# *Ex Ante* Returns and Occupational Choice*

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

First version: October 2014
This version: July 1, 2017

**Abstract**

We show that data on subjective expectations, especially on outcomes from counterfactual choices and choice probabilities, can be used to recover *ex ante* treatment effects as well as the relationship between these treatment effects and individual choices. 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 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 provide clear evidence of sorting across occupations on expected earnings, with the earnings beliefs measured while the individuals were still in college being very informative about their future occupational choices.

*We are grateful to the editor and four anonymous referees for helpful comments. We also thank seminar participants at Arizona State University, Berkeley, Carnegie Mellon, CREST, London School of Economics, New York Fed, Stanford University, University College London, University of Chicago, University of Oslo, University of Tennessee, University of Wisconsin - Madison, Uppsala University, Toulouse School of Economics 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, Kate Maxwell Koegel and Jintao Sun 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.
¶Oxford 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 elicited expectations both on and off the individual's choice path can be used to recover *ex ante* treatments effects as well the relationship between individual choices and expected treatment effects. 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* 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* returns, as opposed to *ex post* ones. Complementing this literature that uses observational data on observed choices, we use data that directly elicit 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 collected 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 estimate the average *ex ante* treatment effect on the treated. Taking the major as given, we compute this parameter as a weighted average of the *ex ante* treatment effects for any given occupation $k$, using as weights the probabilities the individuals report they will work in occupation $k$. This parameter 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 $k$.

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

[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).

working in any given occupation $k$ relative to a baseline occupation. We further estimate the distributions of the *ex ante* treatment effects of working in any given occupation $k$ on the treated and untreated subgroups, by weighting the *ex ante* treatment effects with the occupational choice probabilities. Comparing the weighted distributions of *ex ante* treatment effects with the unweighted ones illustrates how selection on expected returns varies throughout the distribution.

Our results reveal large differences in expected earnings across occupations. Treating the education occupation as the baseline, the average *ex ante* returns 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 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 earnings 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 (95%) 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. 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.

Using the 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 college students hold about their future (choice-specific) labor market outcomes are predictive of the labor market choices they make later in their lives. Furthermore, we show that beliefs about earnings are predictive of what these individuals actually earned seven years later, even after controlling for chosen major and occupation.

We also examine how beliefs evolve over time in several different ways. First, we compare the expected incomes reported by lower- and upper-classmen. For example, we find that the beliefs that students have about the earnings of the average Duke students, which we elicit in the Phase 1 survey, are more homogeneous for upper-classmen than for under-classmen. This finding is consistent with students learning about occupation-specific skill prices before they even complete their undergraduate degrees. 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 find that expected returns to Business increased substantially between both surveys, as did the subjective probabilities of working in that occupation.

We then quantify the importance of sorting across occupations on expected earnings. To do so, we consider a simple framework linking choice probabilities to expected earnings, and preferences for occupations. We find a positive and statistically, as well as economically, significant effect of earnings beliefs on occupational choice. This finding is robust to the inclusion of individual preferences for each occupation, and the corresponding estimates remain qualitatively similar across specifications, as well as whether we use beliefs elicited while the individuals were in college or seven years later. Finally, we find that individuals expect to give up a sizable amount of money as a result of not choosing the highest paying occupation, suggesting that non-pecuniary factors also play an important role in one's choice of occupation.

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, and show that beliefs are predictive of actual labor market outcomes. 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, contrasting earnings beliefs of under-classmen with those of upper-classmen, and documenting how *ex ante* treatment effects changed between Phase 1 and Phase 3 survey. In Section 5, we link subjective choice probabilities to expected earnings and preferences, and quantify the importance of sorting across occupations on expected earnings. Next, we investigate in Section 6 the role that non-pecuniary factors play in occupational choice. We offer some concluding comments in Section 7. Additional details on the data, supplementary estimation results, and an analysis of how students update their beliefs about their own abilities are found in the 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.[6] 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.[7] 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 sciences, humanities, engineering, social sciences, economics, and public policy.[8] 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 sciences majors while under-representing students in public policy. Finally, the sample is slightly tilted towards upper-classmen.

### 2.1.1 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. 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 ten years after graduation. We used the following six broad sectors to characterize possible careers: Science/Technology, Health, Business, Government/Non-Profit, Education and Law.[9] It is important to note that, for all students in the sample, these probabilities and

[6] 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.

[7] 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).

[8] 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.

[9] In most of the paper, we simply refer to these six career groups as occupations. In practice we chose this classification based on the main groups of careers in which Duke graduates worked upon graduation, using data from the Duke Senior Exit Survey of 2007.

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% |
| Sample Size | 173 | |

DATA SOURCES: 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?"

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 ten 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 $j$ ten years after graduating from Duke, again elicited during Phase 1.

Table 2 reports the means of the expected incomes for the various major-occupation combinations collected in Phase 1 of the DuCMES (the $Y_i(j,k,1)$'s).[10] Note that each cell contains averages of the responses by each of the 173 students. Expected incomes exhibit sizable variation both across majors and occupations. For instance, majoring in the natural sciences or engineering is perceived to lead to higher earnings in Science and Health careers, while expected earnings in Business are, on average, higher for economics majors. Differences across occupations are even starker. In particular, average 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.

[10] In our sample, only 1.6% of the expected earnings are missing. For these cases, expected earnings, for each major and occupation, are set equal to the predicted earnings computed from a linear regression of log-earnings on major and occupation indicators, interaction between major and occupation, individual-specific average log-earnings across all occupations and majors and an indicator for whether the subjective probability of working in this occupation is equal to zero ($p_i(j,k,1) = 0$). 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 rescaled his expected income accordingly.

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.

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 1). Overall, the likelihood of working in the various occupations appear to be selectively different across individuals, even after conditioning on a college major.[11]

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

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

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

Table 3: Mean 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: Probabilities were elicited for each possible major-occupation pair at Phase 1, regardless of the respondents' chosen or intended major.

