Ex Ante Returns and Occupational Choice*

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

First version: October 2014
This version: March 25, 2017

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 returns 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 some college majors and occupations. Finally, we find clear evidence of sorting on expected returns, with the *ex ante* returns measured while the individuals were still in college being very informative about both their actual occupational choices and their subsequent earnings.

*We are grateful to the editor and four anonymous referees for helpful comments. We also thank Seth Zimmerman, and 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 2017 AEA Meetings (Chicago, IL), 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 Duke Empirical Micro Jamboree for helpful comments and suggestions. Luis Candelaria, Ashley DeVore and Kate Maxwell Koegel 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, Hotz, and Kang, 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 returns 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, Hansen, and Heckman, 2003; Cunha, Heckman, and Navarro, 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 on observed choices, we use data that directly elicits agents' *ex ante* returns. This allows 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 also have 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*, returns. As such, our paper complements the literature using observational data to show that individuals sort on *ex post* returns. Notable recent examples in the schooling context include Heckman, Humphries, and Veramendi (forthcoming) and Kirkeboen, Leuven, and Mogstad (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 a study of a sample of male undergraduates who participated in the Duke College Major and Expectations Survey (DuCMES). In Phase 1 of the DuCMES, conducted between February and April 2009, we elicited expectations about students' *ex ante* monetary returns to a set of possible occupations and the likelihood of their being in these occupations ten years after graduation.[5] Importantly, these occupation probabilities and expected incomes were elicited not only for each student's chosen major but also for counterfactual majors, i.e., the majors they did not choose. As we discuss below, these elicited expectations allow us to quantify the importance of sorting across occupations based on *ex ante* monetary returns and make it possible to identify how the returns to different occupations vary across majors and to examine the importance of complementarities between majors and occupations.

In particular, the data we collect 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, but also the *ex ante* treatment effects of particular majors on the probabilities of working in any given occupation. In order to quantify the role played by (expected) selection across occupations on the basis of expected returns, 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. This weighted average *ex ante* treatment effect 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. It is larger than the average *ex ante* treatment effect of occupation k if individuals expect to sort across occupations based on expected returns. Similarly, we are able to identify the *ex ante* treatment effect on the untreated, where we weight elicited expected returns by the declared probability that the individual will not work in occupation k.

[4]See Altonji, Blom, and Meghir (2012) and Altonji, Arcidiacono, and Maurel (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, Hotz, and Kang (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).

These data allow 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 a baseline occupation. 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 illustrates how sorting on expected returns varies throughout the distribution of *ex ante* returns.

Our results reveal large 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, consistent with selection into occupations with higher expected 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.

Next, we investigate the relationship between *ex ante* returns and beliefs and *actual* labor market outcomes, using two additional sources of data collected after the students in our sample completed their undergraduate degrees. Specifically, we collected the occupations that sample members were working in as of July 2015 for the vast majority of the DuCMES sample members, using data from the social network, *LinkedIn*, and the Duke Alumni Database. We also conducted a follow-up survey of all DuCMES sample members that was administered between February and April of 2016. In this follow-up survey we asked members of our sample about all of the occupations they held since graduating from Duke, including their current one, and about their current annual earnings.[6]

Using the *LinkedIn* and Alumni Database follow-up data on occupations, we can directly estimate the average *ex ante* treatment effects on the treated and untreated for each occupation. We find similar selection patterns for occupations using the declared probabilities of occupational choice elicited when our sample members were undergraduates or using their actual choices. Overall, this provides evidence that the beliefs that college students hold about their future (choice-specific) labor market outcomes are predictive of the labor market choices they make later in their lives.

We also examine how beliefs evolve over time in several different ways. First, we compare

[6]The respondents were contacted via email, *LinkedIn* message and/or text message and we obtained responses from 117 individuals or about 68% of the initial sample.

the expected incomes reported by lower- and upper-classmen when they were all still undergraduates. For example, we find that the beliefs that students have about the expected earnings of the average Duke students, which we elicit in the Phase 1 survey, are more homogeneous (i.e., less variable) for upper-classmen than are the earnings expectations of under-classmen. This finding is consistent with students learning about occupation-specific skill prices even before they even complete their undergraduate degrees.[7] Second, we explore how beliefs about the treatment effects of different occupations have evolved from college to early careers using data from our follow-up survey. In particular, we document the existence of sizable shifts in the expected returns to occupations such as Science, Business and Law, which relate to changes between both surveys in the subjective probabilities of choosing those occupations.

Finally, we quantify the importance of sorting across occupations on expected earnings returns. To do so, we consider a simple framework linking choice probabilities to expected earnings, and preferences for occupations. We find a robust positive, and statistically, as well as economically significant effect of *ex ante* monetary returns on occupational choice. Furthermore, we show that beliefs about earnings are predictive of what these individuals actually earned seven years later. The elasticity of realized earnings with respect to expected earnings remains sizable and significant after controlling for chosen major and occupation, thus suggesting that earnings beliefs at the time of college matter beyond their effect on the choice of occupation. Finally, we find large *ex ante* earnings losses implied by the subjective probabilities of working in each occupation, suggesting that non-pecuniary factors also play an important role in one's choice of occupations.

The rest of the paper is organized 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 the means and distributions of *ex ante* treatment effects given the data, and then discuss the estimation results. We then examine how beliefs evolve over time in Section 4, documenting how *ex ante* treatment effects changed between Phase 1 and Phase 3 survey, and contrasting earnings beliefs of under-classmen with those of upper-classmen. In Section 5, we examine how expected earnings and treatment effects differ based on the actual occupational choices of our respondents. In Section 6, we link subjective choice probabilities to expected earnings and preferences, and quantify the importance of sorting across occupations on expected earnings returns. Using data from the Phase 3 survey, we then examine the relationship between earnings beliefs and realized earnings. Next, we investigate in Section 7 the role

[7] In Appendix A.2, we outline a set of assumptions under which the data collected from under- and upper-classmen can be combined with beliefs about own earnings to infer how students update their beliefs about their own occupation-specific abilities.

that nonpecuniary factors play in occupational choice. We offer some concluding comments in Section 8. Additional details on the data, supplementary estimation results, and an analysis of how students update their beliefs about their own abilities are found in appendices.

2 Data

2.1 Phase 1 data

The data used in this paper is from the Duke College Major and Expectations Survey (DuCMES). The DuCMES first collected data from a sample of male undergraduate students at Duke University between February and April 2009.[8] We refer to this as Phase 1 of the DuCMES. 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.[9] All 173 students who completed the survey were paid $20.

Phase 1 of 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.[10] 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.

[8]Arcidiacono, Hotz, and Kang (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.

[9]A copy of the questionnaire used in the Phase 1 survey can be found at `public.econ.duke.edu/~vjh3/working_papers/college_major_questionnaire_ph1.pdf` and is discussed further in Kang (2009).

[10]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 is reported in Appendix A.1.

Table 1: Descriptive Statistics for Phase 1 Sample

	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 SOURCE: Phase 1 of DuCMES for the sample characteristics and Campus Life and Learning (CLL) Project at Duke University for Duke Male Student Body. See Arcidiacono et al. (2011) for a detailed description of the CLL dataset.

NOTE: 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?"

2.2 Expected choice probabilities and earnings

In Phase 1, the DuCMES elicited from the students their expectations about their likelihood of choosing future careers, and how much they expected to earn in them if the were to choose 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 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.[11] 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?".

Let $p_i(j, k, 1)$ denote the probability elicited from individual i of their choosing occupation k conditional on majoring in field j and where the last entry, 1, denotes that the elicitation was in Phase 1 of our study.

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?"

Let $Y_i(j, k, 1)$ denote individual i's elicited future income if he worked in occupation k and had majored in field j 10 years after graduating from Duke, again elicited during Phase 1. Below, we refer to this conditioning as being on *possible* occupations, since all of our respondents were still in college as Phase 1.

Table 2 reports the means of the expected incomes for the various major-occupation

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

combinations collected in Phase 1 of the DuCMES (the $Y_i(j,k,1)$'s).[12] 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 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 averages for the subjective probabilities of working in each occupation that were elicited from students in the Phase 1 survey (the $p_i(j,k,1)$'s). The subjective probabilities of entering each occupation vary substantially across majors. At the same time, 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 2.2). Overall, the likelihood of working in the various occupations appear to be selectively different across individuals, even after conditioning on a college major.[13]

Finally, Table A.3 in Appendix A.3 reports the prevalence of students reporting that the probability they would choose a particular occupation was zero for each major-occupation combination ($p_i(j,k,1) = 0$).[14] While some combinations display a large share of zero subjective probabilities, the shares are well below one, suggesting that particular majors do not rule out certain occupations for all individuals.

[12] 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 ($p_i(j,k,1) = 0$). 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.

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

[14] 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 2: Mean of Phase 1 expected incomes for different major/occupation combinations 10 years after graduation (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

DATA: Sample who completed Phase 1 survey ($N = 173$).
NOTE: Expected earnings were elicited for each possible major-occupation pair at Phase 1, regardless of the respondents' chosen or intended major.

Table 3: Mean of Phase 1 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

DATA: Sample who completed Phase 1 survey ($N = 173$).
NOTE: Expected earnings were elicited for each possible major-occupation pair at Phase 1, regardless of the respondents' chosen or intended major.

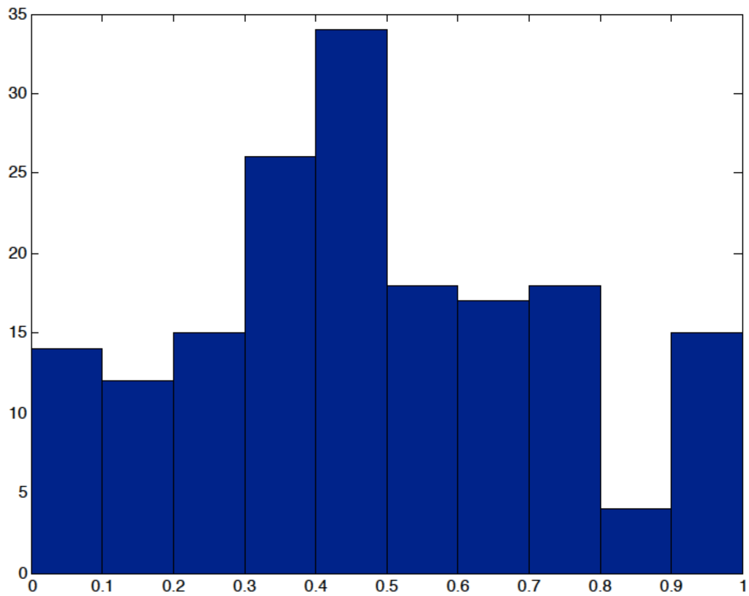

Figure 1: Frequency distribution of subjective probabilities for Economics major, Business occupation (Phase 1 data)

2.3 Phase 2 and Phase 3 data

In order to assess whether beliefs about future labor market outcomes predictive of the actual choices made by the individuals after graduating from college, we collected we collected data on the actual occupational choices and earnings of our sample members several years after all of them completed their BA degrees. These data were collected in two additional phases. We describe each in turn.

