Phase 2 data on the actual choices of occupations, we can investigate how the *ex ante* treatment effects of working in particular occupations vary with the actual, rather than the expected, treatment status. Table 10 below reports the *ex ante* treatment effects on the treated as well as the *ex ante* treatment effects on the untreated for all occupations relative to Education. Comparing the treatment effects for the treated with those for the untreated point to similar sorting patterns to the ones obtained earlier when using the expected treatment status (Table 7, Section 3.1). Consistent with positive sorting into Science, Health, Business and Government, individuals who end up working in any of these occupations anticipate on average higher earnings premia relative to those who work in another occupation. Among this set of occupations, sorting effects are highest for Health while they are lowest for Government, echoing our earlier findings based on expected rather than actual choices. On the other hand for careers in Law, the *ex ante* treatment effects

[23] See also Kinsler and Pavan (2015) on the importance of major-specific human capital. They find, using data from the Baccalaureate and Beyond Longitudinal Study, that individuals have higher wages when working in an occupation related to one's field of study compared to working in non-related occupations.

Table 8: *Ex Ante* Treatment Effects of Occupations by Chosen Major (Annual Earnings, in dollars)

		Chosen Major:					
Occupation:		Economics	Engineering	Humanities	Natural Sciences	Public Policy	Social Sciences
Science	TT	18,607	38,125	17,354	28,844	25,515	14,631
		(6,746)	(8,109)	(6,601)	(8,166)	(11,238)	(3,074)
	TUT	18,053	27,290	7,069	36,036	15,732	19,604
		(7,101)	(6,694)	(4,806)	(11,761)	(8,109)	(6,295)
	ATE	18,092	31,642	7,620	33,710	15,982	18,968
		(6,801)	(6,867)	(4,736)	(10,070)	(8,010)	(5,599)
Health	TT	89,752	84,002	53,978	182,781	38,354	69,137
		(22,916)	(17,260)	(13,455)	(43,391)	(11,733)	(16,417)
	TUT	60,800	57,061	59,513	106,834	21,218	55,753
		(19,922)	(10,295)	(13,283)	(27,038)	(6,658)	(10,421)
	ATE	63,272	61,945	58,664	136,452	22,146	57,774
		(19,241)	(10,566)	(11,831)	(32,277)	(6,813)	(9,758)
Business	TT	120,434	71,691	66,116	112,066	81,288	124,648
		(33,521)	(13,723)	(22,874)	(24,603)	(24,661)	(37,628)
	TUT	120,451	69,920	56,639	107,139	63,834	84,611
		(32,147)	(12,562)	(19,073)	(27,532)	(14,154)	(16,062)
	ATE	120,441	70,309	57,875	107,581	68,393	93,484
		(30,872)	(12,335)	(18,882)	(26,291)	(15,693)	(19,925)
Government	TT	26,740	11,327	16,249	66,656	31,164	16,751
		(14,765)	(4,149)	(5,213)	(28,998)	(15,088)	(8,268)
	TUT	25,775	12,120	23,877	33,673	22,406	36,306
		(7,338)	(4,978)	(9,566)	(12,139)	(9,798)	(16,070)
	ATE	25,882	12,072	22,813	35,323	24,822	33,645
		(7,841)	(4,856)	(8,693)	(12,894)	(10,970)	(14,261)
Law	TT	91,587	57,724	94,926	116,578	136,915	114,266
		(22,839)	(11,077)	(28,309)	(42,514)	(55,369)	(32,543)
	TUT	93,632	67,060	62,091	88,931	131,354	63,003
		(26,729)	(13,864)	(13,566)	(22,230)	(45,257)	(9,845)
	ATE	93,382	66,296	70,688	90,161	133,214	75,323
		(25,632)	(13,066)	(15,371)	(22,690)	(47,102)	(15,221)

Note: Standard errors are reported in parentheses.

Table 9: *Ex Ante* Treatment Effects of Occupations by Counterfactual Major (Annual Earnings, in dollars)

		Counterfactual Major:					
Occupation:		Economics	Engineering	Humanities	Natural Sciences	Public Policy	Social Sciences
Science	TT	5,570	46,103	19,363	42,674	18,541	12,142
		(5,872)	(4,678)	(5,085)	(7,656)	(4,126)	(3,227)
	TUT	17,162	46,160	10,639	32,557	17,607	13,910
		(8,350)	(6,779)	(3,056)	(4,356)	(3,737)	(3,377)
	ATE	16,361	46,137	11,314	36,182	17,665	13,757
		(7,992)	(4,920)	(3,105)	(4,663)	(3,627)	(3,184)
Health	TT	63,261	108,575	83,483	86,114	73,373	74,115
		(35,669)	(21,031)	(22,479)	(9,723)	(23,333)	(21,741)
	TUT	48,796	74,727	50,606	75,443	57,697	50,634
		(9,097)	(7,746)	(7,884)	(9,175)	(8,140)	(7,915)
	ATE	49,889	81,420	54,589	78,689	59,665	53,929
		(10,325)	(9,773)	(9,014)	(8,771)	(9,392)	(9,156)
Business	TT	141,157	84,753	66,887	62,638	100,135	92,047
		(17,154)	(15,689)	(11,055)	(12,406)	(23,612)	(15,227)
	TUT	97,168	78,751	57,145	55,987	83,906	62,078
		(12,148)	(12,565)	(8,993)	(7,929)	(11,186)	(9,499)
	ATE	119,097	79,868	59,478	56,837	87,506	69,576
		(12,307)	(12,212)	(8,657)	(8,251)	(12,263)	(9,687)
Government	TT	20,154	28,556	24,362	24,886	49,602	33,178
		(9,356)	(8,282)	(10,164)	(8,467)	(18,272)	(11,848)
	TUT	23,885	24,663	19,079	18,656	35,465	19,968
		(7,921)	(4,716)	(4,182)	(3,624)	(6,252)	(3,788)
	ATE	23,268	24,968	19,851	19,130	40,055	22,670
		(7,930)	(4,749)	(4,691)	(3,921)	(7,444)	(5,060)
Law	TT	88,413	99,691	75,877	72,712	78,152	73,929
		(18,743)	(42,003)	(10,838)	(19,074)	(11,089)	(13,926)
	TUT	76,764	97,171	78,252	67,972	87,326	81,725
		(11,221)	(26,185)	(9,327)	(8,658)	(10,778)	(9,949)
	ATE	78,248	97,343	77,791	68,339	85,572	80,201
		(11,015)	(26,988)	(9,160)	(8,910)	(10,467)	(10,042)

Note: Standard errors are reported in parentheses.