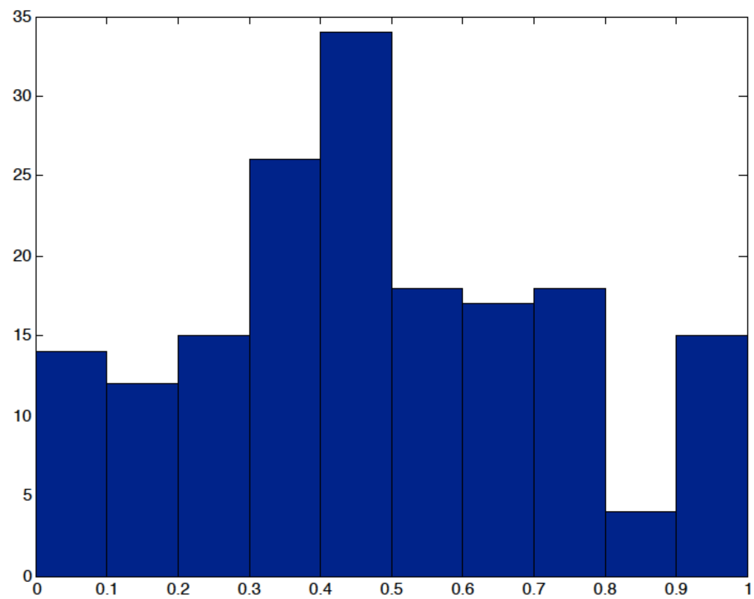


Figure 1: Frequency distribution of subjective probabilities for Economics major, Business occupation (Phase 1 data)

## 2.2 Phase 2 and Phase 3 data

In order to assess whether beliefs about future labor market outcomes are predictive of the actual choices made by the individuals after graduating from college and future labor market outcomes, 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. 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 were able to find the occupations of 143 out of the 173 individuals from our original sample on *LinkedIn*. For another eighteen 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 four more respondents directly from updated information in the Alumni Directory. Thus, our Phase 2 data collection produced current occupations for 165 of the 173 members of our original sample.

The occupation data obtained from these Phase 2 sources were mapped into each 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. In the following, we 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* labor market outcomes, and updated our sample members' expectations about careers, in a follow-up survey administered between February and April of 2016.[13] The respondents were contacted via email, *LinkedIn* message and/or text message.[14] 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 survey 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 ten years after college graduation.[15] Consistent with the notations introduced 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 nineteen 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 two additional individuals. Overall, we ended up with non-missing data on current occupations for a total of 167 of the 173 (96.5%) 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.

[13] 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`.

[14] 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.

[15] 8.3% of the expected earnings elicited in the Phase 3 survey are missing for the 112 individuals who completed this follow-up 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.

### 2.2.1 Subjective choice probabilities versus actual occupations

We next explore the relationship between the data on subjective occupational choice probabilities that we elicited in the Phase 1 and 3 surveys, conditional on 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 shares 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 13.7 percent chance they would in 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 though 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. In particular, 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.[16]

Column (3) of Table 4 shows beliefs about future occupations that were elicited in the Phase 3 follow-up survey, seven 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 supported by the fact that Phase 3 expectations about careers are very similar to the actual choices obtained in Phase 2. Individuals in Phase 3,

[16] 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.

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.

Finally, the last two columns of Table 4 show the expected probability of working in each career elicited in Phase 3, conditional on currently working and not working in that career. In all cases, the average perceived probability of working in their current careers in three to six years is over fifty percent which is significantly higher than the correspondingly probabilities that were elicited at Phase 1 (Column (4)). This suggests that, at this stage, 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.

**Decision to work in Business** While the previous results show 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 several pairs were not chosen by our sample. However, we can examine in particular the decision to work in Business, since, 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 (indexed by $k = 3$) 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 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 results in a lower $R^2$ (0.125) than conditioning 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.

It is possible that the findings in Table 5 are driven by the fact that the estimation sample includes seniors for whom many already had jobs lined up upon graduation.[17] 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 inclusion 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 one at any standard level. As with the full sample, the results show that the subjective probabilities are informative about future career choices.

### 2.2.2 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 and the expected earnings elicited in the initial Phase 1 survey. To do so, we use data on 81 individuals who reported having positive current annual earnings in the Phase 3 survey.[18]

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

[18]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. The vast majority (more than 80%) of those individuals were medical interns or residents, who did not consider these positions

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: Full sample includes 167 individuals for whom we obtained their current occupation from Phase 2, augmented with some Phase 3 data. 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.

In Table 6 below, 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, in the spirit of the analysis conducted earlier of the relationship between subjective probability and decision to work in business (Table 5). The estimated elasticity (0.64) is positive, sizable and statistically 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.[19] Taken together, the results presented in this section provide evidence that beliefs about future occupations as well as earnings are predictive of future labor market outcomes.

Table 6: 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.

# 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

as jobs, or were enrolled in a MBA program at the time of the survey.

[19]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.

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 distributions 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 $\Delta Y_{ik} := Y_i(j^c, k, 1) - Y_i(j^c, 1, 1)$, where, as before, $Y_i(j^c, k, 1)$ denotes individual $i$'s expected earnings in occupation $k$ given the chosen major, $j^c$, and expectations were elicited at Phase 1.

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 earnings ten years out relative to working in the base occupation.[21] This population parameter is

[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. Beyond college majors, there is less variance overall in expected earnings for Education than for the other occupations, which makes it a natural reference alternative. 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.