In what we refer to as Phase 2 of the DuCMES, we used information obtained from the social network *LinkedIn* in July 2015 by matching . 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 also can update their profile in the *LinkedIn* 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 subsequently gathered for 5 more respondents directly from updated information in the Alumni Directory. Thus, our Phase 2 data collection produced current occupations for 166 of the 173 members of our original sample.

The occupation data obtained from these Phase 2 sources was mapped into of the six occupation classifications used in Phase 1: Science, Health, Business, Government, Education and Law. For example, engineers and software developers were mapped into Science careers; doctors, residents 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. Let $d_{ik}, k = 1, ..., 6$, denote indicator variables for whether individual i's *actual occupation* was k.

In what we refer to as Phase 3 of the DuCMES, we collected additional data on *ex post* outcomes, as well as to update our samples' expectations about careers, in a follow-up survey administered between February and April of 2016.[15] The respondents were contacted via email, *LinkedIn* message and/or text message.[16] A total of 117 individuals – about 68% of the initial sample of 173 individuals – replied to the follow-up survey, and 112 individuals completed the survey. In Table A.6 in Appendix A.3, we compare the characteristics of the individuals who completed the Phase 3 sample with those of the baseline Phase 1 sample. On average, individuals who are followed in the Phase 3 survey have very similar characteristics to the initial Phase 1 sample, including in terms of occupation-specific earnings beliefs and subjective probabilities of choosing each type of occupation. Overall, the comparison presented in Table A.6 suggests that the non-response for the Phase 3 survey is largely ignorable.

The Phase 3 survey collected information on their past and current occupations and their current earnings. Respondents also were asked to update their expectations about what they expect their occupations and earnings to be 10 years after college graduation.[17]

[15] A copy of the questionnaire used in the Phase 3 survey can be found at `public.econ.duke.edu/~vjh3/working_papers/college_major_questionnaire_ph3.pdf`.

[16] All individuals who completed the survey received a coupon for a Duke Basketball Championship T-shirt that could be redeemed through the Duke University Bookstore's website.

[17] 8.3% of the expected earnings elicited in the Phase 3 survey are missing for the 112 individuals who completed this follow-uup 3 survey. For these cases, occupation-specific expected earnings are imputed as the predicted earnings computed from a linear regression of log expected earnings on chosen major and occupation indicators, interaction between major and occupation, individual-specific average log expected earnings in Phase 3 across all occupations, occupation-specific log expected earnings in Phase 1, and an indicator for whether the subjective probability of working in this occupation is equal to zero ($p_i(j^c, k, 3) = 0$).

Consistent with the notations defined above, we let $p_i(j^c, k, 3)$ and $Y_i(j^c, k, 3)$, respectively, denote these occupational choice probabilities and expected earnings elicited in the Phase 3 follow-up survey, for the *chosen major*, j^c.

Finally, we used the Phase 3 data on respondents' current occupation to supplement and adjust the information on chosen occupations collected in Phase 2 as follows. Some 19 individuals declared an occupation in the Phase 3 survey that did not match the occupation imputed using the information obtained from Phase 2 data collected from *LinkedIn*, Duke Alumni database and an internet search. For those cases, we used the occupation that respondents provided in the Phase 3 survey. At the same time, from the Phase 3 survey data we were able to find the occupation of 2 additional individuals. Overall, we ended up with non-missing data on current occupations for a 167 of the 173 original sample members. Unless otherwise indicated, we use these data for this "augmented" Phase 2 sample in all of the tabulations and analysis of chosen occupations presented below.

2.4 Subjective choice probabilities versus actual choices

We next explore the relationship between the data on subjective occupational choice probabilities that we elicited from respondents who completed both the Phase 1 and 3 surveys, conditional on their chosen major and the occupation individuals actually chose, at least as of 4-7 years after they completed their undergraduate degrees. Columns (1) and (2) of Table 4 display the average probabilities for occupations elicited at Phase 1 and the average of the actual chosen occupations obtained at Phase 2. A much greater share have ended up in a Business career than what they predicted at the time they were undergraduates, while smaller shares are seen in several occupations, including Government and Law. (All of these differences are significant at the 1% level). More (fewer) individuals also are pursuing a career in Health (Education) relative to what would be predicted from the subjective probabilities, although the differences are only significant at 10%.

Although the beliefs are, on average, off for some of the occupations, the fourth and fifth columns of Table 4 show that the elicited probabilities do have informational content. Column (4) shows the average elicited probability of working in a career, conditional on actually choosing that career. For example, among those who actually chose a Science career, the average subjective probability of choosing Science was about 35 percent. Column (5) show the average elicited probability of working in a career, conditional on *not* working in that career. Hence, those who *ex post* did not end up in a Science career, on average, thought there was a 14.5 percent chance they would back at Phase 1. 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 a tight association between actual occupational choice and elicited probabilities.

Even thought the *ex ante* and *ex post* shares of individuals in the alternative occupations do not match, they still may be consistent with individuals having rational expectations. There are at least two reasons, other than violation of rational expectations, as to why these shares are not closer. First, while we elicited expectations about outcomes as of 10 years after graduation in Phase 1, in fact, we are measuring their occupations 4 to 7 years after graduation. Thus, it could be the case that students currently working in Business intend to transition into Government or Law in the next few years. Second, it could be that intervening aggregate shocks to the labor market led to differences between the *ex ante* and *ex post* occupational shares. For example, there is evidence that entry into the legal profession was affected by a post-Great Recession negative shock that may have not been fully anticipated.[18]

Column (3) of Table 4 shows beliefs about future expectations that were elicited in the Phase 3 follow-up survey, some 7 years after our Phase 1 elicitation. We expect that many, if not most, of the individuals in our sample are already in their preferred occupations at this stage of their careers. This is not surprising, given that the Phase 3 expectations about careers are very similar to the actual choices obtained in Phase 2. Individuals in Phase 3, on average, reported a much higher probability of working in Business and correspondingly lower probabilities of working in Law or Government, patterns that are consistent with actual occupational choices.

The last four columns of Table 4 show the expected probability of working in each career elicited in Phase 1 and 3, respectively, conditional on currently working and not working in that career based on our Phase 2 data. 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 that were elicited at Phase 1 (column (4). This suggests that 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%).These findings are consistent with a very high cost of switching into these two occupations. Nevertheless, the probabilities are all significantly lower than one, suggesting that, while uncertainty has been reduced, some of the young men in our sample, 4 to 7 years after graduation, still perceive a significant chance of moving to another career in the near future.

[18]As noted in Barton (2015) and Lee (2015), while the number of LSAT takers was increasing prior to the Great Recession, this number peaked in 2009-10 and has fallen by 45% between then and 2014-15.

Table 4: Chosen Occupations and Elicited Beliefs about Occupations

	Phase 1 Beliefs: $p_i(j^c, k, 1)$	Phase 2 Chosen: d_{ik}	Phase 3 Beliefs: $p_i(j^c, k, 3)$	$p_i(j^c, k, 1)$, given: $d_{ik} = 1$	$d_{ik} = 0$	$p_i(j^c, k, 3)$, given: $d_{ik} = 1$	$d_{ik} = 0$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
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

DATA: Columns (1), (2), (4) and (5) are based on 167 individuals for whom we obtained their current occupation from Phase 2, augmented with some Phase 3 data.
Columns (3), (6) and (7) are based on 112 individuals from Phase 1 who completed the Phase 3 follow-up survey.

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 majors of the former students in our sample. The link between all possible major-occupation pairs cannot be assessed with our data, given that a number of pairs were not chosen by our sample. However, we can examine the decision to work in Business, where, for each major, at least one individual chose Business as their career. Table 5 shows estimates of a linear probability model of choosing an occupation in Business. Column (1) controls for the elicited probability of choosing Business ($k = 3$) at Phase 1 conditional on the student's actual major, i.e., $p_i(j^c, 3, 1)$. Conditioning only on this one variable results in an R^2 close to 0.16 and the coefficient on it (0.840) is not statistically different from one, which is what would be expected under rational expectations. Column (2) estimates the differences in choosing a Business occupation by one's chosen chosen major, j^c. Compared to having graduated with a major in the natural Sciences, all other majors have a higher probability of being in a Business occupation as of Phase 2, with economics majors having the highest relative probability (48.5%). However, accounting for one's major has a lower R^2 (0.125) than conditional on the elicited probability at Phase 1 of having a career in Business. Column (3) includes both the Phase 1 elicited probability of Business and one's chosen major. While the coefficient on the elicited 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.

Table 5: Linear probability model of whether Phase 2 occupational choice is Business

	Full Sample			Excluding Seniors		
	(1)	(2)	(3)	(4)	(5)	(6)
$p_i(j^c, 3, 1)$	0.840		0.733	0.659		0.569
	(0.152)		(0.198)	(0.196)		(0.248)
Chosen Major (j^c):						
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)
R^2	0.158	0.125	0.194	0.094	0.082	0.127

DATA: Phase 1 data. Full Sample includes all 167 respondents. The Excluding Seniors sample consists of the 113 respondents who were not seniors.
NOTES: Subjective probability of choosing Business is conditional on their chosen major, j^c. Standard errors in parentheses. All specifications include a constant term.

It is possible that the findings in Table 5 are driven by the fact that they are estimated over the full sample, which include seniors for whom many already had jobs lined up upon graduation.[19] In Columns (4)-(6) of Table 5, we perform the same analysis as in Columns (1)-(3) but remove seniors from the sample. The same patterns emerge: the elicited probability of choosing a career in Business 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.

[19] Recall that Phase 1 was conducted during the 2009 Spring Semester, only 1-2 months before Duke's commencement.

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 *Ex ante* treatment effects and their Means

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 $k = 1$.[20] For any given individual i, the *ex ante* treatment effect of occupation $k \in \{2, 3, 4, 5, 6\}$, conditional on chosen (or intended) major, is simply given by $\Delta Y_{ij^c k\ell} := Y_i(j^c, k, \ell) - Y_i(j^c, 1, \ell)$, where, as before, $Y_i(j^c, k, \ell) := Y_{ij^c k\ell}$ denotes individual i's expected earnings in occupation k given the chosen major, j^c, and expectations we elicited at Phase $\ell, \ell = 1, 3$. In what follows, we suppress the major subscript, j, throughout and, for now, the ℓ subscript for when expectations. Below, we distinguish between treatment effects based on Phase 1 versus Phase 3 elicited data.