Table 10: *Ex Ante* Treatment Effects of Occupations Conditional on the Actual Treatment Status (Annual Earnings, in dollars)

Occupation	TT	TUT
Science	34,808	18,733
	(7,627)	(3,275)
Health	122,000	50,776
	(14,155)	(7,289)
Business	90,726	83,596
	(11,462)	(10,101)
Government	37,111	23,415
	(16,676)	(3,980)
Law	88,667	88,515
	(30,692)	(9,642)

Note: Estimation based on the subsample of 167 matched individuals. Standard errors are reported in parentheses.

on the treated is marginally lower than the *ex ante* treatment effects on the untreated, contrasting with the previous estimates that were obtained using subjective probabilities as opposed to the actual choice of occupation. One possible explanation for this finding is that nonpecuniary components, some of them being negatively correlated with the expected earnings, play an important role in the decision to pursue a career in Law.

Turning to the distributions of treatment effects, Figures 5 and 6 below reports the densities of the *ex ante* treatment effects conditional on the actual treatment status for Health and Business occupations, respectively. We chose to focus on these two occupations since these are the two most popular ones in the sample. Comparing Figure 5 with the conditional distributions based on the expected choices (Figure 3, Section 3) reveals that both sets of distributions are very similar. For this occupation, using the *ex ante* choice probabilities rather than conditioning on the actual choices does not make much of a difference throughout the whole distribution of *ex ante* treatment effects. The distributions for Business are not as similar whether we use the expected or actual choices, but the overall patterns in terms of selection into treatment are qualitatively similar.

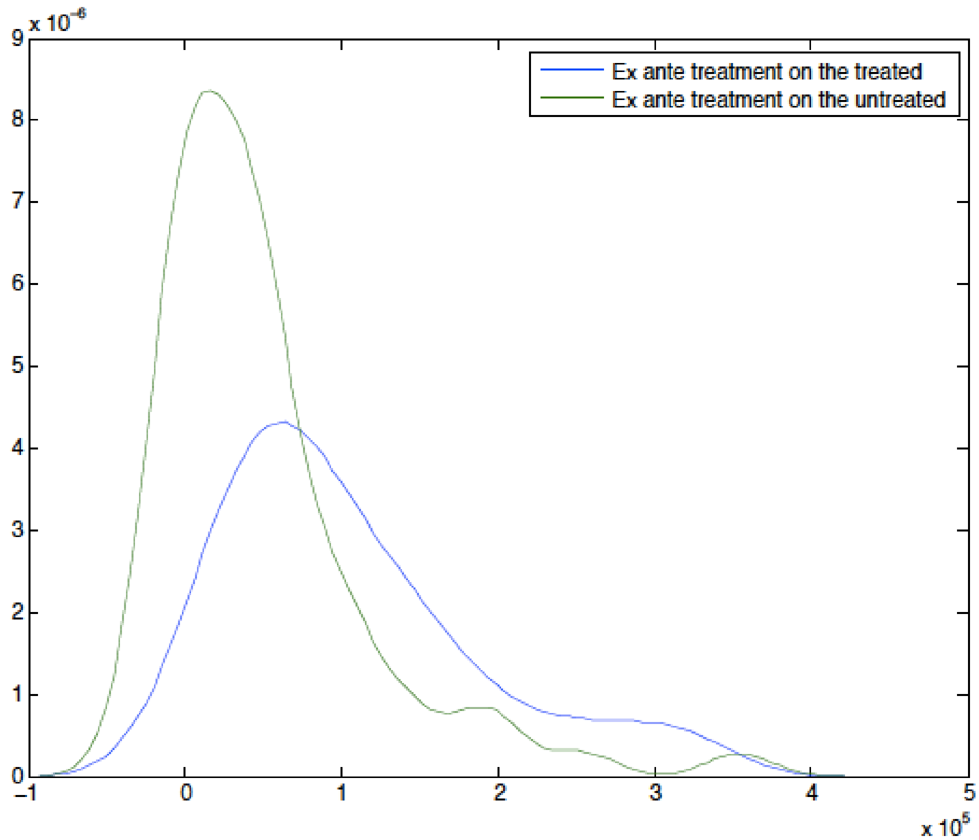

Figure 5: Densities of *Ex Ante* Treatment Effects Conditional on the Actual Treatment Status: Health

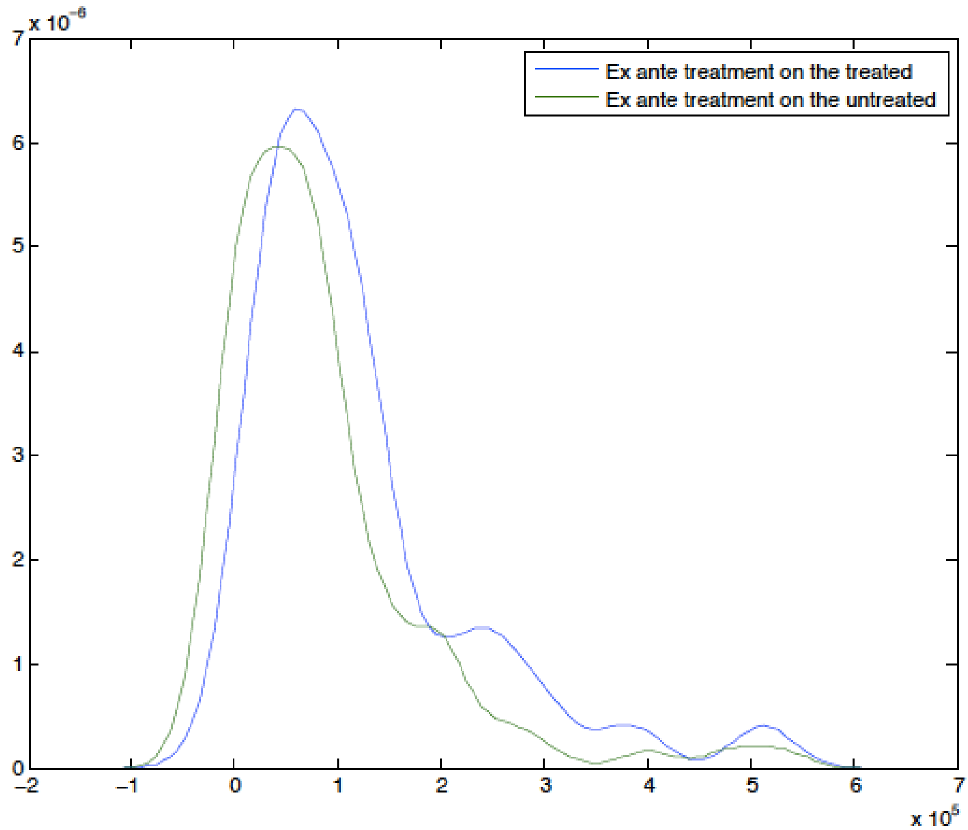

Figure 6: Densities of *Ex Ante* Treatment Effects Conditional on the Actual Treatment Status: Business

4 Evolution of beliefs

4.1 Learning in college

In practice, expected incomes likely evolve over the course of college as students obtain new information about their own major and occupation-specific abilities, as well as about the average wages and returns to ability within each occupation and major. While for each student we only elicit expectations at a given point in time, students in our sample are enrolled in different years of college. In the following we use a synthetic cohort approach and examine how students update their beliefs by comparing the distributions of expected incomes for lower-classmen with those of upper-classmen.