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

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 work in occupation $k$ ten years after graduation, 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 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 the assumption that individuals form rational expectations about their future occupational choices. Under this assumption, Equation (3.3) still holds after replacing the weights $\omega_{ik}^{TT}$ by $d_{ik}/E(d_{ik})$, where $d_{ik}$ is the indicator for whether $i$ works in occupation $k$ ten years after graduating, so that:

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

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

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, assuming that students form rational expectations over their future choices, the *ex ante* treatment effect of occupation $k$

on the *untreated* is given by:

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

where $\omega_{ik}^{TUT} = (1-p_{ik})/E\,(1-p_{ik})$. 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 $\widehat{\omega}_{ik}^{TUT} = (1-p_{ik})/[N^{-1}\sum_i(1-p_{ik})]$.[22]

Table 7 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. 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 positive selection on expected gains across occupations, the estimated $TUT$'s in Table 7 are lower than the $TT$'s 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 $TUT$'s and $TT$'s turn out to be much weaker, and only significant at 10%, 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 role of selection with our data is to construct the *ex ante* analogues of occupation-specific earnings, both 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, as would be observed in national data sets such as the American Community Survey (ACS). Using our expectations data, we can produce *ex ante* analogues of both measures. Namely, define the selected *ex ante*

[22] In the following we will somewhat abusively refer to the population parameter $TT(k)$ ($TUT(k)$) as the *ex ante* treatment effect of occupation $k$ on the treated (untreated). One should keep in mind that, if individuals do not form rational expectations over their future choices, this parameter should be interpreted as a weighted average *ex ante* treatment effect.

[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 7: *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 ($N = 173$).
NOTES: Standard errors are reported in parentheses. $TT$ is significantly different from $TUT$ for Science (p-value = 0.051), and Health (p- value = $3.10^{-4}$).

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 *ex ante* treatment effect on the treated, this parameter upweights the expected earnings by individuals' subjective probabilities of being in occupation $k$, thereby mimicking the realized earnings of those who 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:

$$\Delta\widehat{SE}(k) \quad = \quad \widehat{SE}(k) - \widehat{SE}(1) \tag{3.9}$$

This is the *ex ante* equivalent of the raw earnings premium of occupation $k$ relative to Education, which would be estimated in a dataset such as the ACS through a simple difference in mean earnings across occupations.[24] We can then compare these estimates with the average *ex ante* treatment effects to quantify how much of the selected *ex ante* earnings premium

[24]It follows from the definitions of the selected *ex ante* earnings difference and the *ex ante* treatment effect on the treated that $\Delta SE(k) = TT(k) - E\left((\omega_{i1}^{TT} - \omega_{ik}^{TT})Y_{i1}\right)$. Thus, unlike the earlier comparisons between conditional and unconditional *ex ante* treatment effects, the discrepancy between $\Delta SE(k)$ and $ATE(k)$ depends on the correlation between the normalized subjective probabilities of working in Education ($\omega_{i1}^{TT}$) and the expected earnings in that occupation ($Y_{i1}$), and as such also reflects selection into Education.

$\Delta\widehat{SE}(k)$ is due to selection.

Panel A of Table 8 performs this decomposition using Phase 1 data. Row (1) displays estimates for the selected earnings $SE(k)$ and, as a point of reference, the simple unweighted means of the *ex ante* earnings for occupation $k$, $\overline{Y}(k) := N^{-1}\sum_i Y_{ik}$, are displayed in Row (2). The fact that the means of selected *ex ante* earnings are all greater than $\overline{Y}(k)$ is indicative 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 average *ex ante* treatment effects, $\widehat{ATE}(k)$. The nature of selection into occupations based on *ex ante* returns is illustrated by 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 the share of that difference with respect to the selected *ex ante* earnings differentials $\Delta\widehat{SE}(k)$, 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. Put differently, our sample members, on average, expect to choose occupations in which they expect to earn more. Furthermore, the selection share estimates show that selection is much stronger for the Health occupation, least important for the Business occupation, with the other occupations somewhere in between. These results are consistent with Table 7 in that the gap between $TT$ and $TUT$ is especially large in Health and small in Business.

Panel B of Table 8 presents the same statistics as in Panel A, using data from Phase 3 on respondents' expectations about the expected earnings and probabilities of being in each occupation $k$ ten years after graduation. Using these more recent elicited expectations allows 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 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 a very large increase in selection for Business careers, smaller increases in Health and Science careers and actual declines in selection for careers in Government and 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.[25]

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

Table 8: 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) $\widehat{ATE}(k)$ (Aver. *ex ante* 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) $\widehat{ATE}(k)$ (Aver. *ex ante* 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 ($N = 173$) and Phase 3 ($N = 113$).

## 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.

The density of the *unconditional* distribution of the *ex ante* treatment effects for occupation $k$, i.e., $\Delta Y_{ik}$, 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:

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

where $\omega_{ik}^{TT}(u) := g(u)/E(p_{ik})$ and $g(u) = E(p_{ik}|\Delta Y_{ik} = u)$. If individuals form rational expectations over their future occupational choices, 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.[26] 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 $g(u)$. In the following we will use the same abuse of language as for the mean *ex ante* treatment effect parameters, and simply refer to $f_{TE,k}^{Treated}(.)$ as the density of the *ex ante* treatment effects on the treated for occupation $k$. Finally, the distribution of the *ex ante* treatment effects on the untreated 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.[27] (The distributions of the *ex ante* treatment effects for Science and Law are displayed in Figures

[26]Note that this remains true in the presence of unanticipated aggregate shocks, provided that these shocks affect the shares of workers in each occupation in a multiplicative fashion.

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

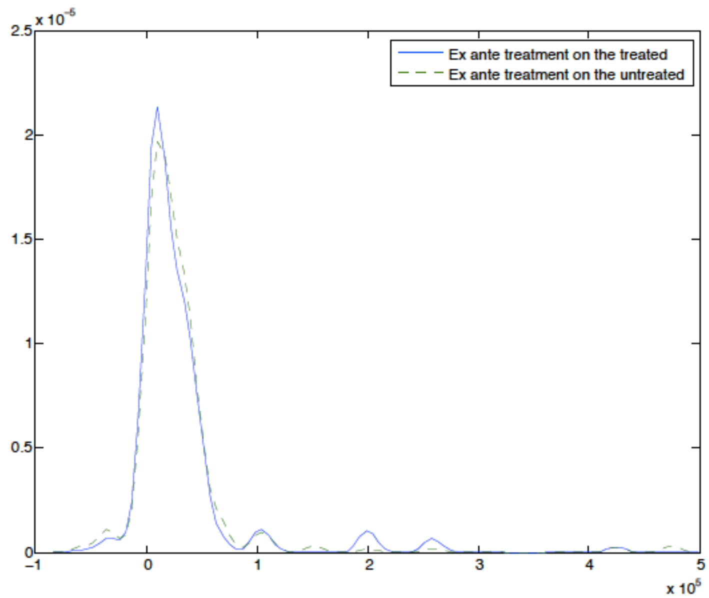


Figure 2: Densities of *Ex Ante* Treatment Effects on the Treated & Untreated: Government

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.[28] This latter pattern suggests that there is a 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 show 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 on each of the majors that respondents chose. Let $m_{ij}$ denote an indicator variable for

[28]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).