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:

$$ATE(k) := E\left(\Delta Y_{ik}\right). \tag{3.1}$$

Note that the parameter, $ATE(k)$, does not incorporate any (expected) differences in direct and opportunity costs across occupations; in fact, such costs may be significant, since some careers, such as Law, typically require an advanced degree. That is, $ATE(k)$ is not an *ex ante* internal rate of return, but rather the expected effect of working in occupation k on

[20] 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. 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 labor supply and spousal earnings.

earnings ten years out relative to that for working in the base occupation.[21] This population parameter is consistently estimated using its sample analog:

$$\widehat{ATE}(k) = N^{-1} \sum_i \Delta Y_{ik}, \tag{3.2}$$

where N is the sample size.

As with the more traditional treatment effects literature, we also are interested in investigating the heterogeneity in the *ex ante* treatment effects by choice of occupation. We define the following mean *ex ante* treatment effect parameter:

$$TT(k) := E\left(\omega_{ik}^{TT} \Delta Y_{ik}\right), \tag{3.3}$$

where $\omega_{ik}^{TT} := p_{ik}/E(p_{ik})$, and $p_{ik} := p_i(j^c, k, 1)$ is the elicited probability from individual i that he would choose occupation k, given his chosen major is j^c. $TT(k)$ is a weighted average *ex ante* treatment effect of occupation k, which "upweights" the *ex ante* treatment effects for the individuals with higher subjective probabilities of choosing occupation k and "downweights" those with lower probabilities. Note that $TT(k)$ will be larger than $ATE(k)$ in (3.1) if individuals expect, on average, to sort into occupations with higher expected returns. A consistent estimator of $TT(k)$ is given by:

$$\widehat{TT}(k) = N^{-1} \sum_i \widehat{\omega}_{ik}^{TT} \Delta Y_{ik}, \tag{3.4}$$

where $\widehat{\omega}_{ik}^{TT} = p_{ik}/(N^{-1} \sum_i p_{ik})$.

The parameter in (3.3) is more directly interpretable under two additional assumptions: (i) individuals form rational expectations about their future occupational choices and (ii) unanticipated occupation-specific aggregate shocks, denoted by α_k, either do not exist ($\alpha_k = 0$) or take the form: $E(d_{ik}) = e^{\alpha_k} E(p_{ik})$.[22] Under these assumptions, it follows that the weights in (3.3), ω_{ik}^{TT}, are also equal to $d_{ik}/E\left(d_{ik}\right)$, where d_{ik} is the indicator for whether i works in occupation k 10 years after graduating so that:

$$TT(k) = E\left(\Delta Y_{ik} | d_{ik} = 1\right), \tag{3.5}$$

is just the *ex ante* treatment effect of occupation k on the treated.

[21] Note that if individuals form rational expectations over their future outcomes, and in the absence of unanticipated aggregated shocks, this parameter coincides with the mean (*ex post*) effect of working in occupation k, relative to Education, on earnings ten years out.

[22] In Section 6 we provide sufficient conditions under which unanticipated shocks on the occupation-specific earnings lead to these types of shocks.

Finally, the *ex ante* treatment on the treated, $TT(k)$ in (3.5), has a natural analogue, namely, the *ex ante* treatment effect on the untreated. And, under analogous assumptions to those given above for $TT(k)$, the *ex ante* treatment effect of occupation k on the *untreated* is given by:

$$\begin{aligned} TUT(k) &= E\left(\Delta Y_{ik} | d_{ik} = 0\right) \\ &= E\left(\omega_{ik}^{TUT} \Delta Y_{ik}\right), \end{aligned} \tag{3.6}$$

where $\omega_{ik}^{TUT} = (1 - d_{ik})/E\left(1 - d_{ik}\right)$. A consistent estimator of $TUT(k)$ is given by

$$\widehat{TUT}(k) = N^{-1} \sum_i \widehat{\omega}_{ik}^{TUT} \Delta Y_{ik}, \tag{3.7}$$

where $\widehat{\omega}_{ik}^{TUT} = (1 - p_{ik})/[N^{-1} \sum_i (1 - p_{ik})]$.

None of the estimators, $\widehat{ATE}(k)$, $\widehat{TT}(k)$ and $\widehat{TUT}(k)$, defined in (3.2), (3.4), and (3.7), respectively, are conditioned on *actual* occupational choices, since the individuals in our sample were students at the time we elicited their expected earnings and probabilities of choosing different occupations. In Subsection 5 below, we compute the *ex ante* treatment effect parameters conditional on the actual choice of occupation, and, in Section 6, we investigate the extent to which the choice of occupation sorts on expected gains in earnings.

Table 6 presents estimates of the three *ex ante* treatment effects of working in particular occupations on earnings 10 years after graduation using the estimators defined above. Recall that in eliciting these earnings expectations conditional on the possibility that respondents were in each occupation. Relative to the Education occupation, 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 91% or more, while those working in a Science or Government occupation expect a much smaller premium of 30.0% to 34.8% ten years after graduation.[23]

Consistent with sorting to occupations with higher expected earnings, the estimated TUTs in Table 6 are lower than the TTs for each occupation. The difference is particularly large (and significant at 1%) in the case of Health occupations, where the expected premium is more than two times smaller for the untreated compared to the treated. However, differences between the TUTs and TTs turn out to be much weaker, and only significant at 10%,

[23]Table A.5 in Appendix A.3 presents estimates of the average *ex ante* treatment effects separately for under-classmen and upper-classmen. While the estimates for all occupations are larger for upper-classmen compared to under-classmen, none of them are significantly different at standard statistical levels.

Table 6: *Ex Ante* Treatment Effects by Occupation (Earnings in 2009 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)	

DATA: Sample who completed Phase 1 survey ($N = 167$).
NOTES: Standard errors are reported in parentheses. TT is significantly different from TUT for Science (p-value = 0.051), and Health (p- value = 3.10^{-4}).

for Science careers, with the untreated expected to earn only 69% as much as the treated, and are negligibly small for Government careers.

Another way of assessing the potential importance of selection in our *ex ante* data is to construct the analogues of *ex post* (realized) earnings, unadjusted and adjusted for the selectivity of choosing a particular occupation. Unadjusted *ex post* earnings are just the observed earnings of individuals working in a particular occupation, while selection-adjusted ones are constructed using some selection-correction estimation procedure. Using our expectations data, we can produce *ex ante* analogue of both measures. Define the "selected *ex ante* earnings" for occupation k to be $SE(k) = E\left(\omega_{ik}^{TT} Y_{ik}\right)$ for which a consistent estimator is:

$$\widehat{SE}(k) := N^{-1} \sum_{i} \widehat{\omega}_{ik}^{TT} Y_{ik}). \tag{3.8}$$

As with the $TT(k)$ effect in (3.3), the $SE(k)$ and its estimator $\widehat{SE}(k)$ upweights the expected earnings by individuals' (relative) *ex ante* probabilities of being in occupation k, thereby mimicking the unadjusted ex post earnings of those who *ex post* chose to work in occupation k. The corresponding estimator for the selected *ex ante* earnings difference between occupation

k and Education ($k = 1$) is given by:

$$\Delta\widehat{SE}(k) := \widehat{SE}(k) - \widehat{SE}(1) = N^{-1}\sum_i \big[\widehat{\omega}_{ik}^{TT} Y_{ik} - \widehat{\omega}_{ik}^{TT} Y_{i1}\big]. \quad (3.9)$$

Note that $\Delta\widehat{SE}(k)$ is similar to $TT(k)$ in (3.4), except that it weights Y_{ik} and Y_{i1} differently, with $\Delta\widehat{SE}(k)$ weighing each of them by the respective subjective probabilities that individual i would choose to work in these two occupations. Again, this mimics the calculation of an unadjusted (for selection) difference in observed *ex post* earnings for two occupations. Finally, the adjusted for selection estimator for the differences in *ex ante* earnings between occupations k and 1 is just the average *ex ante* treatment effect of occupation k, $ATE(k)$, in (3.1).

We present, in Panel A of Table 7, estimates of the unadjusted and selection-adjusted *ex ante* earnings using Phase 1 data $(Y_{ik\ell}, p_{ik\ell}, k = 1, ..., K, \ell = 1)$. Row (1) displays estimates for $SE(k)$ and, as a point of reference, the simple unweighted means of the *ex ante* earnings for occupation k, $\overline{Y}(k) := N^{-1}\sum_i Y_{ik}$, in row (2). The fact that the means of selected *ex ante* earnings are all greater than the $\overline{Y}(k)$s is indicative, although not proof, of positive sorting on expected earnings.

In rows (3) and (4), we display the occupation-specific estimates of the selected *ex ante* earnings differentials, $\Delta\widehat{SE}(k)$, and the selected-adjusted differentials, $\widehat{ATE}(k)$. The nature of the selection and sorting in occupations based on *ex ante* data is the relationship between $\Delta\widehat{SE}(k)$ and $\widehat{ATE}(k)$. In rows (5) and (6) we show the simple difference between the two, which we label as the "Selection amount," and share of that difference with respect to the selection-unadjusted estimates, $\Delta\widehat{SE}(k)$, which we label as the "Selection share." As one can see, the selection amounts for all occupations are positive, which is consistent with positive sorting on *ex ante* earnings in *ex ante* occupational choice. Put differently, our sample members, on average, suggest that they expect to sort to occupations in which they expect to earn more, for whatever reason. Furthermore, the selection share estimates that this selection is much larger for the Health occupation, least important for the Business occupation, with the other occupations somewhere in between. These results are consistent with Table 6 in that the gap between the TTs and TUT is especially large in Health and small in Business.

Panel B of Table 7 presents the same statistics as in Panel A, using data from Phase 3 on respondents' expectations about the expected earnings and probabilities of being for each occupation k 10 years after graduation $(Y_{ik\ell}, p_{ik\ell}, k = 1, ..., K, \ell = 3)$. Use these more recent elicited expectations allow us to assess the role selection plays after Educational decisions are essentially finalized. Comparing row (1) across the two Panels, there is a sizable increase in respondents' selected *ex ante* earnings expectations for Business occupations (\$176,393 vs \$294,728), a noticeable decline for careers in Government (\$109,419 vs \$76,514) and Law (\$190,072 vs \$138,062), and almost no change for those in Education and Health. Comparing the Selection Amounts and Shares in rows (5) and (6) across the two Panels, we see an extremely large increase in selection for earnings in Business careers, smaller increases in Health and Science careers and actual declines in selection for careers in Government and the Law. These changes may reflect our respondents learning more about their prospects in these careers over the 7 years between the two surveys as well as changes that may have occurred to the relative demands and wages across different occupations.[24] At the same time, selected *ex ante* earnings in Health remain unchanged across the surveys even though average *ex ante* earnings fall. The rationale for this likely comes from medical school. In Phase 1, respondents associated working in Health with going to medical school. At the time of the Phase 3 survey, however, individuals have already committed to going or not going to medical school.