4.1.1 Beliefs about the average student

As students learn about the average incomes within each occupation and major, one should expect the within-sample dispersion of income beliefs about the average Duke student to decline over time. Table 11 below reports, for all pairs of occupation and major, the relative change between lower and upper-classmen in the variance of log-expected incomes for the average Duke student. Consistent with students learning about the average incomes as they progress through college, the distribution of individual beliefs about the average Duke student is tighter among upper-classmen for the vast majority of occupation-major pairs (34 out of 36, albeit significantly so for 9 of them only).

Table 11 also shows that the magnitude of those changes in variances tends to be substantial. The variance of the log-expected incomes decreases indeed by as much as 35%, on average across all occupations and majors for which the variance declines over time. There is however quite a bit of heterogeneity across majors and occupations.

Table 11: Variances of log expected incomes for the average Duke student: differences between upper- and lower-classmen

	Occupation:						
Major:	Science	Health	Business	Government	Education	Law	All
Natural Sciences	**-0.16***	-0.05	-0.07	**-0.15***	-0.14	-0.08	**-0.11***
Humanities	-0.11	-0.01	-0.06	**-0.19***	**-0.27***	-0.09	**-0.12***
Engineering	**-0.14**	-0.17	-0.05	**-0.13**	-0.14	-0.01	**-0.11**
Social Sciences	-0.12	-0.05	0.00	-0.14	**-0.18**	-0.09	**-0.10**
Economics	-0.04	-0.01	-0.03	-0.10	0.37	-0.08	0.02
Public Policy	-0.07	-0.06	-0.08	**-0.10**	**-0.18**	-0.02	-0.09
All	**-0.11**	-0.06	-0.05	**-0.13***	**-0.09**	-0.06	**-0.08****

Note: Major can either be the chosen major or a counterfactual major. Differences that are statistically significant at the 10% level are reported in bold. $*$ and $**$ indicate statistical significance at the 5% level and 1% level, respectively. "All" indicates average across majors (rows) and occupations (columns).

4.1.2 Beliefs about own abilities

In the following, we show how changes in beliefs about the average student can be combined with changes in beliefs about own incomes to identify the evolution of individual-level uncertainty about own abilities. To do so, we need to impose some restrictions on the income processes, as well as on how individuals form their expectations. Specifically, for any given individual i, major j and occupation k , we assume that the income $(Y^i_{j,k})$ can be decomposed as follows:

$$\begin{aligned} Y^i_{j,k} &= \exp(\mu^i_{j,k} + \overline{\mu}_{j,k}) & (4.1) \\ &= \exp(\mu^i_{j,k})\overline{Y}_{j,k} & (4.2) \end{aligned}$$

where $\mu^i_{j,k}$ and $\overline{\mu}_{j,k}$ denote the (major,occupation)-specific individual and mean ability, and $\overline{Y}_{j,k}$ is the income of the average Duke student for that same major-occupation pair. Under the assumption that individuals have normally distributed prior beliefs on $(\mu^i_{j,k}, \overline{\mu}_{j,k})$ in each period, we write the beliefs about $Y^i_{j,k}$, $Y^i_{j,k,t}$, as:

$$\begin{aligned} Y^i_{j,k,t} &= E\left(\exp(\mu^i_{j,k} + \overline{\mu}_{j,k})|\mathcal{I}_{i,t}\right) & (4.3) \\ &= \exp\left(\mu^i_{j,k,t} + \overline{\mu}^i_{j,k,t} + (\sigma^i_{j,k,t})^2/2 + (\overline{\sigma}^i_{j,k,t})^2/2 + \rho^i_{j,k,t}\right) & (4.4) \end{aligned}$$

where $\mathcal{I}_{i,t}$ denotes individual i's information set at t, $\left(\mu^i_{j,k,t}, \overline{\mu}^i_{j,k,t}\right)$, $\left(\sigma^i_{j,k,t}, \overline{\sigma}^i_{j,k,t}\right)$ are the means and standard deviations of the prior distributions of the individual and mean ability, and $\rho^i_{j,k,t}$ is the covariance of the prior joint distribution of individual and mean ability. Similarly, the beliefs about the average Duke student's income are given by:

$$\begin{aligned} \overline{Y}^i_{j,k,t} &= E\left(\exp(\overline{\mu}_{j,k})|\mathcal{I}_{i,t}\right) & (4.5) \\ &= \exp\left(\overline{\mu}^i_{j,k,t} + (\overline{\sigma}^i_{j,k,t})^2/2\right) & (4.6) \end{aligned}$$

Taking the logs and computing the difference between beliefs about own income and beliefs about the average Duke student yields:

$$\ln(Y^i_{j,k,t}) - \ln(\overline{Y}^i_{j,k,t}) = \mu^i_{j,k,t} + (\sigma^i_{j,k,t})^2/2 + \rho^i_{j,k,t} \quad (4.7)$$

This equality plays a key role in our analysis. It is important to note that, while the derivation above implicitly assumes that students are making rational expectations over their own earnings and over those of the average Duke student, this assumption is stronger than necessary. For instance, the specification in (4.7) still holds if we relax the rational expectations assumption and write instead the individual earnings beliefs as $\mathcal{E}(Y^i_{j,k}|\mathcal{I}_{i,t}) = \kappa \times E(Y^i_{j,k}|\mathcal{I}_{i,t})$, where $\kappa \neq 1$.