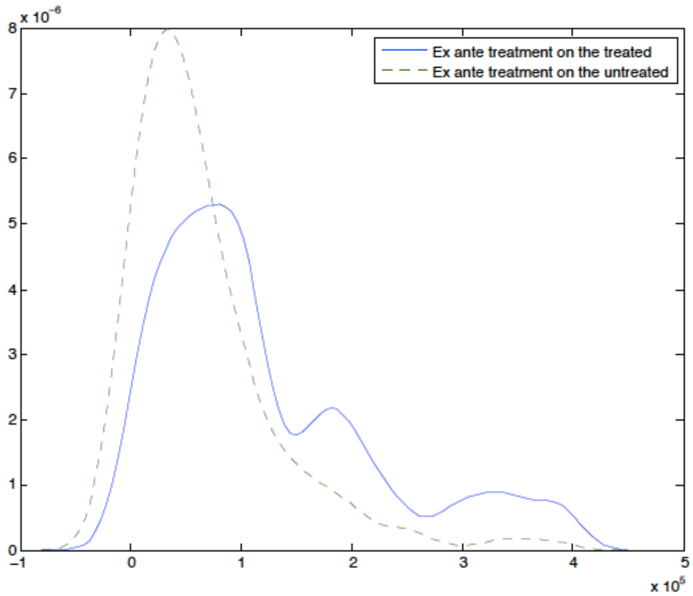


Figure 3: Densities of *Ex Ante* Treatment Effects on the Treated & Untreated: Health

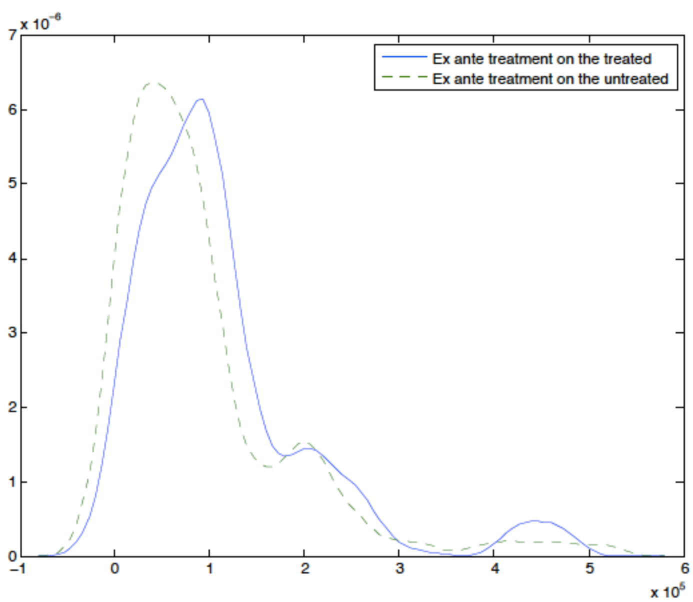


Figure 4: Densities of *Ex Ante* Treatment Effects on the Treated & Untreated: Business

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, $j^c = j$, are given by:

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

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

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

Given that we also elicit the subjective expectations for all counterfactual majors, we can 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.

In Table 9, we present the estimates of these *ex ante* treatment effects conditional on particular majors, $j^c = j$, 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 Sciences 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 Sciences 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 Sciences and Health, as well as engineering or natural Sciences 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 Sciences major. We will use below the *ex ante* treatment effects conditional on non-chosen majors to tell these two mechanisms apart.

As can be seen in Table 9, *ex ante* treatment effects on the untreated by students' majors generally are lower than the treatment effects on the treated, similar to the results obtained without conditioning on the major (Table 7). There are, however, a couple of exceptions. For instance, *ex ante* returns to Science careers are higher for the untreated in social Sciences 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 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 natural Sciences majors, where selection accounts for around half of the expected premium.

Finally, Table A.4 in Appendix A.3 provides estimates of the three *ex ante* treatment effects by counterfactual (non-chosen) 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 Sciences 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.[29]

## 3.4 *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 working in particular occupations vary with the occupations they actually chose.[30]

In Table 10 we present estimates for 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, where the treatment status is defined based 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 10 with those in Table 7 that were using subjective probabilities of choosing particular occupations, 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

[29]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.

[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 9: *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 ($N = 173$).
NOTE: Standard errors are reported in parentheses.

Table 10: *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 2.2 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 = $10^{-5}$).

average higher earnings premia for those occupations relative to those who work in another occupation. Estimates of *ex ante* treatment effects on the treated are significantly different from those on the untreated at the 1% level for Health careers, and marginally significant at 10% for Science.