3.2 Distributions of *ex ante* treatment effects

Our elicited expectations data not only allow us to estimate the means of the *ex ante* treatment effects defined in the previous section, but also estimate their distributions. We first consider the estimation of the unconditional distribution of the *ex ante* treatment effects and then turn to the conditional distributions of the *ex ante* treatment effects on the treated and untreated subpopulations. All of the *ex ante* treatment effects are computed for students' chosen college majors, j^c, using data from Phase 1 ($\ell = 1$).

The density of the *unconditional* distribution of the *ex ante* treatment effects for occupation k, i.e., ΔY_{ik}, in 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 $\widehat{f}_{TE,k}(\cdot)$.

Next consider a weighted version of $f_{TE,k}(\cdot)$, where the weights are functions of the elicited probabilities of choosing the various occupations. This density function is defined as:

$$f_{TE,k}^{Treated}(u) = \omega_{ik}^{TT}(u) \cdot f_{TE,k}(u), \tag{3.10}$$

[24]As noted above, there appears to ahve been a sizable decline in the demand for Lawyers over this period.

Table 7: Relative Importance of Selection in *Ex Ante* Earnings Returns: Phases 1 and 3 (Earnings in 2009 dollars)

	Occupation:					
	Science	Health	Business	Government	Law	Education
Panel A: Phase 1						
(1) $\widehat{SE}(k)$ (Selected *ex ante* earnings)	102,699	188,189	176,393	109,419	190,072	72,725
(2) $\overline{Y}(k)$ (Aver. *ex ante* earnings)	96,793	142,538	164,006	100,348	163,223	74,473
(3) $\Delta\widehat{SE}(k)$ (Selected earnings difference from Educ.)	29,974	115,463	103,668	36,694	117,346	
(4) $\widehat{ATE}(k)$ (Ave. *ex ante* expected earnings effect)	22,320	68,065	89,533	25,875	88,750	
(5) $\Delta\widehat{SE}(k) - \widehat{ATE}(k)$ (Selection amount)	7,654	47,399	14,135	10,819	28,596	
(6) $\frac{\Delta\widehat{SE}(k)-\widehat{ATE}(k)}{\Delta\widehat{SE}(k)}$ (Selection share)	25.5%	41.1%	13.6%	29.5%	24.4%	
Panel B: Phase 3						
(1) $\widehat{SE}(k)$ (Selected *ex ante* earnings)	137,631	183,852	294,728	76,514	138,062	64,380
(2) $\overline{Y}(k)$ (Aver. *ex ante* earnings)	123,301	125,557	211,147	81,379	137,328	71,333
(3) $\Delta\widehat{SE}(k)$ (Selected earnings difference from Educ.)	73,251	119,472	230,347	12,134	73,681	
(4) $\widehat{ATE}(k)$ (Ave. *ex ante* expected earnings effect)	51,968	54,224	139,814	10,046	65,995	
(5) $\Delta\widehat{SE}(k) - \widehat{ATE}(k)$ (Selection amount)	21,283	65,248	90,534	2,087	7,687	
(6) $\frac{\Delta\widehat{SE}(k)-\widehat{ATE}(k)}{\Delta\widehat{SE}(k)}$ (Selection share)	29.1%	54.6%	39.3%	17.2%	10.4%	

DATA: Data from Phase 1 ($N = 167$) and Phase 3 ($N = 113$).

where $\omega_{ik}^{TT}(u) := g(u)/E(p_{ik})$ and $g(u) = E(p_{ik}|\Delta Y_{ik} = u)$. As discussed in the previous section for the average treatment effect parameters, 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. The following plug-in estimator:

$$\widehat{f}_{TE,k}^{Treated}(u) = \widehat{\omega}_{ik}^{TUT}(u) \cdot \widehat{f}_{TE,k}(u), \tag{3.11}$$

is a consistent estimator of $f_{TE,k}^{Treated}(u)$, where $\widehat{\omega}_{ik}^{TUT}(u) = \widehat{g}(u)/(N^{-1}\sum_i p_{ik})$ and $\widehat{g}(u)$ is the Nadaraya-Watson estimator of the nonparametric regression $g(u)$. Finally, the distribution of the *ex ante* treatment effects on the untreated, $\widehat{f}_{TE,k}^{Untreated}(u)$, can be estimated in a similar fashion by replacing p_{ik} with $1 - p_{ik}$ in Equation (3.11).

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.[25] (The distributions of the *ex ante* treatment effects for Science and Law are displayed in Figures A.1 and A.2 in Appendix A.4.) Each of 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 selection on expected returns throughout the distribution. 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.[26] 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.

3.3 Heterogeneity in *ex ante* treatment effects across majors

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

[25] 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.

[26] While, for Business, the average *ex ante* treatment on the treated is not significantly different from the average *ex ante* treatment on the untreated, one can indeed reject at 5% the equality of the first quartiles of these two distributions (p-value of 0.015).

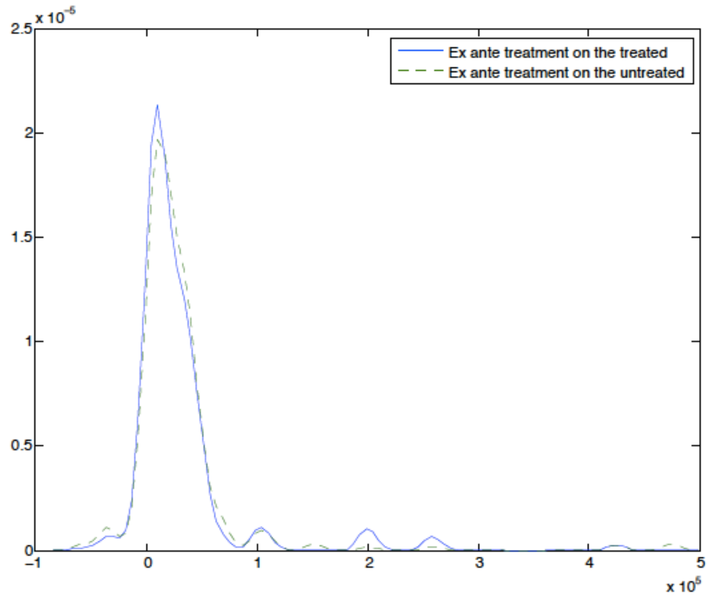

Figure 2: Densities of *Ex Ante* Treatment Effects on the Treated & Untreated: Government

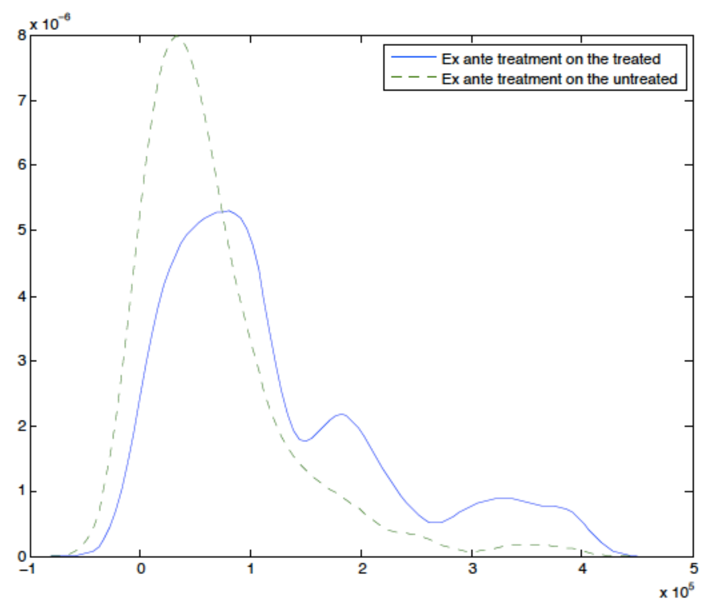

Figure 3: Densities of *Ex Ante* Treatment Effects on the Treated & Untreated: Health

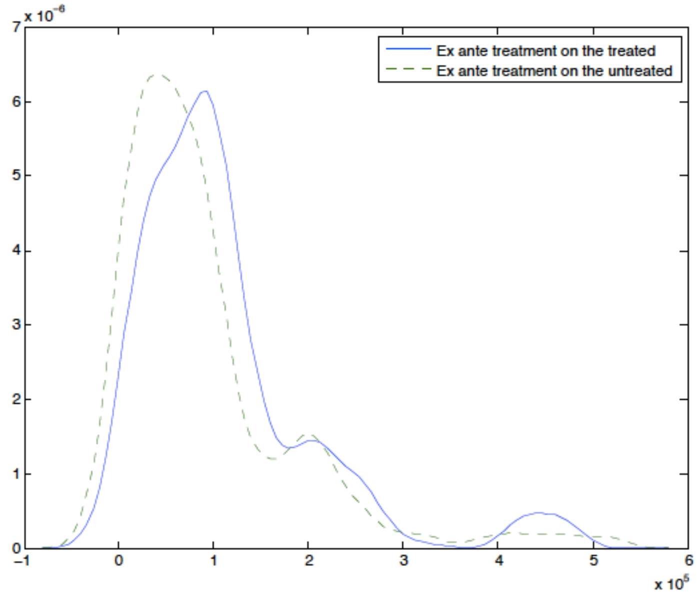

Figure 4: Densities of *Ex Ante* Treatment Effects on the Treated & Untreated: Business

on each of the majors that respondents chose ($j = j^c$). Let m_{ij} denote an indicator variable for whether i chose major j. Then the estimators for the average *ex ante* treatment effect, *ex ante* treatment on the treated and treatment on the untreated for each chosen major, j^c, are given by:

$$\widehat{ATE}(k|j^c) := \frac{\sum_i m_{ij}[Y_i(j,k,1) - Y_i(j,1,1)]}{\sum_i m_{ij}}, \tag{3.12}$$

$$\widehat{TT}(k|j^c) := \frac{\sum_i m_{ij} p_i(j,k,1)[Y_i(j,k,1) - Y_i(j,1,1)]}{\sum_i m_{ij} p_i(j,k,1)}, \tag{3.13}$$

$$\widehat{TUT}(k|j^c) := \frac{\sum_i m_{ij}(1 - p_i(j,k,1))[Y_i(j,k,1) - Y_i(j,1,1)]}{\sum_i m_{ij}(1 - p_i(j,k,1))}.[27] \tag{3.14}$$

In Table 8, we present the estimates of these *ex ante* treatment effects conditional on respondents' chosen major, j^c, using data from Phase 1. 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. Examining some of the other average *ex ante* returns, economics majors have the highest premium for Business occupations, while engineering and natural

[27]Given that we also elicit the subjective expectations for all counterfactual majors, we could estimate the *ex ante* treatment effects for those who did *not* choose major j by replacing m_{ij} with $1 - m_{ij}$ in the estimators above.