We are specifically interested here in the evolution over time of the uncertainty about individual-specific ability, that is how $\sigma^i_{j,k,t}$ changes between lower and upper-classmen. Assuming that individuals are forming rational expectations over their own abilities, $E(\mu^i_{j,k,t})$ will remain constant across t. If we further assume that the covariance terms $\rho^i_{j,k,t}$ are equal to zero, then we can identify the evolution of uncertainty over time using a difference-in-differences strategy.[24] Namely:

$$E\left(\ln(Y^i_{j,k,t+1}) - \ln(\overline{Y}^i_{j,k,t+1})\right) - E\left(\ln(Y^i_{j,k,t}) - \ln(\overline{Y}^i_{j,k,t})\right) \tag{4.8}$$
$$= E\left((\sigma^i_{j,k,t+1})^2/2 - (\sigma^i_{j,k,t})^2/2\right) \tag{4.9}$$

It follows that the evolution between upper and lower-classmen of the uncertainty about individual-specific abilities is directly identified from the data and can be consistently estimated from the empirical counterpart of the left hand-side.

Table 12: Change between upper and lower-classmen in the variances of own beliefs

	Occupation:						
Major:	Science	Health	Business	Government	Education	Law	All
Natural Sciences	-0.04	**-0.13**	-0.06	-0.02	-0.05	-0.07	**-0.06***
Humanities	-0.07	-0.08	-0.08	-0.03	0.01	-0.08	**-0.05***
Engineering	-0.05	**-0.13**	-0.01	-0.03	-0.03	-0.07	**-0.05**
Social Sciences	-0.07	-0.05	-0.11	-0.06	-0.06	**-0.17****	**-0.08****
Economics	-0.08	-0.04	**-0.11**	-0.07	0.05	**-0.11****	**-0.06****
Public Policy	-0.05	-0.05	**-0.14***	-0.08	-0.06	-0.07	**-0.08****
All	**-0.06***	**-0.08****	**-0.08****	**-0.05**	-0.02	**-0.10****	**-0.07****

Note: Major can either be the chosen major or a counterfactual major. Differences that are statistically significant at the 10% level are reported in bold. $*$ and $**$ indicate statistical significance at the 5% level and 1% level, respectively. "All" indicates average across majors (rows) and occupations (columns).

The estimation results are reported In Table 12. A first important takeaway is that, with the exception of the pairs Education-Economics and Education-Humanities, all of the entries from this table are negative. These results are consistent with students learning about their own occupation and major-specific abilities as they progress through college. A second takeaway from this table is that the absolute decrease in the posterior variance of the individual beliefs is faster for occupations such as law, business and health, while it is slower for occupations such as education and government. This pattern is consistent with

[24]This condition is stronger than necessary as the equality below holds as long as the covariance terms to remain on average constant over time ($E(\rho^i_{j,k,t}) = E(\rho^i_{j,k,t+1})$).

individuals being initially more uncertain about their own abilities in the former occupations. To illustrate this point, consider a simple two-period learning model where individuals update their ability beliefs in a Bayesian fashion after receiving a noisy signal. All else equal the decrease in prior variance is larger in magnitude if individuals are initially more uncertain about their ability, since, assuming normally distributed prior and noise distributions:

$$|\sigma_1^2 - \sigma_0^2| = \frac{1}{1 + \sigma_\epsilon^2/\sigma_0^2} \tag{4.10}$$

where σ_0^2 and σ_1^2 are the prior ability variances in period $t = 0$ and $t = 1$, and σ_ϵ^2 is the noise variance.

Finally, the evolution of uncertainty about individual-specific ability relative to a baseline major-occupation pair is identified under milder assumptions. Specifically, assuming that the evolution over time of the covariance terms $\rho_{j,k,t}^i$ is the same across all pairs of majors and occupations, we can identify the evolution of uncertainty over time (relative to a baseline major-occupation) using a triple differences strategy. Namely:

$$\Delta\left(E\left(\ln(Y_{j,k,t+1}^i) - \ln(\overline{Y}_{j,k,t+1}^i)\right) - E\left(\ln(Y_{j,k,t}^i) - \ln(\overline{Y}_{j,k,t}^i)\right)\right) \tag{4.11}$$

$$= \Delta\left(E\left((\sigma_{j,k,t+1}^i)^2/2 - (\sigma_{j,k,t}^i)^2/2\right)\right) \tag{4.12}$$

where $\Delta(.)$ denotes the difference between (j, k) and a baseline (major,occupation) pair (j_0, k_0). It follows that the evolution between upper and lower-classmen of the uncertainty about individual-specific beliefs (relative to (j_0, k_0)) is directly identified from the data and can be consistently estimated from the empirical counterpart of the left hand-side. Estimation results using Humanities-Education as a baseline alternative are reported in Table 20 in the Appendix. Overall, this table supports the same generalizations as the ones discussed above. These results further strengthen previous evidence suggesting that the speed of learning is heterogeneous across major-occupation pairs, with the decrease in posterior variance of individual beliefs being statistically significantly faster for occupations and majors such as Law and Social Sciences, Law and Economics as well as Business and Public Policy relative to Education and Humanities.

4.2 *Ex ante* treatment effects seven years later

The previous section showed how beliefs evolve over the course of the college career. We now examine how beliefs about the treatment effects of different occupations have evolved from when the students were in college to the present. For each occupation, Table 13 shows

Table 13: *Ex Ante* Treatment Effects of Occupations Seven Years Later (Annual Earnings, in 2009 dollars)

Occupation	TT	TUT	ATE
Science	61,879	49,942	51,968
	(14,337)	(6,070)	(5,786)
Health	119,588	35,131	54,224
	(26,631)	(5,352)	(8,897)
Business	220,938	82,518	139,815
	(28,211)	(10,380)	(17,843)
Government	18,008	9,524	10,046
	(3,932)	(1,979)	(1,927)
Law	54,175	66,990	65,995
	(16,723)	(8,168)	(7,763)

Note: Sample is 112 respondents to Phase 3 survey

the treatment on the treated, the treatment on the untreated and the average treatment effect computed using the beliefs about expected earnings that were elicited in the DuCMES Phase 3 survey. This table replicates the results in Table 7 based on beliefs data collected in the Phase 1 survey. Recall that this survey asked students to give us their current beliefs of what they expect to be earnings 10 years after receiving their BA.[25]

While the patterns are generally similar across the two tables, the treatment on the treated is substantially higher in business, more than doubling between the two surveys. For law, both the treatment on the treated and the treatment on the untreated fall, but the former fall is much larger such that the treatment on the treated is actually lower than the treatment on the untreated. The likely driver for these changes is that those who perceived a high return to law initially saw large changes in their returns to business, shifting them to the treated group in business.

These shifts in treatment effects relate directly to changes in probabilities of choosing occupations. Recall that the Phase 2 and Phase 3 data both revealed significantly higher shares going into business than in the Phase 1 survey, consistent with expected earnings in business rising. Similarly, the largest shifts away from occupations occurred in law and government, both of which saw a decrease in expected treatment effects. In the next section, we focus on sorting across occupations and directly relate these changes in expected earnings to changes in probabilities of choosing particular occupations.