Selection effects are strongest for Health careers, echoing our earlier findings based on expected, rather than actual, choices, while the estimates of the *ex ante* treatment effects on the treated and untreated for Law are similar in magnitude, even though the corresponding estimates in Table 7 are quite different. In Section 4.2 below 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

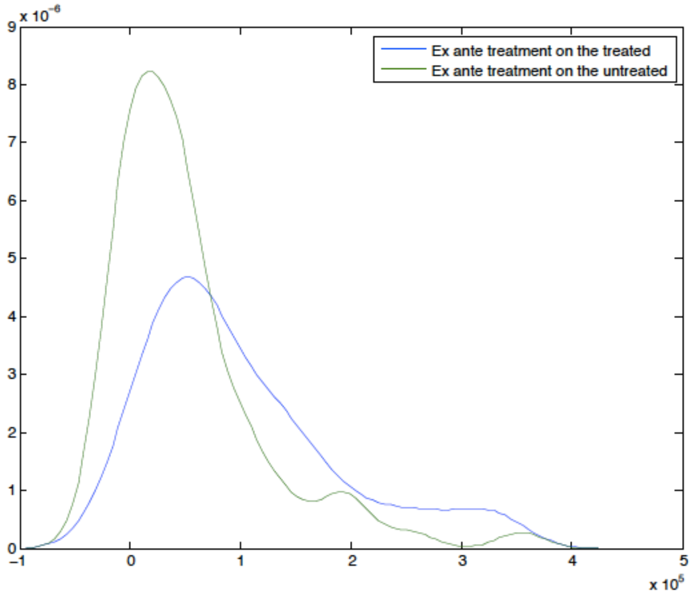


Figure 5: Densities of *Ex Ante* Treatment Effects on Treated & Untreated, conditional on Chosen Occupation: Health

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 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, the distributions of *ex ante* treatment effects for Business that conditions on chosen versus possible occupations (Figures 6 and 4) also point to a qualitatively similar pattern in terms of selection.

# 4 Evolution of beliefs

In practice, expected earnings may evolve over time as individuals obtain new information about their own major- and occupation-specific abilities, as well as about the average earnings and returns to ability within each occupation and major. In this section we discuss how our elicited data on expected earnings can be used to characterize how beliefs evolve over time. First, using elicited data of under- and upper-classmen, we characterize learning while in college about the average earnings within each occupation and major. We then document

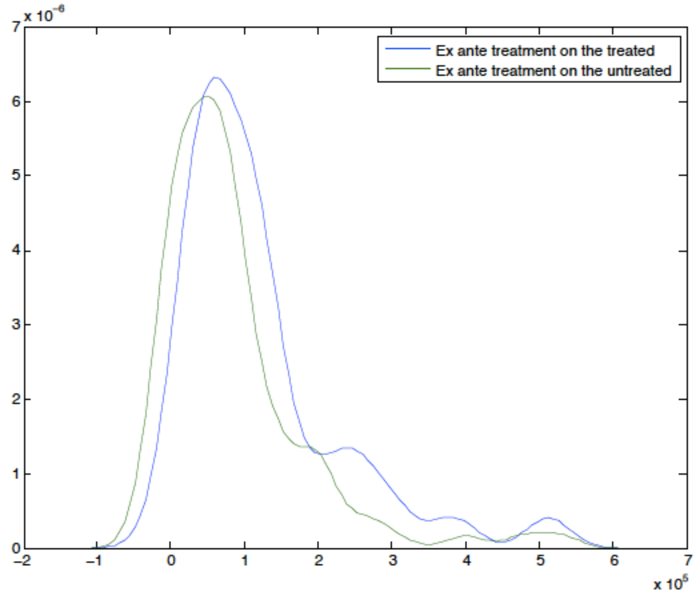


Figure 6: Densities of *Ex Ante* Treatment Effectson Treated & Untreated, conditional on Chosen Occupation: Business

the evolution of the *ex ante* treatment effects of the different occupations between the Phase 1 and Phase 3 surveys.

## 4.1 Learning in college

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

Students may be 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 outlining a set of assumptions under which 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 [Natural 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-cohort dispersion of income beliefs about the average Duke student to decline over time. In Table 11, we report the differences in the variance of log-expected incomes for the average Duke student between upper- and under-classmen, for each possible major-occupation pair. 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).[31] Table 11 also shows that the magnitude of those changes in variances tends to be substantial. The variance of the log-expected incomes decreases indeed by about 35%, on average across all occupations and majors for which the variance declines over time.

Table 11: 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 in 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.

## 4.2 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 12 shows estimates of the *ex ante* treatment effects on the treated ($TT$), treatment

[31] Due to the large number of hypotheses being simultaneously tested, controlling the familywise error rate with a standard Bonferroni correction of the p-values would result in none of the differences being statistically significant at standard levels. That the significance of these estimates is not robust to multiple testing correction is to be expected since most of the estimates are not significant at standard levels even without such a (conservative) correction. It is more appropriate from a statistical standpoint to focus instead on the average estimated difference across all occupations and majors, which is negative and significant at 1%, thus supporting the hypothesis that students learn about the average incomes over the course of college.

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 Phase 3 survey, seven years after those elicited in Phase 1. This table replicates the results in Table 7 (Section 3.1), using the updated beliefs of respondents elicited in Phase 3. Recall that each survey asked the individuals to give their current beliefs of what they expect to be earning 10 years after completing their degree. At enrollment in our study in 2009, students in our sample came from all four classes (freshmen, sophomores, juniors and seniors). Thus, at the time of the Phase 3 survey in 2016, these students were between four and seven years since graduation, giving us variation in how close they currently were to the ten-year benchmark used in our elicitations.

Several comments are in order about the treatment effects in Table 12 and comparison to those in Table 7 based on expectations elicited in Phase 1. First, the $ATE$s are substantially higher in Science and Business relative to the Phase 1 beliefs. However, the estimated $TT$s and $TUT$s 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 $ATE$s for Business is primarily driven by the increase in the $TT$s, resulting in a large and significant discrepancy between the $TT$s and $TUT$s for Business. Second, the $ATE$s 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 $TT$s and $TUT$s fall, but the former decline is much larger such that the $TT$s becomes actually lower than the $TUT$s. 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.[32]

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.

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

Table 12: *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).

# 5 Occupational choice and sorting on expected earnings

The findings in the preceding sections all indicate a positive association between expected earnings and occupational choice. However, these results may partly reflect preferences for occupation-specific non-pecuniary job attributes that are correlated with the expected earnings. In this section, we go a step further and relate the choice of occupations – both on an *ex ante* and *ex post* basis – to the elicited *ex ante* earnings beliefs of our respondents, and the (unobserved) non-pecuniary occupation attributes. Using the data from the three phases of our study, this framework allows us to quantify the importance of sorting on expected earnings.