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 also are 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 6). 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 selection into occupations with the highest expected returns, although 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 A.4 in the Appendix 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.[28]

[28] 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 (Earnings in 2009 dollars)

		Chosen Major (j^c):					
Occupation:	Treat. Eff.	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)

DATA: Data from Phase 1 ($N = 167$).
NOTE: Standard errors are reported in parentheses.

4 Evolution of beliefs

In this section we discuss how our elicited data on expected earnings can be used to characterize how beliefs evolve over time. We first document the evolution of the*ex ante* treatment effects of different occupations between Phase 1 and Phase 3 surveys. Then, we use the elicited data of under- and upper-classmen to characterize learning about the average incomes in occupations while in college. In Appendix A.2 we show how one can model the evolution of respondents' beliefs about their abilities in alternative careers using our elicited beliefs data along with additional parametric assumptions.

4.1 Evolution of *ex ante* treatment effects from college to early careers

In this section, we 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 9 shows estimates of the *ex ante* treatment effects on the treated (TT), treatment effects on the untreated (TUT), and the average treatment effects (ATE), computed using the beliefs about expected earnings and subjective probabilities of choosing each occupation that were elicited in the DuCMES Phase 3 survey, seven years after those elicited in Phase 1. This table replicates the results in Table 6, using the updated beliefs of respondents elicited in Phase 3. Recall that each survey asked to give their current beliefs of what they expect to be earnings 10 years after completing their undergraduate degree. At enrollment in our study in 2009, students in our sample came from all four classes (freshman, sophomores, juniors and seniors). Thus, at the time of the Phase 3 survey in 2016, these 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.

Several comments are in order about the treatment effects in Table 9 and comparison to those in Table 6 based on expectations elicited in Phase 1. First, the ATEs are substantially higher in Science and Business relative to the Phase 1 beliefs. However, the estimated TTs and TUTs point to different patterns in terms of selection for both occupations. While for Science occupations, both the estimated TT and TUT increase from Phase 1, more than doubling between the two surveys, the increase in the ATEs for Business is primarily driven by the increase in the TTs, resulting in a large and significant discrepancy between the TTs and TUTs for Business. Second, the ATEs decrease for some of the occupations, namely Government, Law, as well as, to a lesser extent, Health. Interestingly, in the case of Law, both the TTs and TUTs fall, but the former fall is much larger such that the TTs

Table 9: *Ex ante* treatment effects for each possible occupation (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)

DATA: Data from Phase 3 ($N = 112$). Recall that Phase 3 is conducted 7 years after the Phase 1 survey.
NOTES: Standard errors in parentheses. TT is significantly different from TUT for Health (p-value = 0.001), Business (p-value = 0.000) and Government (p-value = 0.036).

becomes actually lower than the TUTs. A likely explanation for these changes is that those who perceived a high return to Law initially saw large changes in their returns to Business, shifting them from Law to Business.[29]

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.

4.2 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 in college, students in our sample are enrolled in different years of college. In the following we use a synthetic cohort approach

[29]These differences between Phase 1 and Phase 3 *ex ante* treatment effects are statistically significant at the 5% level, with the exception of the ATEs for Health and Law, the latter being marginally significant at 10% only.

and examine how students update their beliefs by comparing the distributions of expected incomes for under-classmen with those of upper-classmen.

Students are learning about both their own abilities as well as about the market. We focus this section on learning about the market with Appendix A.2 showing stronger assumptions where both learning about the market and about own abilities can be disentangled. To address learning about the market, we use students' phase 1 beliefs of what the "average" Duke [male] undergraduate would earn in different major-career combinations 10 years after graduation. In particular students were asked:

> "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?"

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. In Table 10 the differences in the variance of log-expected incomes for the average Duke student between under- and upper-classmen for each possible major-occupation pair using Phase 1 data. 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). Finally, Table 10 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.

for all pairs of occupation and major,

5 *Ex ante* treatment effects conditional on actual occupational choices

Finally, we use our Phase 2 and 3 follow-up data on the *actual* choices of occupations of our sample members to investigate how the *ex ante* treatment effects for *ex ante* earnings of working in particular occupations vary with the occupations they *actually chose*.[30]

[30] As our discussion of the findings in Table 4 found in Section 2.1 indicates, the occupations we record in Phases 2 and 3 may not be the final occupation for all of our respondents; nonetheless, it appears that for many, they have settled on their careers by this point in their lives.

Table 10: Differences in variances of the log of elicited expected incomes for the average Duke student between upper- and under-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***

DATA: Phase 1 data ($N = 173$).
NOTES: Expected earnings were elicited for each possible major-occupation pair at Phase 1, regardless of the respondents' chosen or intended major. "All" indicates average across majors (rows) and occupations (columns). *, **, and *** indicate statistical significance of differences at the 10%, 5%, and 1% level, respectively.

Table 11 below reports the *ex ante* treatment effects on the treated (TT), as well as the *ex ante* treatment effects on the untreated (TUT) for all occupations relative to Education, conditional on respondents' chosen occupation. Note that the estimated *ex ante* treatment effects are based on expected earnings elicited in Phase 1, while respondents' chosen occupations were determined using Phase 2 and 3 data. Comparing the version of these two treatment effects in Table 11 with those in Table 6 that were conditioned on respondents' responses to possible occupations elicited in Phase 1 we find similar selection patterns. Consistent with positive sorting on expected earnings, individuals who end up working in Science, Health, Business, and Government occupations anticipate on average higher earnings premia for those occupations relative to those who work in another occupation. Estimates of TT are significantly different from those for TUT at the 1% level for Health careers and marginally significant(at 10% level) for Science.

Selection effects are highest for Health careers, echoing our earlier findings based on possible, rather than actual, choices, while the estimates of TT and TUT for those who chose careers in Law are similar in magnitude, even though the corresponding estimates in Table 6 are quite different, where we asked respondents to condition their responses on possible occuaptions. Below, in Section 4.1 we argue that a possible explanation for this pattern is that individuals with large *ex ante* returns to Law saw large increases in their expected returns to Business after graduating from college, shifting some of them from Law to Business.

Figures 5 and 6 below display the densities of the *ex ante* treatment effects conditional on chosen occupations for Health and Business occupations, respectively. We focus on these two occupations since these are the two most frequently chosen in our sample. Comparing the distributions of *ex ante* returns for those who are observed to choose a Health occupation

Table 11: *Ex Ante* Treatment Effects, conditional on Actual Occupation Choices

Actual Occupation	*TT*	*TUT*
Science	34,808	19,290
	(7,612)	(3,269)
Health	122,570	52,358
	(14,222)	(7,323)
Business	90,726	84,021
	(11,445)	(10,086)
Government	37,111	23,758
	(16,673)	(3,979)
Law	88,667	89,474
	(30,611)	(9,616)

DATA: Estimation based on the subsample of the 167 respondents from Phase 1 that had information on their chosen occupations from Phase 2 and 3 data. See Section 2.3 for explanation of the construction of chosen occupations.
NOTES: The expected earnings used to define the *TT* and *TUT* effects were elicited in Phase 1, while actual occupations are determined using Phase 2 and 3 data.
Standard errors are reported in parentheses.
TT is significantly different from *TUT* for Science (p-value = 0.061) and Health (p-value = 10^{-5}).

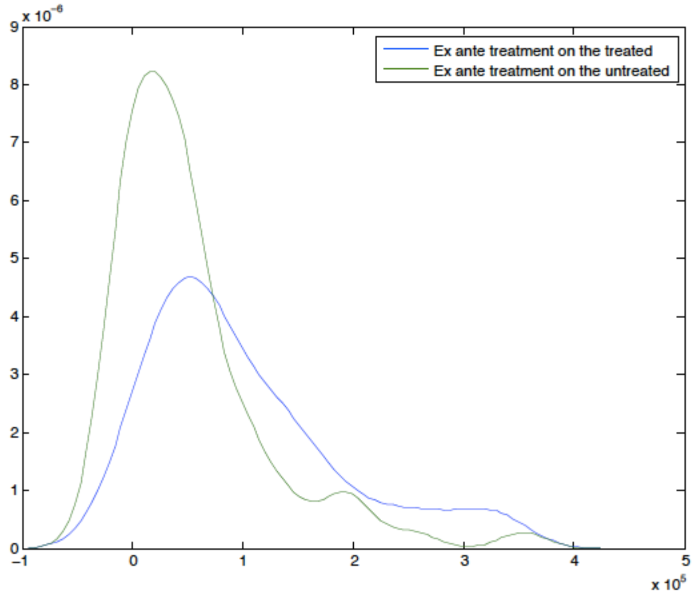

Figure 5: Densities of *Ex Ante* Treatment Effects on Treated & Untreated, conditional on Chosen Occupation: Health

in Figure 5 with the conditional distributions based on respondents' possible choice of Health (Figure 3) reveals that both sets of distributions are very similar. For the Health 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. While not as similar as for Health occupations, the distributions of *ex ante* treatment effects for Business condition on possible versus chosen occupations (Figures 6 and 4) point to a qualitatively similar pattern in terms of selection.

6 Assessing the returns to expected earnings in occupational choice

The findings in the preceding sections all indicate positive sorting of occupational choice on expected earnings. In this section, we go a step further and explicitly model the choice of occupations – both on an *ex ante* and *ex post* basis – as a function of the elicited *ex ante* earnings beliefs of our respondents. We use the data from the three phases of our study to progressively account for the possibility of non-pecuniary occupation attributes, that their influence on occupational choice varies across individuals, and direct returns to majors in order to assess the robustness of the importance of earnings expectations on the choice of

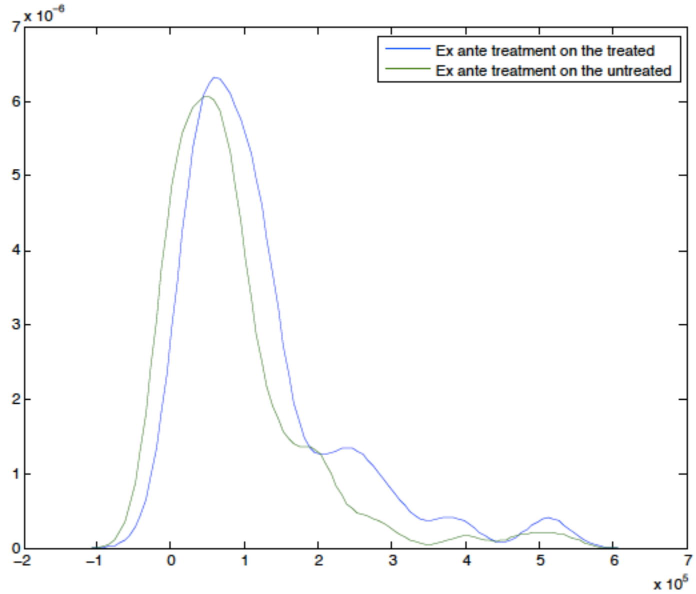

Figure 6: Densities of *Ex Ante* Treatment Effectson Treated & Untreated, conditional on Chosen Occupation: Business

occupations.