[25]Recall that in the Phase 1 survey in 2009, students in our sample cam from all four classes (freshman, sophomores, juniors and seniors), so at the time of the Phase 3 survey in 2016, the students were 4-6 years since graduation, giving us variation in how close they currently were to the 10-year benchmark used in our Phase 3 elicitations.

5 Sorting on expected gains

5.1 Choice of occupation

While the results discussed in Section 3 point to a positive and significant association between expected earnings and choice of occupation, they may partly reflect preferences for occupation-specific non-pecuniary job attributes that are correlated with the expected earnings. In this section we go step further and address this issue by modeling the choice of occupations, linking it to expected earnings and (unobserved) non-pecuniary job attributes.

The self-reported probabilities of working in each occupation reflect uncertainty. Suppose the uncertainty takes the form of a Type 1 extreme value shock to each occupation. We can then view the self-reported probabilities as individuals knowing that they will choose their occupation given major j to maximize

$$d^*_{ij} = \arg\max_{(d_{ik})_k} \sum_k d_{ik}(u_{ijk} + \epsilon_{ijk}) \tag{5.1}$$

where u_{ijk} denotes the expected utility at the time of the initial survey associated with occupation k and major j.

With the unknown information in each occupation drawn from a Type 1 extreme value distribution, the self-reported probabilities will follow a multinomial logit. We can then invert the self-reported probabilities to obtain estimates of u_{ijk} using:

$$\ln(p_{ijk}) - \ln(p_{ij1}) = u_{ijk} - u_{ij1} \tag{5.2}$$

We can further expand on equation (5.2) to link expected earnings to the choice of occupation. Namely, we express u_{ijk} as the sum of an occupational dummy, α_k, and the log of the elicited expected earnings in the occupation given the major, $\ln(Y_{ijk})$, and a residual, ζ_{ijk}, that is known to the individual but not to the econometrician:

$$u_{ijk} = \alpha_k + \beta \ln(Y_{ijk}) + \zeta_{ijk} \tag{5.3}$$

Assuming the sum of ζ_{ijk} and ϵ_{ijk} follows a Type 1 extreme value distribution we estimate a conditional logit on the ex-post choice. Results for both the full sample and the sample excluding seniors are displayed in Column (1). We estimate the model without seniors as some seniors may already have jobs lined up at the time of the survey. For both samples, the estimates show a significant and strong relationship between expected future earnings and occupational choice.[26] This provides clear evidence of positive sorting on expected earnings gains, consistent with a Roy model of occupational choice.[27]

A similar relationship also holds between the self-reported probabilities and expected future earnings. Expanding on equation (5.2), we can estimate the following regression conditional on the chosen major:

$$\begin{aligned} \ln(p_{ijk}) - \ln(p_{ij1}) &= u_{ijk} - u_{ij1} \\ &= (\alpha_k - \alpha_1) + \beta\left[\ln(Y_{ijk}) - \ln(Y_{ij1})\right] + (\zeta_{ijk} - \zeta_{ij1}) \end{aligned} \tag{5.4}$$

where α_1 is normalized to zero.[28] We deal with the zero self-reported probabilities by replacing them by an arbitrarily small number, as proposed by Blass et al. (2010). We then estimate the flow utility parameters using a least absolute deviation (LAD) estimator.[29] Results of the LAD estimation of (5.4) are given in Column (2) of Table 14. The estimated coefficient on log income (1.25) is economically and statistically significant for this specification. The coefficient is smaller when excluding seniors (1.14), but remains overall of similar order of magnitude.

Having a more continuous dependent variable in the self-reported probabilities case makes it possible to control for major times occupation dummies even with our small sample. Results are presented in Column (3). Adding the major-occupation interactions leads to a decrease in the coefficient on log income. This is especially true when excluding seniors where the coefficient on log income falls by 40%.

But even more information is available using the self-reported probabilities as we also have information on expectations regarding counterfactual majors. Using both data on actual and counterfactual majors makes it possible to account for individual-occupation fixed effects in addition to the major-occupation interactions.[30] The results reported in Column (5)

[26] If one maintains the assumption that students form rational expectations over their future choice of occupation, ruling out aggregate shocks on $u_{ijk} - u_{ij1}$ is more restrictive than necessary to identify the marginal utility of log-earnings β from the actual choices of occupation. Occupation-specific shocks affecting log-earnings additively would be absorbed by the occupation dummies, implying that the conditional logit will still consistently estimate the earnings coefficient β in the presence of these types of shocks. Note also that, with this specification of the flow utility, the multiplicative shocks affecting the occupation shares which are discussed in Section 3) can be derived from these types of additive shocks on log-earnings.

[27] We also considered an alternative specification where we assumed that preferences are linear, as opposed to logarithmic, in the expected occupation-specific earnings. Results from a Vuong test for non-nested model selection lead to rejection of the null hypothesis at the 1% level (P-value of 0.004), indicating that the specification with log expected earnings fits the data better.

[28] Note that for the self-reported probabilities specification the sum of ζ_{ijk} and ϵ_{ijk} need not be Type 1 extreme value to justify the specification.

[29] The resulting estimator is consistent, for a fixed number of majors, under a zero median restriction on the residuals.

[30] See Wiswall and Zafar (2016b) who provide evidence from a sample of NYU students that preferences

are obtained after eliminating the individual-occupation fixed effects by applying a within-transformation to the data where we now cluster the standard errors at the individual-occupation level. Due to the sample size increase, the standard error on log income falls substantially despite the rich set of additional controls. Estimates of the coefficient on log income are now slightly less than one.

To quantify the sensitivity of subjective choice probabilities to expected earnings, we calculate for the specification in Column (4) the percentage change in the probability of choosing an occupation given a percentage change in expected earnings, denoted by e_{ijk}. The elasticity formula for our specification is (Train, 2003):

$$e_{ijk} = (1 - p_{ijk})\beta. \tag{5.5}$$

Note that this formula only applies for the intensive margin, that is for those individuals such that the subjective probability p_{ijk} is strictly bounded between 0 and 1. For those individuals in our sample, the subjective probabilities of entering a given career conditional on a given major range from 0.003 to 0.962, yielding elasticities from 0.04 to 0.95. Taking the chosen (or intended) major from the data as given, we can estimate the average elasticity for occupation k using:

$$\hat{e_k} := \frac{\sum_i \sum_j d_{1ij}(1 - p_{ijk})\widehat{\beta}}{N} \tag{5.6}$$

where $\widehat{\beta}$ denotes the LAD estimate of β ($\widehat{\beta} = 0.953$ for our preferred specification). These occupation-specific elasticities range from 0.65 (for business) to 0.82 (for education), resulting in a mean elasticity across all occupations equal to 0.74. It is worth noting that these elasticities are sizable, especially in comparison with the very low earnings elasticities which have been found in the literature on college major choices (see, e.g., Beffy et al., 2012; Long et al., 2015; Wiswall and Zafar, 2015; and Altonji et al., 2016, for a survey).