We use the following simple framework to model occupational choice. Assume individuals choose their occupations to maximize their expected utility. Conditional on major $j$, individual $i$ chooses among $K$ mutually exclusive occupations. Let $d_{ijk} = 1$ if $k$ is 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:

$$\max_{(d_{ijk})_k} \sum_{k=1}^{K} d_{ijk}\big(u_{ijk} + \varepsilon_{ijk}\big), \tag{5.1}$$

where $u_{ijk}$ denotes the (expected) utility that is observable by the researcher up to a vector of parameters and $\varepsilon_{ijk}$ is unobserved to the researcher, assumed to be drawn from a standard Type 1 extreme value distribution and independent across occupations. We further assume that $u_{ijk}$ is a function of log expected earnings in the occupation-major pair $(k, j)$, $y_{ijk}$, given by

$$u_{ijk} = \alpha_k + \beta y_{ijk} \tag{5.2}$$

where we allow for occupation-specific utility payoffs, $\alpha_k$, and normalize $\alpha_1$ to zero. The specification given by (5.1) and (5.2) constitutes our basic model of occupational choice. We first use the above specification to model the actual 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. Namely, we set $y_{ijk} = \ln[Y_i(j, k, 1)]$.[33] Estimates for the $\beta$ parameter in (5.2) are displayed in Column (1) of Table 13, both for the full sample of all respondents and the sample excluding seniors.[34] We also estimate the model without seniors as some seniors may already have jobs lined up at the time of the survey. For both samples, the estimates in Column (1) show a significant, positive and strong relationship between expected future earnings and respondents' actual choice of occupations, a result that is consistent with positive sorting across occupations on expected earnings.

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, using the notations introduced above, some of what is unobserved to the researcher, i.e. the $\varepsilon_{ijk}$'s, is actually known by individual $i$ at that time, but some is not known to either the researcher or individual $i$ back at Phase 1 when they form their beliefs about the probabilities of working and their expected earnings in each of the occupations. It is the lack of knowledge about this latter part of $\varepsilon_{ijk}$ that makes 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.

[33] We also considered an alternative specification where we assumed that preferences are linear, as opposed to logarithmic, in the expected occupation-specific earnings. This specification yielded positive and significant estimates of the earnings coefficient. However, 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 Arcidiacono (2004, 2005) who uses a similar specification of the expected utility of future labor market outcomes. This specification of the indirect utility function can be derived from utility maximization when flow utility is given by the logarithm of current consumption, and assuming perfect credit markets.

[34] Note that if one maintains the assumption that students form rational expectations over their future choice of occupation, this specification allows for aggregate occupation-specific earnings shocks. Aggregate shocks affecting log-earnings additively would be absorbed by the occupation dummies, and the conditional logit would therefore still consistently estimate the earnings coefficient $\beta$ in this case.

To acknowledge this source of individuals' uncertainty, let

$$\varepsilon_{ijk} = \phi_{ijk} + \zeta_{ijk}, \tag{5.3}$$

where the $\phi_{ijk}$'s are known to $i$ at Phase 1 but the $\zeta_{ijk}$'s are unknown. Assuming that $\zeta_{ijk}$ has a standard Type 1 extreme value distribution and that students form rational expectations over their future choice of occupation, we can invert the self-reported probabilities to obtain:[35]

$$\begin{aligned} \ln\left[p_i(j,k,1)\right] - \ln\left[p_i(j,1,1)\right] &= u_{ijk} - u_{ij1} + \phi_{ijk} - \phi_{ij1} \\ &= \alpha_k + \beta \Delta y_{ijk} + \Delta \phi_{ijk} \end{aligned} \tag{5.4}$$

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

We first estimate $\beta$ using Equation (5.4), 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), and then estimate the flow utility parameters using a least absolute deviation (LAD) estimator.[36]

Results of the LAD estimation of (5.4) are given in Column (2) of Table 13. The estimate of $\beta$ for the full sample (1.371) is economically and statistically significant, while the estimate excluding seniors from the sample (1.337) is very similar in magnitude.

Moreover, our elicited occupational choice probabilities allow one to control for major-occupation dummies even with our relatively small sample. That is, we can replace the $\alpha_k$'s with $\alpha_{jk}$'s in the regression specification in (5.4), otherwise using the same data as used for Column (2). These new estimates are presented in Column (3) of Table 13. Adding the major-occupation interactions reduces the magnitudes of the estimated coefficients on log income, especially those for the sample that excludes seniors, where the estimate declines by 40%. Nonetheless, both estimates remain positive with this specification, as well as statistically and economically significant.

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 individual-occupation specific fixed effects in addition to the major-occupation interactions.[37]

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

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

[37] See Wiswall and Zafar (2016b) who provide evidence from NYU students that preferences for non-

Table 13: 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) |
| *Controls:* | | | | |
| Occupation | Y | Y | N | N |
| Major $\times$ Occupation | N | N | Y | Y |
| Individual $\times$ Occupation | N | N | N | Y |

DATA: *Full Sample* includes 167 individuals while the *Excluding Seniors* sample contains 113 individuals. Major-occupation-specific expected earnings and occupational choice probabilities are from Phase 1, and actual occupational choices are from the augmented Phase 2 data.
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 estimate regressions of the form given in (5.4). Columns (2) and (3) use observations corresponding to chosen majors $j^c$ only. 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 $\times$ occupation level.

The results for this final specification, where the individual-occupation fixed effects are eliminated by within-transformation, are reported in Column (4) of Table 13, 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 $\beta$. And, while the estimate of $\beta$ falls slightly for the full sample, the results provide clear evidence of positive sorting on expected earnings across occupations.