We use the following simple framework to model occupational choice. Assume individuals choose their occupations to maximize their expected utility. Conditional on major j, i chooses among K mutually exclusive occupations. Let $d_{ijk} = 1$ if k was chosen and one's major is j and zero otherwise. The values for $d_{ijk}, k = 1, ..., K$, for individual i and major j are chosen to satisfy:

$$\max_{(d_{ijk})_k} \sum_{k=1}^{K} d_{ijk}\big(u_{ijk} + \varepsilon_{ijk}\big), \tag{6.1}$$

where u_{ijk} denotes the (expected) utility that is observable by the researcher up to a vector of parameters and ε_{ijk} is unobserved to the researcher and assumed to be drawn from a standard Type 1 extreme value distribution. We further assume that u_{ijk} is a function of log expected occupation-specific earnings,[31] $y_{ijk} := \ln\left[Y_i(j, k, \ell)\right]$, i.e.,

$$u_{ijk} = \alpha_k + \beta y_{ijk} \tag{6.2}$$

[31] 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 (on the actual choices of occupations) better. See also Arcidiacono (2004, 2005) who uses a similar specification of the expected utility of future labor market outcomes.

where we allow for occupation-specific payoffs, α_k, normalize α_1 to zero and suppress, for now the ℓ subscript. The specification in (6.1) and (6.2) constitute our basic model of occupational choice.

We first use the above specification to model the *observed* choices of occupations, i.e., *ex post* occupations, that we measured in Phase 2 to estimate a conditional logit where we substitute in for y_{ijk} with the Phase 1 beliefs on earnings. Estimates for the β parameter in (6.2) are displayed in Column (1) of Table 12, both for the full sample of all respondents and the sample excluding seniors. We estimate the model without seniors as some seniors may already have jobs lined up at the time of the survey. Finally, we use the expected earnings elicited for our respondents' chosen major, j^c. For both samples, the estimates in column (1) show a significant positive relationship between expected future earnings and respondents' actual choice of occupations,[32] a result that provides further evidence of positive sorting across occupations on expected earnings returns. This finding would be predicted by a Roy model of occupational choice.

We next examine the relationship between respondents' earnings beliefs elicited in Phase 1 and their Phase 1 beliefs about the probabilities of working in each of the occupations $[p_i(j,k,1)]$. These elicited probabilities are our *ex ante* measures of occupational choice. In Phase 1, some of what is unobserved, i.e., the ε_{ijk}s to the researcher is actually known by individual i at that time, but some not known to either i or the researcher back at Phase 1 when they form their $[p_i(j,k,1)]$s and their expected earnings in each occupation. It is the lack of knowledge about this latter part of ε_{ijk} that make individuals uncertain about which occupation is best for them and presumably why they do not just report ones and zeros for the $[p_i(j,k,1)]$s.

To acknowledge this source of individuals' uncertainty, let

$$\varepsilon_{ijk} = \phi_{ijk} + \zeta_{ijk}, \tag{6.3}$$

$k = 1, ..., K$, where the ϕ_{ijk}s are is known to i at Phase 1 but the ζ_{ijk}s are unknown. Assuming that ζ_{ijk} has a standard Type 1 extreme value distribution, we can invert the self-reported

[32]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.

probabilities to obtain:[33]

$$\begin{aligned} \ln\left[p_i(j,k,1)\right] - \ln\left[p_i(j,1,1)\right] &= u_{ijk} - u_{ij1} + \phi_{ijk} - \phi_{ijk1} \\ &= \alpha_k + \beta \Delta y_{ijk} + \Delta \phi_{ijk} \end{aligned} \tag{6.4}$$

for $k = 2, ..., K$, and where, as before, Δ, the differencing operator, is taken with respect to the baseline occupation $k = 1$ (education).

Given respondents' elicited $p_i(j,k,1)$s for all occupations obtained in Phase 1, we can estimate β by treating (6.4) as a regression function, conditional on the individual's chosen major, j^c. We deal with the zero self-reported probabilities by replacing them by an arbitrarily small number, as proposed by Blass, Lach, and Manski (2010). We then estimate the flow utility parameters using a least absolute deviation (LAD) estimator.[34]

Results of the LAD estimation of (6.4) are given in Column (2) of Table 12. The estimate of β for the full sample is 1.371 and is (statistically) and the estimate excluding seniors from the sample is almost the same value, 1.337. Moreover, these estimates, based on a regression specification using elicited occupational choice probabilities at Phase 1 are almost identical – and within the precision of the estimates – to the estimates of β obtained using actual occupational choices measured 7 years later and estimated with conditional logit methods!

Moreover, our elicited occupational choice probabilities allows one to control for major-occupation dummies even with our relatively small sample. That is, we can replace the α_ks with $\alpha_{j^c k}$s in the regression specification in (6.4), otherwise using the same data as used for column (2). These new estimates are presented in column (3) of Table 12. Adding the major-occupation interactions reduce the magnitudes of the estimated βs, especially those for the sample that excludes seniors from the sample, where the estimate declines by 40%. Nonetheless, both estimates of β are positive and fairly to very precisely estimated, providing further evidence of the robustness of the evidence for positive sorting on expected earnings across occupations.

Finally, using the occupational choice probabilities elicited at Phase 1 not only for respondents' chosen major, j^c, but also for their counterfactual majors, allows us to include not only major-occupation dummies, α_{jk}s, but also individual-occupation specific fixed effects,[35] θ_{ik}s in the regression function (6.4), which are eliminated by within-transformation.

[33] To be fully consistent with the generic model, we would also need to assume that the sum of ϕ_{ijk} and ζ_{ijk} follows a Type 1 extreme value distribution. See Cardell (1997) for possible distributions of ϕ_{ijk} such that $\phi_{ijk} + \zeta_{ijk}$ follows a Type 1 extreme value distribution.

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

[35] See Wiswall and Zafar (2016b) who provide evidence from a sample of NYU students that preferences for job attributes are highly heterogeneous across individuals.

Table 12: Estimates of returns to (log of) expected earnings in occupational choice

Occupations:	*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)
Other variables included:				
Occupation dummies	Y	Y	N	N
Major × Occupation dummies	N	N	Y	Y
Individual × Occupation fixed effects	N	N	N	Y

DATA: *Full Sample* includes 167 individuals while the *Excluding Seniors* sample contains 113. Major-occupation-specific expected earnings and occupational choice probabilities are from Phase 1 and actual occupational choices are from Phase 2.

NOTES: All 4 columns use expected earnings elicited in Phase 1. Column (1) models chosen occupations, conditional on chosen majors, j^c, with a conditional logit. Columns (2) – (4) use elicited occupational choice probabilities to estimates regressions of the form given in (6.4). Columns (2) and (3) condition choices of respondents' chosen major, j^c. Column (4) uses data on respondents elicitations of expected earnings and occupational choice probabilities for each possible major-occupation pair, providing 6 times the number of observations in the sample.

Standard errors in parentheses.

For specification (4), standard errors are clustered at the individual × occupation level.

The results for this final specification are reported in Column (4) of Table 12, again using occupational choice probabilities and expected earnings elicited in Phase 1. (The standard errors were clustered at the individual-occupation level.) Note that the multiple observations per respondent help to produce more precise estimates of β, even accounting for the additional controls in these regressions. And, while the estimates of β fall slightly, they still are consistent with positive sorting on expected earnings across occupations.

To quantify the responsiveness of subjective occupational choice probabilities to expected earnings, we calculate the percentage change in the probability of choosing an occupation given a percentage change in expected earnings using the estimmates in column (4) in Table 12. These elasticities, denoted by e_{ijk}, (Train, 2003) are give by:

$$e_{ijk} = [1 - p_i(j, k, 1)]\widehat{\beta}, \tag{6.5}$$

for each individual i and major-occupation pair and where $\widehat{\beta}$ denotes the LAD estimate of β ($\widehat{\beta} = 0.953$ for our preferred specification). Note that this formula only applies for the intensive margin, that is for variation in the subjective probability $p_i(j, k, 1)$ over the strictly bounded between 0 and 1. Hence, we estimate this elasticity only with the data on individuals who provided non-zero $p_i(j, k, 1)$'s. 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. Conditioned on individuals' chosen major, j^c, the average elasticity for occupation k using:

$$\hat{e_k} := N^{-1} \sum_i (1 - p_{ik})\widehat{\beta}. \tag{6.6}$$

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, Fougere, and Maurel, 2012; Long, Goldhaber, and Huntington-Klein, 2015; Wiswall and Zafar, 2015; and Altonji, Arcidiacono, and Maurel, 2016, for a survey).

Thus far, our analysis has made use of the beliefs respondents have for what their earnings and occupations would be 10 years after graduation. As we have already discussed, the beliefs we elicited in Phase 3, conducted in 2016, give us another source of data to assess the importance of sorting on expected earnings for occupational choice. In Table 13, we repeat the *ex ante* occupational choice analyses found in Table 12 with these more recent beliefs. Recall that this follow-up elicitation also asked about expected outcomes for 10 years after graduation. We consider three different specifications, controlling for occupation dummies [column (1)], controlling for major-occupation dummies [column (2)] and one that adds dummy variable indicating whether the particular occupation is the respondent's actual occupation as 2016 [column (3)].

For all three of these specifications, 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 12 provides evidence that beliefs about future choice of occupation tend to be more tightly associated with expected earnings gathered in the Phase 3 follow-up survey than in the initial Phase 1 survey that was collected when 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 decline in the magnitude of the return to expected earnings for occupational choice is consistent with the existence of costs to switching between occupations.

Finally, the fact that we elicited beliefs at two points in time makes it possible to estimate the association between changes in subjective probabilities of choosing particular occupations

Table 13: Estimates of returns to (log of) expected earnings in occupational choice (Phase 3 data)

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

Data: The data is for the sample of 112 individuals who responded to the Phase 3 survey.
Standard errors in parentheses.

Table 14: 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)
Other variables included:			
Current Occupation	N	Y	Y
Major × Occupation	N	N	Y

Notes: This estimation sample includes data on 112 individuals who responded to the Phase 3 survey.
Standard errors in parentheses.

and changes in the occupation-specific expected earnings. Table 14 reports the LAD estimation results which correspond to the specification in (6.4), but here differencing between Phase 3 and Phase 1 beliefs. 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 13 (Column 3, 0.914) and in Table 12 [0.953 in column (4)]. 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.

6.1 Expected versus actual earnings

We conclude this section by examining the relationship between the *actual* earnings of the respondents in our study that were collected in our Phase 3 follow-up survey conducted in 2016 and the *expected* earnings elicited in the Phase 1 survey conducted in 2009 when the individuals were still enrolled in college. To do so, we use data on 81 individuals who reported having positive *current* annual earnings in the Phase 3 survey.[36] We are in particular interested in testing whether earnings beliefs are predictive of future labor market outcomes, beyond their effects documented earlier on occupational choice.