In Table 15, we repeat the ex ante analysis using instead the subjective probabilities of choosing particular occupations and the occupation-specific expected earnings that were collected in the follow-up survey. We consider three different specifications where we control for occupation (Column 1), major interacted with occupation (Column 2), and add a dummy variable indicating whether the occupation is the current one (Column 3). In all cases, the estimated coefficient associated with log expected earnings is positive and significant at any standard level. Comparing the first two columns with Columns (2) and (3) from Table 14 provides evidence that beliefs about future choice of occupation tend to be more tightly associated with expected earnings in the follow-up survey than in the initial survey when

for job attributes are highly heterogeneous across individuals.

Table 14: Conditional Logit of Occupational Choice

	Ex Post	Ex Ante		
	(1)	(2)	(3)	(4)
Full Sample				
Log Income	1.484	1.371	1.000	0.953
	(0.299)	(0.271)	(0.332)	(0.148)
Excluding Seniors				
Log Income	1.589	1.337	0.688	1.014
	(0.346)	(0.310)	(0.333)	(0.177)
Occupation Dummies	Y	Y	N	N
Major × Occupation	N	N	Y	Y
Individual × Occupation	N	N	N	Y

Standard errors in parentheses. For specifications 4, standard errors are clustered at the individual X occupation level.
167 (113) individuals are included in the full sample (excluding seniors). We use observations corresponding to chosen majors only for specifications 3 and 4. Specification 4 is estimated using observations for chosen as well as non-chosen majors, giving us 6 times the number of observations in the sample.

the individuals were all still enrolled in college. The estimated coefficient decreases in the final specification where we control for current occupation (from 1.257 to 0.914). This pattern is consistent with the existence of switching costs, as well as with sorting across occupations on nonpecuniary benefits. We will focus on the later mechanism in Section 6.

Finally, the longitudinal aspect of the data makes it possible to estimate the association between changes in subjective probabilities of choosing particular occupations and changes in the occupation-specific expected earnings. Table 16 reports the corresponding LAD estimation results. In all three specifications the estimated earnings coefficient remains positive and significant, both statistically and economically. Besides, focusing on Specification (2) where we control for whether the occupation is the actual occupation from the follow-up survey, it is interesting to note that the estimated earnings coefficient (1.020) is of similar order of magnitude to the estimates that were obtained for the most comparable specifications in Table 15 (Column 3, 0.914) and in Table 14 (Column 4, 0.953). While the magnitude decreases once we allow aggregate preferences for majors and occupations to vary over time by adding occupation-major fixed effects, the estimated coefficient (0.783) remains statistically significant and sizable.

Table 15: Occupational Choice (Beliefs from the Follow-Up Survey)

	(1)	(2)	(3)
Log Income	1.848	1.257	0.914
	(0.202)	(0.182)	(0.159)
Occupation Dummies	Y	N	N
Major × Occupation	N	Y	Y
Current Occupation	N	N	Y

Standard errors in parentheses.
112 individuals are included in the estimation sample.

Table 16: Changes in Subjective Probabilities of Choosing Occupations

	(1)	(2)	(3)
Δ Log Income	1.274	1.020	0.783
	(0.239)	(0.235)	(0.206)
Current Occupation	N	Y	Y
Major × Occupation	N	N	Y

Standard errors in parentheses.
112 individuals are included in the estimation sample.

5.2 Beyond the choice of occupation: expected versus actual earnings

We conclude this section by examining the relationship between the actual earnings that were collected in our Phase 3 follow-up survey conducted in 2016 and the expected earnings elicited in 2009 when the individuals were still enrolled in college. In Table 17, we report the estimation results from a linear regression of log (actual) earnings on log expected earnings in chosen occupation, with and without controlling for chosen major and chosen occupation. While the estimates are imprecise as a result of the small number of individuals for whom we measure realized earnings, the estimated elasticities are positive and significant at the 1% level for Specifications (1) (no controls) and (2) (control for chosen major). It is interesting to compare our results from Specification (1) with Wiswall and Zafar (2016a) who estimate in a different context the association between log realized earnings and log expected earnings. Among males, they find a positive but insignificant relationship between these two quantities, with a smaller estimated elasticity of 0.167 and a R^2 equal to 0.018 (against 0.187 for our estimates).[31]

Turning to Specification (3), the estimated coefficient is only significant at the 10% level when we control for chosen major and occupation. However, even for this specification, the magnitude of the expected earnings elasticity for realized earnings (0.448) remains sizable. Taken together, these results show that beliefs about earnings are predictive of actual future earnings. That the elasticity of realized earnings with respect to expected earnings remains non-negligible, even after controlling for a respondent's chosen major and chosen occupation, is suggestive evidence that earnings beliefs matter above and beyond their effect on the choice of occupation.[32]

Table 17: Relationship between Actual and Expected Earnings in Chosen Occupation

	Log Earnings		
	(1)	(2)	(3)
Log Expected Earnings	0.859	0.661	0.448
	(0.253)	(0.251)	(0.244)
Chosen Major	N	Y	Y
Chosen Occupation	N	N	Y
R^2	0.187	0.377	0.534

Standard errors in parentheses.
52 individuals are included in the estimation sample. All specifications include a constant term.

[31] Wiswall and Zafar (2016a) show that beliefs are much more predictive of the actual earnings among females, with a significant estimated elasticity of 0.521 and a R^2 of 0.153.

[32] We also ran a linear regression of log earnings on log expected earnings in all (chosen and counterfactual) occupations, controlling for chosen major and chosen occupation. The coefficients associated with the occupation-specific log expected earnings are jointly significant at the 5% level (P-value of 0.044).

6 Efficiency and nonpecuniary benefits

Given that we have expected earnings across all occupations, we can see how much income individuals expect to give up as a result of not choosing the occupation that maximizes their income. This allows us to provide direct evidence on the role played by non-monetary factors in the choice of occupation. Importantly, this expected willingness-to-pay for non-monetary factors is directly identified from the data, and it does not require making any distributional assumption on the non-monetary factors affecting the choice of occupation.[33] An issue with using the Phase 1 data to address this question is that some of the occupations such as health and law likely require additional schooling. But the Phase 3 data likely does not suffer from this issue. Individuals have either completed their education or will do so soon. Hence we use the Phase 3 data to see how much income is lost by individuals expecting to choose occupations where their expected earnings are lower than the occupation with the highest expected earnings.