To quantify the responsiveness of subjective occupational choice probabilities to expected

pecuniary job attributes are highly heterogeneous across individuals. Our estimator remains consistent in the presence of unobserved individual-occupation specific characteristics which may be correlated with earnings beliefs. On the other hand, major-occupation heterogeneity in preferences operates through the unobserved preference term, which is assumed to be orthogonal to the covariates. Incorporating unobserved heterogeneity on this dimension that would be correlated with other variables (expected earnings in particular) requires eliciting expected earnings at multiple points in time. The estimation results which correspond to Specification 5.4 evaluated in difference between Phase 1 and Phase 3 beliefs (see Table 15) are robust to such unobserved heterogeneity.

earnings, we calculate the percentage change in the probability of choosing an occupation given a percentage change in expected earnings, using the estimates in Column (4) in Table 13. These elasticities, denoted by $e_{ijk}$, are estimated using (Train, 2003):

$$\widehat{e}_{ijk} = [1 - p_i(j,k,1)]\widehat{\beta}, \tag{5.5}$$

for each individual $i$ and major-occupation pair, and where $\widehat{\beta}$ denotes the estimate of $\beta$ ($\widehat{\beta} = 0.953$ for our preferred specification in Column (4)). Note that this formula only applies for the intensive margin, that is for variation in the subjective probability $p_i(j,k,1)$ strictly between 0 and 1. Hence, we estimate this elasticity only with the data on individuals who provided non-zero choice probabilities. 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. We then compute the average elasticity for each occupation $k$ ($\hat{e_k}$) as the sample average of the elasticities conditional on chosen major ($\widehat{e}_{ij^ck}$).

The resulting occupation-specific elasticity estimates range from 0.65 (for Business) to 0.82 (for Education) and yield in a mean elasticity across all occupations of 0.74. That is, on average across individuals and occupations, a 10% increase in the expected earnings for a given occupation is associated with a 7.4% increase in the subjective probability of choosing that occupation. It is worth noting that these elasticities are sizable, especially in comparison with the very low earnings elasticities which have typically 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 recent survey).

Thus far, our analysis has made use of the beliefs elicited from respondents in 2009 when they were still undergraduate students about what their earnings and occupations would be ten years after graduation. As we have already discussed, the beliefs we elicited in the Phase 3 survey (conducted in 2016) give us another source of data to assess the importance of sorting on expected earnings for occupational choice. In Table 14, we repeat the *ex ante* occupational choice analyses found in Table 13 using these more recent occupation choice probabilities. We use the earnings our respondents would expect to receive in each of these occupations ten years after graduation that we elicited from them in the Phase 3 survey to estimate the returns to expected earnings. We present results for three different specifications: one where we control for occupation dummies [Column (1)], another that controls for major-occupation dummies [Column (2)], and one that adds a dummy variable indicating whether the particular occupation is the respondent's actual occupation as of 2016

Table 14: 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) |
| *Controls:* | | | |
| Occupation | Y | N | N |
| Major × Occupation | N | Y | Y |
| Current Occupation | N | N | Y |

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

[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 13 provides evidence that beliefs about future choice of occupation tend to be even more tightly associated with expected earnings gathered in the Phase 3 follow-up survey than in the initial Phase 1 survey. 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 and changes in the occupation-specific expected earnings. Table 15 reports the LAD estimation results which correspond to the specification in (5.4), but here differencing between Phase 3 and Phase 1 beliefs. Note that forming the difference between Phase 3 and Phase 1 beliefs ensures that occupation-major-individual fixed effects cancel out, so that the estimated sorting effects are robust to any occupation-specific preferences and/or major-specific non-pecuniary job attributes that may be correlated with expected earnings. In all three specifications, the estimated earnings coefficient remains positive and significant, both statistically and economically. 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 close in magnitude to the estimates that were obtained for the most comparable specifications in Table 14 (Column 3, 0.914) and in Table 13 (Column 4, 0.953). While the magnitude decreases once we allow aggregate preferences for majors and occupations to vary over time by adding occupation-major fixed effects, the estimated coefficient (0.783) remains statistically significant and sizable. Taken

Table 15: Changes in subjective probabilities of choosing occupations

| | (1) | (2) | (3) |
|---|---|---|---|
| $\Delta$ Log Income | 1.274 | 1.020 | 0.783 |
| | (0.239) | (0.235) | (0.206) |
| *Controls:* | | | |
| Current Occupation | N | Y | Y |
| Major $\times$ Occupation | N | N | Y |

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

together, these results provide yet further evidence that individuals sort across occupations based on their expected earnings.

# 6 The role of non-pecuniary factors

Given our previous findings, it is natural 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 highest paying occupation. These *ex ante* measures of willingness-to-pay for non-earnings-maximizing choices provide evidence on the role played by other, in particular non-monetary, factors in the choice of occupation, and are directly identified from the data. Importantly, this does not require any distributional assumptions, nor does it require to take a stand 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 typically require additional schooling. But the Phase 3 data likely does not suffer from this issue. Individuals at this stage have either completed their education or will do so soon. By using choice and earnings expectations from the Phase 3 data, we get measures of *ex ante* willingness-to-pay that are most likely not contaminated by tuition payments.

[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, it 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 in the absence of subjective expectations.

In Table 16 we display estimates of the *ex ante* earnings lost due to individuals making their occupational choices based on factors other than expected earnings. Column (1) of the table presents estimates of the mean, median, first and third quartiles and standard deviation for the distribution of the expected earnings for that occupation in which sample members expected to earn the most, i.e.,

$$Y_i^{max}(j^c, 3) := \max\{Y_i(j^c, 1, 3), Y_i(j^c, 2, 3), ..., Y_i(j^c, 6, 3)\}. \tag{6.1}$$

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.,

$$\overline{Y}_i(j^c, 3) := \sum_{k=1}^{6} Y_i(j^c, k, 3) p_i(j^c, k, 3) \tag{6.2}$$

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., $G_i(j^c, 3) := Y_i^{max}(j^c, 3) - \overline{Y}_i(j^c, 3)$. For any given individual $i$, $G_i(j^c, 3)$ is our estimate of the *ex ante* willingness-to-pay for non earnings-maximizing choices, measured 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.[39] Note that these estimates are lower bounds on *ex ante* income losses as they do 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 expected earnings, which is why the first quartile of $G_i(j^c, 3)$ is zero.[40] Note that this holds even though there is a non-negligible difference between the first quartile of $Y_i^{max}(j^c, 3)$ and the first quartile of $\overline{Y}_i(j^c, 3)$, as rank invariance does not hold here in the sense that 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 expected income, i.e. $Y_i^{max}(j^c, 3) > \overline{Y}_i(j^c, 3)$. Note that these individuals as a whole tend to have lower maximum earnings

[39]Here we use the terminology "maximum earnings" as a shorthand for "maximum occupation-specific expected earnings".