In Table 15, we report the estimation results from a linear regression of log (actual) earnings on log expected earnings in chosen occupation. Column (1) displays the estimation results when we restrict the sample to the individuals who work in business and control for chosen major, j^c, in the spirit of the analysis conducted earlier of the relationship between subjective probability and decision to work in business (Subsection 2.4.1, Table 5). While the estimates are imprecise as a result of the small sample size, the estimated elasticity (0.64) is positive, sizable and significant at the 5% level.

Even though small cell sizes prevent us from repeating this analysis for the other occupations, we can nonetheless use the (expected and actual) earnings data for all occupations and control for chosen major as well as chosen occupation. The corresponding results are reported in Column (2). The estimated elasticity (0.423) is smaller, but remains positive and significant at the 1% level.[37] Taken together, these results suggest that earnings beliefs matter beyond their effect on the choice of occupation.

7 The role of nonpecuniary factors

Given our findings of sorting across occupations on expected returns, or treatment effects, in both expected and actual choices, it also is useful to assess the role that non-pecuniary benefits may play in occupational choice. To address this, we use the measures of expected earnings associated with all occupations that we elicited from our sample to estimate how much income individuals expect to give up as a result of not choosing the occupation that

[36] Some 30 individuals out of the 112 individuals who completed the Phase 3 survey indicated that they did not have a current job and, thus, were not asked about their current annual earnings. More than 80% of those individuals were medical interns or residents, who did not consider these positions as jobs, or were enrolled in a MBA program at the time of the survey.

[37] It is interesting to compare these results 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 an estimated elasticity of 0.167 without controlling for majors or occupations. They show that beliefs are much more predictive of the actual earnings among females.

Table 15: Relationship between actual and expected earnings in chosen occupation

	Log Actual Earnings	
	Business Only	All Occupations
	(1)	(2)
Log Expected Earnings	0.640	0.423
	(0.257)	(0.149)
Other variables included:		
Chosen Major	Y	Y
Chosen Occupation	-	Y
R^2	0.396	0.521

DATA: Column (1) (Column (2)) is based on 37 (81) individuals who reported they had a current job and provided their current annual earnings for that job.
NOTES: Standard errors in parentheses. All specifications include a constant term.

maximizes their income. These measures of willingness-to-pay for non-earnings-maximizing choices provide direct evidence on the role played by other, in particular non-monetary, factors in the choice of occupation, are directly identified from the data, and do not require any additional assumptions on exactly what non-monetary factors affect the choice of one's occupation.[38]

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. Some individuals may report high expected earnings in health because if they were in a health occupation ten years after graduation then that would mean they would have gone to medical school. But the Phase 3 data likely does not suffer from this issue. Individuals have either completed their education or will do so soon. By using expectations from the Phase 3 data, we get measures of lost income that are most likely not contaminated by tuition payments, eliciting what individuals would expect to make without changing their educational decisions.

In Table 16 we display estimates on the earnings lost due to individuals making their expected occupational choices based on factors other than expected earnings. Column (1) of the table presents estimates of the mean, median and 1st and 3rd quartiles for the distribution of the expected earnings for that occupation in which sample members expected to earn the

[38]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. While their approach does not require direct measures of subjective expectations about future returns and does not require exclusion restrictions, does rely on stronger assumptions concerning the non-pecuniary factors. See also Eisenhauer, Heckman, and Vytlacil (2015), who use exclusion restrictions between monetary returns and non-pecuniary factors to separately identify these two components.

most, i.e.,

$$Y_i^{max}(j^c, 3) := \max\{Y_i(j^c, 1, 3), Y_i(j^c, 2, 3), ..., Y_i(j^c, 6, 3)\}. \tag{7.1}$$

Column (2) characterizes the distribution of expected earnings elicited in Phase 3, $\overline{Y}_i(j^c, 3)$, where elicited expected earnings in each occupation are weighted by the elicited probabilities that the individual would work in each of these occupations, i.e.,

$$\overline{Y}_i(j^c, 3) := \sum_{k=1}^{K} Y_i(j^c, k, 3) p_i(j^c, k, 3) \tag{7.2}$$

Column (3) displays the distribution of the difference, or gap, between $Y_i^{max}(j^c, 3)$ and $\overline{Y}_i(j^c, 3)$, i.e., $G_i(j^c, 3) = Y_i^{max}(j^c, 3) - \overline{Y}_i(j^c, 3)$. $G_i(j^c, 3)$ is our estimate of the lost income associated with not working in the occupation that maximizes their expected earnings in Phase 3 given their chosen major j^c.

Panel A of Table 16 shows the distributions of $Y_i^{max}(j^c, 3)$, $\overline{Y}_i(j^c, 3)$, and $G_i(j^c, 3)$ for the full sample of respondents to the Phase 3 survey. The average gap of slightly less than $30,000 represents about fourteen percent of the maximum earnings individuals expect to receive. Note that these estimates are lower bounds on income losses as it does not take into account any sorting into jobs within an occupation category. Almost 27% of respondents report with certainty that they will be working in the career that maximizes their earnings, which is why the first quartile of $G_i(j^c, 3)$ is zero.[39] Note that this holds even though there is a differences between the first quartile of $Y_i^{max}(j^c, 3)$ and the first quartile of $\overline{Y}_i(j^c, 3)$ as those who maximize their own income do not match one-to-one with those who have the highest maximum incomes.

Panel B of Table 16 repeats Panel A, but does so for the 73% of respondents who were not certain of choosing the career that maximizes their income, $Y_i^{max}(j^c, 3) > \overline{Y}_i(j^c, 3)$. Note that these individuals as a whole have lower maximum earnings than those who are certain of choosing the income-maximizing career: at all quartiles, the maximum earnings are lower in Panel B than those in Panel A. These results point to large *ex ante* earnings losses. On average, this group expects to give up almost $41,000 of earnings ten years after college as a result of not choosing the income-maximizing occupation, or a little over 21% of their maximum expected earnings. The distribution is skewed, however, with a median loss of $24,000. These results provide evidence that non-monetary components play a key role in explaining the choice of occupation.

[39] An additional 10% of respondent are certain they will be working in a career where their income is *not* maximized. Overall, Phase 3 respondents report a 57.6% chance of working in the occupation where their expected earnings are the highest.

Table 16: Distribution of Maximum and Expected Earnings: Phase 3 Data, 2009 dollars

	Max Earnings $[Y_i^{max}(j^c, 3)]$ (1)	Expected Earnings $[\overline{Y}_i(j^c, 3)]$ (2)	Difference (3)
Panel A: Full Sample			
Mean	212,946	183,020	29,926
1^{st} quartile	118,815	84,041	0
Median	158,419	143,370	14,258
3^{rd} quartile	237,629	210,405	35,124
Standard Dev.	165,133	148,179	48,427
Panel B: Conditional on $Y_i^{max}(j^c, 3) > \overline{Y}_i(j^c, 3)$			
Mean	193,000	152,126	40,874
1^{st} quartile	111,060	68,120	11,881
Median	158,419	128,478	23,961
3^{rd} quartile	198,024	187,727	47,526
Standard Dev.	154,956	120,947	52,543

DATA: Sample is 112 respondents to Phase 3 survey

8 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* monetary 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* returns 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 selection into occupations based on higher expected earnings. Clear complementarities exist between majors and occupations. For example, expected returns for business careers are highest for economics major, which mirrors the existence of higher subjective probabilities of pursuing a business occupation in the (sometimes) hypothetical case that they were an economics major. Comparing the distributions of expected earnings between under- and upper-classmen further suggests that students learn about the average returns to the various occupation-major pairs over the course of college.

Linking occupational choice probabilities to expected earnings and preferences for occupations, we provide evidence of sorting on *ex ante* returns, with the existence of significant and quantitatively large estimated elasticities of occupational choice with respect to expected earnings. Using data from a follow-up survey, we find that beliefs about earnings are strong predictors of actual occupational choices and realized earnings sever years later. Interestingly, expected earnings are positively associated with realized earnings after controlling for chosen major and occupation, which suggests that 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 estimate the determinants of sorting across alternatives.

References

Altonji, Joseph, Peter Arcidiacono, and Arnaud Maurel. 2016. "The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects." In *Handbook of the Economics of Education*, vol. 5, edited by Eric Hanushek, Stephen Machin, and Ludger Wößmann. Elsevier, 305–396.

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

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

Arcidiacono, Peter. 2004. "Ability sorting and the returns to college major." *Journal of Econometrics* 121:343–375.

———. 2005. "Affirmative action in higher education: how do admission and financial aid rules affect future earnings?" *Econometrica* 73:1477–1524.

Arcidiacono, Peter, Esteban. Aucejo, Hanming Fang, and Kenneth Spenner. 2011. "Does Affirmative Action Lead to Mismatch? A New Test and Evidence." *Quantitative Economics* 2:303–333.

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

Barton, Benjamin H. 2015. *The Glass Half Full: The Decline and Rebirth of the Legal Profession*. Oxford University Press.

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

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

Cardell, N. Scott. 1997. "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity." *Econometric Theory* 13:185–213.

Carneiro, Pedro M., Karsten T. Hansen, and James J. Heckman. 2003. "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice." *International Economic Review* 44:361–422.

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

———. 2008. "A new framework for the analysis of inequality." *Macroeconomic Dynamics* 12:315–354.

Cunha, Flávio, James J. Heckman, and Salvador Navarro. 2005. "Separating uncertainty from heterogeneity in life cycle earnings." *Oxford Economic Papers* 57:191–261.

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

Delavande, Adeline and Basit Zafar. 2014. "University Choice: The Role of Expected Earnings, Non-pecuniary Outcomes and Financial Constraints." Working paper.

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

Dominitz, Jeff and Charles F. Manski. 1996. "Eliciting student expectations of the returns to schooling." *Journal of Human Resources* 31:1–26.

———. 1997. "Using expectations data to study subjective income expectations." *Journal of the American Statistical Association* 92:855–867.

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

Heckman, James J., John E. Humphries, and Gregory Veramendi. forthcoming. "Returns to Education: The Causal Effects of Education on Earnings, Health and Smoking." *Journal of Political Economy* .

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

Hurd, Michael and Kathleen McGarry. 1995. "Evaluation of the subjective probabilities of survival in the HRS." *Journal of Human Resources* 30:S268–S292.

———. 2002. "The predictive validity of subjective probabilities of survival." *Economic Journal* 112:966–985.

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

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

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

Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad. 2016. "Field of Study, Earnings, and Self-Selection." *Quarterly Journal of Economics* 131:1057–1111.