The first panel of Table 18 shows lost income due to sorting into occupations across factors besides salary. The first column of Table 18 averages the earnings associated with the (individual-specific) highest paying occupation across individuals. The second column calculates average expected earnings using the Phase 3 self-reports on the probabilities of working in each of the occupations as weights. The difference between the first two columns, given in column 3, then gives the lost income associated with not working in the occupation that maximizes their income. The average gap of $33,624 represents about fourteen percent of the maximum earnings individuals expect to receive. Note that this is still a lower bound on efficiency losses as it does not take into account any sorting into jobs within an occupation category.

The second and third panels of Table 18 provides information how the earnings losses are spread across the respondents. The second panel shows that almost 27% of respondents reported that they would be working in the occupation that maximized their expected earnings with certainty with the overall probability of working in the occupation with highest expected earnings at almost 58%. Almost 10% report with certainty that they will be working in one occupation at that occupation is not the one that maximizes their expected earnings. Hence, around 37% of individuals have some uncertainty regarding their occupational choices in the next three to six years. The third panel shows lost incomes for the

[33] Related work by D'Haultfoeuille and Maurel (2013) investigates the relative importance of ex ante monetary returns versus non-pecuniary factors in the context of an extended Roy model applied to the decision to attend college. Their approach, which can be used in the absence of subjective expectations data and does not require exclusion restrictions, relies on stronger assumptions on the non-pecuniary factors. See also Eisenhauer et al. (2015), who use exclusion restrictions between monetary returns and non-pecuniary factors to separately identify these two components.

Table 18: Maximum and Expected Earnings: Phase 3

	Full Sample		
	Max Earnings	Expected Earnings	Gap
Mean	239,266	205,641	33,624
1^{st} quartile	133,500	94,429	0
Median	178,000	161,090	16,020
3^{rd} quartile	267,000	236,411	39,465
Standard Dev.	(185,544)	(166,494)	(54,413)
	Share Income Maximizing	Prob. Of Max Earn	Share Certain But Not Earnings Max
Mean	0.268	0.576	0.098
	Conditional on Expected Earnings<Max		
	Max Earnings	Expected Earnings	Gap
Mean	216,855	170,929	45,926
1^{st} quartile	124,787	76,540	13,350
Median	178,000	144,358	26,923
3^{rd} quartile	222,500	210,930	53,400
Standard Dev.	(174,108)	(135,896)	(59,037)

Note: Sample is 112 respondents to Phase 3 survey

63% of the sample who reported at least some probability greater than zero of working in an occupation that did not maximize their earnings. This group had lower maximum earnings than respondents as a whole. On average, this group expects to give up $45,926 of earnings as a result of not choosing the income-maximizing occupation, or a little over 21% of their maximum earnings.

7 Conclusion

This paper uses elicited beliefs from a sample of male undergraduates at Duke University on the expected earnings in different occupations as well as on the probabilities of working in each of those occupations, to recover the distributions of the *ex ante* returns (or *ex ante* treatment effects on earnings) for particular occupations, and to quantify the importance of sorting on expected gains. Importantly, these beliefs were asked not only for the college major the individual chose or intended to choose, but also for all counterfactual majors, thus making it possible to examine the complementarities between majors and occupations.

The distributions of the *ex ante* monetary returns (or *ex ante* treatment effects on earnings) for particular occupations, conditional on each college major, are directly identified from our subjective expectations data. We find large differences in expected earnings across occupations, with a substantial degree of heterogeneity across individuals. The estimates also suggest that those who place high probabilities on working in particular occupations also tend to expect the greatest monetary returns from those occupations, consistent with sorting on (expected) gains. Clear complementarities exist between majors and occupations. For example, expected returns for business careers are highest for economics major, which in turn leads individuals to report higher probabilities of pursuing a business occupation in the (sometimes) hypothetical case that they were an economics major. Comparing the distributions of expected earnings between lower- and upper-classmen further suggests that students learn about the average returns to the various occupation-major pairs over the course of college.

Linking our subjective expectations data with the actual choices of occupations, we find a significant and quantitatively large association between the *ex ante* returns and actual choice of occupation. Occupational choice probabilities that were elicited while the individuals were still enrolled in college turns out to be very informative about actual sorting across occupations on expected gains. We also find that beliefs about earnings are strong predictors of actual future earnings seven years later. Interestingly, expected earnings are positively associated with realized earnings even after controlling for chosen major and occupation, which suggests that earnings beliefs matter beyond their effect on the choice of occupation, possibly through the position level within the occupation. Taken together, our findings illustrate the value of collecting subjective expectations data on choice probabilities and counterfactual outcomes to recover *ex ante* treatment effects, and quantify the importance of sorting on expected gains.

References

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

_ , Peter Arcidiacono, and Arnaud Maurel, "The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects," in Eric Hanushek, Stephen Machin, and Ludger Wößmann, eds., *Handbook of the Economics of Education*, Vol. 5, Elsevier, 2016, pp. 305–396.

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

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

_ , V. Joseph Hotz, and Songman Kang, "Modeling college major choices using elicited measures of expectations and counterfactuals," *Journal of Econometrics*, 2012, *166*, 3–16.

Beffy, Magali, Denis Fougere, and Arnaud Maurel, "Choosing the Field of Studies in Postsecondary Education: Do Expected Earnings Matter?," *Review of Economics and Statistics*, 2012, *94*, 334–347.

Blass, Asher A., Saul Lach, and Charles F. Manski, "Using elicited choice probabilities to estimate random utility models: preferences for electricity reliability," *International Economic Review*, 2010, *51*, 421–440.

Carneiro, Pedro M., Karsten T. Hansen, and James J. Heckman, "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*, 2003, *44*, 361–422.

Cunha, Flávio and James J. Heckman, "Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling," *Labour Economics*, 2007, *14*, 870–893.

_ and _ , "A new framework for the analysis of inequality," *Macroeconomic Dynamics*, 2008, *12*, 315–354.

_ , _ , and Salvador Navarro, “Separating uncertainty from heterogeneity in life cycle earnings,” *Oxford Economic Papers*, 2005, *57*, 191–261.

Delavande, Adeline, “Pill, patch, or shot? Subjective expectations and birth control choice,” *International Economic Review*, 2008, *49*, 999–1042.

_ and Basit Zafar, “University Choice: The Role of Expected Earnings, Non-pecuniary Outcomes and Financial Constraints,” 2014. Working paper.