[40]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

than those who are certain of choosing the income-maximizing career: at all quartiles, the maximum earnings are lower (or equal) 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 their (*ex ante*) income-maximizing occupation, or a little over 21% of their maximum expected earnings. The distribution is skewed, however, with a median loss of about $24,000. Taken together, these results show that individuals have large *ex ante* willingness-to-pay for non earnings-maximizing choices, consistent with non-monetary factors playing a key role in explaining the choice of occupation.

# 7 Conclusion

This paper uses elicited beliefs from a sample of male undergraduates at Duke University on the expected earnings in different occupations as well as on the probabilities of working in each of those occupations, to recover the distributions of the *ex ante* 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 then provide evidence of sorting on expected earnings, 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 also are strong predictors of actual occupational choices, and that the association between expected earnings and subjective probabilities of occupational choice is even tighter using beliefs that we elicited seven years later. However, non-pecuniary components also play an important role in the choice of occupation, with individuals expecting to give up sizable amounts of money as a result of not choosing the highest paying 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 $\mu_{ijk}$ 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 $\rho_{it}$ 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 abilities, that is how $\sigma_{it}$ 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, $\rho_{it}$, are equal to zero, then we can identify the evolution of uncertainty over time using a difference-in-differences strategy.[41] Namely:

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

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 abilities, since, assuming normally distributed prior and noise distributions:

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

[41] 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., $E(\rho_{it}) = E(\rho_{i,t+1})$.

where $\sigma_0^2$ and $\sigma_1^2$ are the prior ability variances in period $t = 0$ and $t = 1$, and $\sigma_\epsilon^2$ 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 abilities relative to a baseline major-occupation pair is identified under milder assumptions. Specifically, assuming that the evolution over time of the covariance terms $\rho_{it}$ 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:

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

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.[42] These results support the idea 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 (Humanities, Education) pair.

[42]Note, however, that one should exercise caution when interpreting these estimates as these are not statistically significant at standard levels once we correct the critical values to account for multiple testing using a Bonferroni correction.

Table A.2: Change between upper- and under-classmen in the variances of own beliefs relative to (Humanities, Education) 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 (natural sciences, humanities, engineering, social sciences, 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

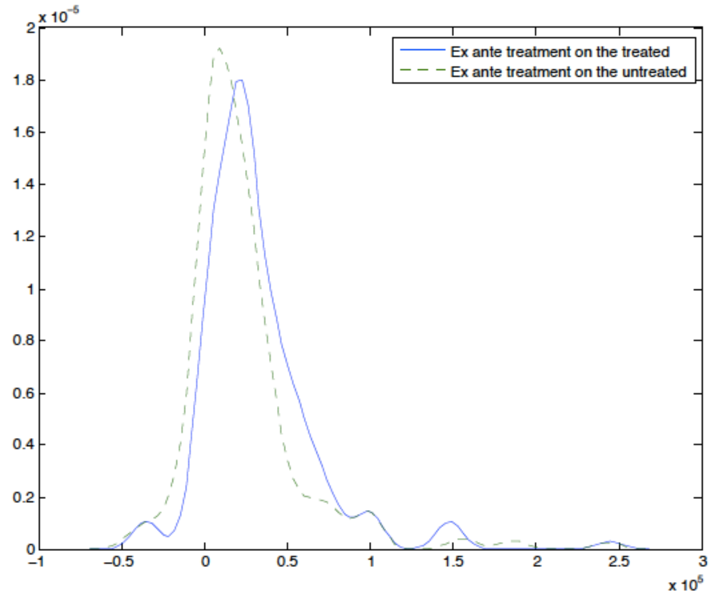


Figure A.1: Densities of *Ex Ante* Treatment Effects: Science

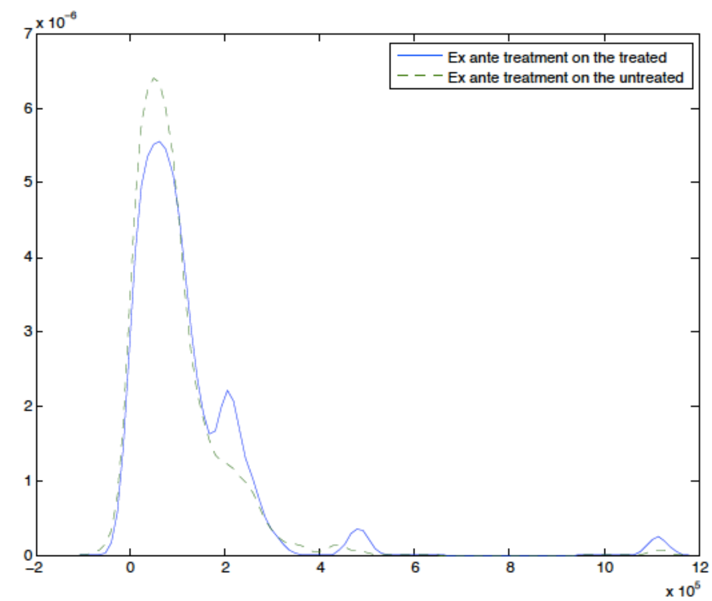


Figure A.2: Densities of *Ex Ante* Treatment Effects: Law