Lee, Keith. 2015. "The Continuing Decline in LSAT Takers." URL `http://abovethelaw.com/2015/08/the-continuing-declining-in-lsat-takers/`.

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

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

———. 2004. "Measuring Expectations." *Econometrica* 72:1329–1376.

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

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

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

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

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

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

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

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

———. 2016a. "Human Capital Investments and Expectations about Career and Family." NBER Working Paper No. 22543.

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

Zafar, Basit. 2013. "College major choice and the gender gap." *Journal of Human Resources* 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 Evolution of beliefs about own abilities

In this section, we show how changes in beliefs of members of our sample of Duke students about their own future earnings can be combined with their beliefs about the future earnings of the average student at Duke to identify the evolution of individual-level uncertainty about a sample member's own abilities in different majors and careers.

To characterize the evolution of individual-level uncertainty about own abilities, 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, we assume that the potential income associated with occupation k and major j (Y^0_{ijk}) can be decomposed as follows:

$$\begin{aligned} Y^0_{ijk} &= \exp(\mu_{ijk} + \overline{\mu}_{jk}) \\ &= \exp(\mu_{ijk})\overline{Y}_{jk}, \end{aligned} \tag{A.1}$$

where μ_{ijk} and $\overline{\mu}_{jk}$ denote the (major, occupation)-specific individual and mean ability, and $\overline{Y}_{jk}$ is the income of the average Duke student for that same major-occupation pair. For notational convenience we omit the major and occupation indices (j, k) throughout the rest of the section, with the understanding that all of the variables are specific to that major-occupation pair.

Under the assumption that individuals have normally distributed prior beliefs on $(\mu_i, \overline{\mu})$ in each period, we write the beliefs about Y^0_i, denoted by Y_{it}, as:

$$\begin{aligned} Y_{it} &= E\left(\exp(\mu_i + \overline{\mu}) \mid \mathcal{I}_{it}\right) \\ &= \exp(\mu_{it} + \overline{\mu}_{it} + \sigma^2_{it}/2 + \overline{\sigma}^2_{it}/2 + \rho_{it}), \end{aligned} \tag{A.2}$$

where $\mathcal{I}_{it}$ denotes individual i's information set at t, $(\mu_{it}, \overline{\mu}_{it})$, $(\sigma_{it}, \overline{\sigma}_{it})$ are the means and standard deviations of the prior distributions of the individual and mean ability, and ρ_{it} 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:

$$\begin{aligned} \overline{Y}_{it} &= E(\exp(\overline{\mu}) \mid \mathcal{I}_{it}) \\ &= \exp(\overline{\mu}_{it} + \overline{\sigma}^2_{it}/2). \end{aligned} \tag{A.3}$$

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

$$\ln Y_{it} - \ln \overline{Y}_{it} = \mu_{it} + \sigma^2_{it}/2 + \rho_{it}. \tag{A.4}$$

The equality in (A.4) plays an important role in this 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 (A.4) still holds if we relax the rational expectations assumption and write instead the individual earnings beliefs as

$$\mathcal{E}(Y_i^0 \mid \mathcal{I}_{it}) = \kappa \times E(Y_i^0 \mid \mathcal{I}_{it}), \tag{A.5}$$

where $\kappa \neq 1$.

Specifically, we are interested in the evolution over time of the uncertainty about individual-specific ability, that is how σ_{it} changes between under- and upper-classmen. Assuming that individuals are forming rational expectations over their own abilities, $E(\mu_{it})$ will remain constant across t. If we further assume that the covariance terms, ρ_{it}, are equal to zero, then we can identify the evolution of uncertainty over time using a difference-in-differences strategy.[40] Namely:

$$E(\ln Y_{i,t+1} - \ln \overline{Y}_{i,t+1}) - E(\ln Y_{it} - \ln \overline{Y}_{it}) = E(\sigma^2_{i,t+1}/2 - \sigma^2_{it}/2). \tag{A.6}$$

It follows that the evolution between upper- and under-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.

The estimation results are reported in Table A.1. The 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.

The second takeaway from Table A.1 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 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

[40] This condition is stronger than necessary as the equality below holds as long as the covariance terms to remain, on average, constant over time, i.e., $E(\rho_{it}) = E(\rho_{i,t+1})$.

distributions:

$$| \sigma_1^2 - \sigma_0^2 | = \frac{1}{1 + \sigma_\epsilon^2/\sigma_0^2}, \tag{A.7}$$

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.

Table A.1: Change between upper- and under-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***

Notes: Major can either be the chosen major or a counterfactual major. *, **, and *** indicate statistical significance of differences at the 10%, 5%, and 1% level, respectively. "All" indicates average across majors (rows) and occupations (columns).

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 ρ_{it} 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:

$$\Delta\Big[E\big(\ln Y_{i,t+1} - \ln \overline{Y}_{i,t+1}\big) - E\big(\ln Y_{it} - \ln \overline{Y}_{it}\big)\Big] = \Delta\Big[E\big(\sigma_{i,t+1}^2/2 - \sigma_{it}^2/2\big)\Big] \tag{A.8}$$

where $\Delta(.)$ denotes the difference between the major-occupation (j, k) and a baseline (major, occupation) pair (j_0, k_0). It follows that the evolution between upper- and under-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.

In Table A.2, we present estimation results using Humanities-Education as a baseline alternative. 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 major-occupation pairs, such as (Social Sciences, Law), (Economics, Law), and (Public Policy, Business) relative to the (Education, Humanities) pair.

Table A.2: Change between upper- and under-classmen in the variances of own beliefs relative to (Education, Humanities) major-occupation pair

	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

Notes: Major can either be the chosen major or a counterfactual major.
*, **, and *** indicate statistical significance of differences at the 10%, 5%, and 1% level, respectively.
"All" indicates average across majors (rows) and occupations (columns).

A.3 Additional Tables

Table A.3: Incidence of elicited zero probabilities of choosing occupations in Phase 1, 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.

Table A.4: *Ex Ante* Treatment Effects of Occupations by Counterfactual Major (Earnings in 2009 dollars)

		Counterfactual Major ($j^{\sim c}$):					
Occupation:	Treat. Eff.	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 A.5: Average *Ex Ante* Treatment Effects (ATE) of Occupations: Under-Classmen versus Upper-Classmen (Annual Earnings, in dollars)

Occupation	Under-classmen	Upper-classmen	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 under-classmen and upper-classmen.

Table A.6: Comparison of Phase 3 and Phase 1 Samples

	Phase 3 Sample	Phase 1 Sample
Current/Intended Major:		
Sciences	17.9%	17.9%
Humanities	8.9%	9.3%
Engineering	21.4%	19.1%
Social Sciences	15.2%	17.9%
Economics	21.4%	19.7%
Public Policy	15.2%	16.2%
Class/Year at Duke:		
Freshman	21.4%	20.8%
Sophomore	18.8%	20.2%
Junior	26.8%	27.2%
Senior	33.0%	31.8%
Characteristics of Students:		
White	70.5%	66.5%
Asian	20.5%	20.2%
Hispanic	3.6%	4.6%
Black	1.8%	4.0%
Other	3.6%	4.6%
U.S. Citizen	96.4%	94.8%
Receives Financial Aid	41.1%	40.5%
*Mean Subjective Probability (Phase 1):**		
Science	0.182	0.180
Health	0.181	0.171
Business	0.273	0.266
Government	0.136	0.124
Education	0.086	0.095
Law	0.142	0.169
*Mean Expected Earnings (Phase 1):**		
Science	$92,598	$96,790
Health	$143,036	$142,540
Business	$160,420	$164,010
Government	$97,813	$100,350
Education	$75,929	$74,470
Law	$150,214	$163,220
*Mean Realized Earnings (7 years later):***		
	$131,527	
Sample Sizes	112	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 et al. (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?"'.
*Conditional on chosen/intended major.
** Earnings expressed in 2009 dollars, average over 81 individuals with non-missing earnings in Phase 3.

A.4 Additional Figures

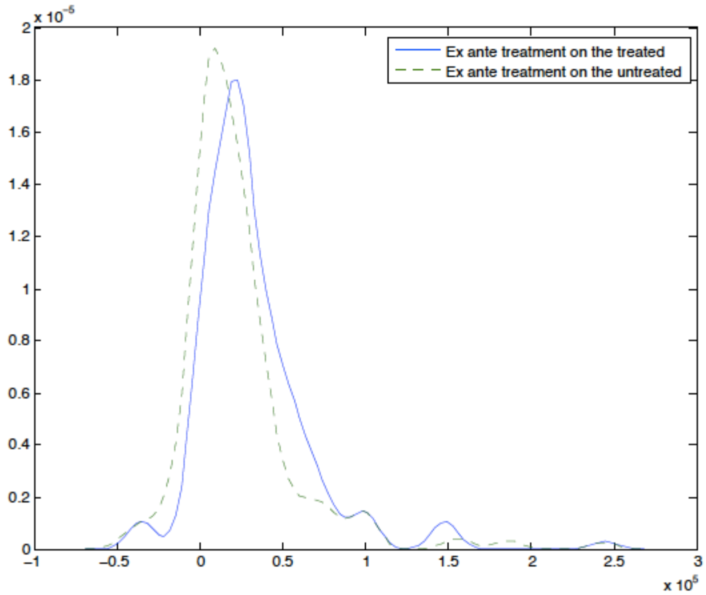

Figure A.1: Densities of *Ex Ante* Treatment Effects: Science

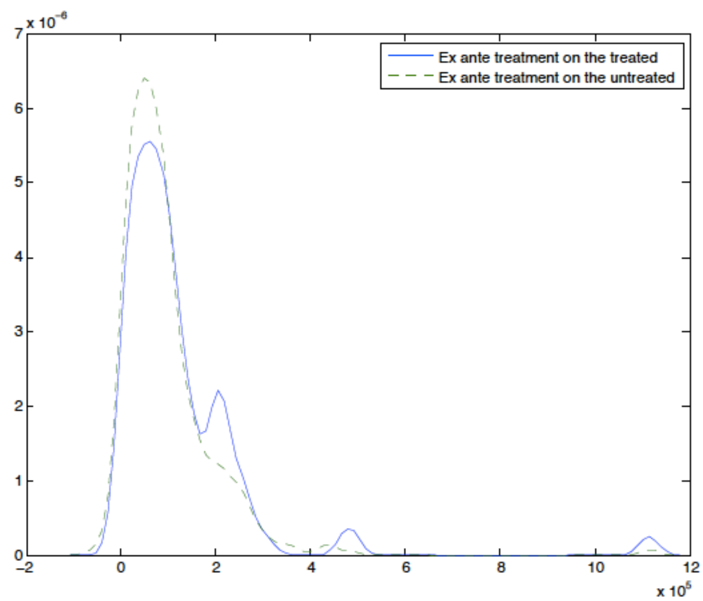

Figure A.2: Densities of *Ex Ante* Treatment Effects: Law