D’Haultfoeuille, Xavier and Arnaud Maurel, “Inference on an Extended Roy Model, with an Application to Schooling Decisions in France,” *Journal of Econometrics*, 2013, *174*, 95–106.

Dominitz, J. and C. Manski, “Eliciting student expectations of the returns to schooling,” *Journal of Human Resources*, 1996, *31*, 1–26.

Dominitz, Jeff and Charles F. Manski, “Using expectations data to study subjective income expectations,” *Journal of the American Statistical Association*, 1997, *92*, 855–867.

Eisenhauer, Philipp, James J. Heckman, and Edward Vytlacil, “The generalized Roy model and the cost-benefit analysis of social programs,” *Journal of Political Economy*, 2015, *123*, 413–443.

Heckman, James J., John E. Humphries, and Gregory Veramendi, “Returns to Education: The Causal Effects of Education on Earnings, Health and Smoking,” *Forthcoming in the Journal of Political Economy*, 2016.

Hurd, Michael, “Subjective Probabilities in Household Surveys,” *Annual Review of Economics*, 2009, *1*, 543–562.

_ and Kathleen McGarry, “Evaluation of the subjective probabilities of survival in the HRS,” *Journal of Human Resources*, 1995, *30*, S268–S292.

_ and _ , “The predictive validity of subjective probabilities of survival,” *Economic Journal*, 2002, *112*, 966–985.

Kang, Songman, “Eliciting expectations and counterfactuals about college major choice: Data Documentation,” 2009. Working paper.

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

Kinsler, Joshua and Roni Pavan, "The specifity of general human capital: evidence from college major choice," *Journal of Labor Economics*, 2015, *33*, 933–972.

Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad, "Field of Study, Earnings, and Self-Selection," *Forthcoming in the Quarterly Journal of Economics*, 2016.

Long, Mark C., Dan Goldhaber, and Nick Huntington-Klein, "Do Students' College Major Choices Reflect Changes in Wages?," *Economics of Education Review*, 2015, *49*, 1–14.

Manski, C., "Measuring Expectations," *Econometrica*, 2004, *72*, 1329–1376.

Manski, Charles F., "Adolescent Econometricians: How Do Youths Infer the Returns to Schooling?," in C.T. Clotfelter and M. Rothschild, eds., *Studies of Supply and Demand in Higher Education*, Chicago: University of Chicago Press, 1993, pp. 43–57.

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

Pantano, Juan and Yu Zheng, "Using Subjective Expectations Data to Allow for Unobserved Heterogeneity in Hotz-Miller Estimation Strategies," 2013. Working paper.

Siow, Alioysus, "Occupational choice under uncertainty," *Econometrica*, 1984, *52*, 631–645.

Stinebrickner, Todd and Ralph Stinebrickner, "A major in science? Initial beliefs and final outcomes for college major and dropout," *Review of Economic Studies*, 2014, *81*, 426–472.

Train, Kenneth, *Discrete Choice Methods with Simulation*, Cambridge University Press, 2003.

van der Klaauw, Wilbert, "On the use of expectations data in estimating structural dynamic choice models," *Journal of Labor Economics*, 2012, *30*, 521–554.

_ **and Kenneth I. Wolpin**, "Social security and the retirement and savings behavior of low-income households," *Journal of Econometrics*, 2008, *145*, 21–42.

Wiswall, Matthew and Basit Zafar, "Determinants of college major choice: identification using an information experiment," *Review of Economic Studies*, 2015, *82*, 791–824.

_ **and** _ , "Human Capital Investments and Expectations about Career and Family," 2016. NBER Working Paper No. 22543.

_ **and** _ , "Preference for the Workplace, Human Capital, and Gender," 2016. NBER Working Paper No. 22173.

Zafar, Basit, "College major choice and the gender gap," *Journal of Human Resources*, 2013, *48*, 545–595.

A Appendix

A.1 Actual Majors at Duke and Major 'Groups'

The following is the list of majors at Duke and the six Groups we used to classify them:

Science	***Engineering***
Biological Anthropology and Anatomy	Computer Science
Biology	Biomedical Engineering
Chemistry	Civil Engineering
Earth & Ocean Sciences	Electrical & Computer Engineering
Mathematics	Mechanical Engineering
Physics	
Humanities	***Social Sciences***
Art History	Cultural Anthropology
Asian and African Languages and Literature	History
Classical Civilization/Classical Languages	Linguistics
Dance	Psychology
English	Sociology
French Studies	Women's Studies
German	
International Comparative Studies	***Economics***
Italian Studies	Economics
Literature	
Medieval & Renaissance Studies	***Policy***
Music	Environmental Science and Policy
Philosophy	Political Science
Religion	Public Policy Studies
Spanish	
Theater Studies	
Visual Arts	

A.2 Tables

Table 19: Average *Ex Ante* Treatment Effects of Occupations: Lower-Classmen versus Upper-Classmen (Annual Earnings, in dollars)

Occupation	Lower-classmen ATE	Upper-classmen ATE	P-value
Science	20,796	23,424	0.66
	(4,652)	(3,733)	
Health	61,657	72,492	0.51
	(13,911)	(8,448)	
Business	75,981	98,961	0.48
	(30,760)	(10,406)	
Government	24,803	26,608	0.83
	(6,333)	(5,627)	
Law	74,450	98,608	0.33
	(19,873)	(15,011)	

Note: Standard errors are reported in parentheses. Reported P-values correspond to a t-test of equality of the average *ex ante* treatment effects between lower-classmen and upper-classmen.

Table 20: Change between upper and lower-classmen in the variances of own beliefs (relative to Education, Humanities)

				Occupation:			
Major:	Science	Health	Business	Government	Education	Law	All
Natural Sciences	-0.05	-0.15	**-0.07**	-0.03	-0.06	-0.08	-0.07
Humanities	-0.08	-0.09	-0.10	-0.04	0.00	**-0.09**	-0.07
Engineering	-0.06	**-0.15***	-0.02	-0.05	-0.05	-0.08	-0.07
Social Sciences	-0.08	-0.06	**-0.12**	-0.07	-0.07	**-0.18****	-0.10
Economics	-0.09	-0.06	**-0.12***	-0.08	0.04	**-0.13***	-0.07
Public Policy	-0.06	-0.07	**-0.15***	-0.09	-0.07	-0.08	-0.09
All	-0.07	-0.09	-0.10	-0.06	-0.04	**-0.11**	-0.08

Note: Major can either be the chosen major or a counterfactual major. Differences that are statistically significant at the 10% level are reported in bold. $*$ and $**$ indicate statistical significance at the 5% level and 1% level, respectively. "All" indicates average across majors (rows) and occupations (columns).