Ex Ante Returns and Occupational Choice*

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

March 2019

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 their 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. We also find clear evidence of sorting across occupations based on expected earnings, with the earnings beliefs measured while the individuals were still in college being very informative about their future occupational choices six years later. However, consistent with a generalized Roy model of occupational choice, non-pecuniary components also play an important role, with individuals expecting to give up sizable amounts of money as a result of not choosing the highest paying occupation.

*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 of Emory University.

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 & McGarry, 1995, 2002; Hurd, 2009; Dominitz & Manski, 1996, 1997), more recent articles have used elicited probabilities of taking particular courses of actions in the future, along with expectations about potential future outcomes corresponding to counterfactual choices (or treatments), to analyze how individuals are making their decisions under uncertainty (see, e.g., Arcidiacono et al., 2012; Zafar, 2013; Stinebrickner & Stinebrickner, 2014; Delavande & Zafar, 2018; Wiswall & Zafar, 2015, 2016, 2017).[1]

In this paper, we use panel data on the subjective expectations of college students to address three sets of questions. First, what is the predictive validity of expectations about future occupational choices and earnings elicited from individuals when they are students in college? The fact that we find substantial predictive validity from this data informs the second question: how can students' beliefs about future earnings and occupational choices be used to quantify selection on gains into occupations? In order to address this question, we first show that elicited expectations both on and off the individual's choice path can be used to recover *ex ante* treatments effects, and then investigate the relationship between individual choices and expected treatment effects. Third and finally, we explore what students' subjective beliefs and actual occupational choices reveal about the importance of expected earnings and non-pecuniary factors in their choices of occupation. To do so, we consider a variant of the generalized Roy model where agents sort across occupations based on expected monetary returns as well as non-pecuniary preferences for occupations.

To address these questions, we use beliefs elicited from male undergraduates who participated in the Duke College Major and Expectations Survey (DuCMES). In Phase 1 of the DuCMES, 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.[2] In Phase 2, we collected the occupations that participants were working in six years later for the vast majority (95%) of the DuCMES sample, using data from the social network *LinkedIn*, and the Duke Alumni Database. Finally, in Phase 3, we conducted a follow-up survey of all DuCMES sample members that was administered seven years after

[1]Several studies also have incorporated subjective expectations about objective events in the estimation of structural dynamic models (Delavande, 2008; van der Klaauw & Wolpin, 2008; van der Klaauw, 2012).

[2]This dataset was previously used to examine the determinants of college major choice by Arcidiacono et al. (2012). Their paper treated occupations as lotteries, where the lotteries were affected by the choice of major. In this paper, we follow a more conventional route and treat occupations as choices, consistent with, e.g., Miller (1984), Siow (1984), Keane & Wolpin (1997), Antonovics & Golan (2012), van der Klaauw (2012) and Wiswall & Zafar (2017).

Phase 1. In this follow-up survey we asked members of our sample about all of their current occupations and earnings as well as asking them to update their beliefs about their probabilities of working in different occupations and the associated earnings ten years after graduation. The respondents were contacted via email, *LinkedIn* message and/or text message, and we obtained responses from 68% of the initial sample.

We first show that the data collected in Phase 1 are strongly correlated with actual occupational choices in Phase 2 as well as actual earnings and their updated beliefs in Phase 3. For each occupation, those who chose that occupation in Phase 2 reported substantially higher probabilities of working in that occupation in Phase 1 than those who did not choose that occupation in Phase 2. For example, those working in a health occupation reported Phase 1 probabilities for that occupation that were over four times larger than the probabilities reported by those not working in a health occupation. Beliefs college students hold about their future occupations are then predictive of their future labor market choices later in their lives. Furthermore, we show that beliefs about earnings in Phase 1 are predictive of what these individuals actually earned seven years later, even after controlling for chosen major and occupation.

Given that the Phase 1 beliefs contain relevant information about later life outcomes, we next turn to what these beliefs tell us about the expected earning premia associated with different occupations. As recently emphasized in a series of papers on schooling decisions in the presence of heterogeneity and uncertainty (see, e.g., Carneiro et al., 2003; Cunha et al., 2005; Cunha & Heckman, 2007; and Cunha & 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.[3]

The data we collected allow us to identify the full distribution of *ex ante* treatment effects of particular occupations (relative to a reference occupation) on earnings. In order to quantify the role played by 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 chosen 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

[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 et al. (Forthcoming) and Kirkeboen et al. (2016).

the untreated, where we weight elicited expected returns by the declared probability that the individual will not work in occupation k.

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. Taken together these patterns are consistent with Roy-type models of occupational choice in which these choices are functions of individual beliefs about earnings returns.

These *ex ante* treatment effects for occupations ten years after graduation evolve from Phase 1 to Phase 3; *ex ante* treatment effects naturally depend on when they are measured. For example, occupation-specific aggregate shocks between Phase 1 and Phase 3 will affect beliefs about earnings in those occupations. The Phase 3 data show much larger *ex ante* treatment effects for business occupations than in Phase 1 and this is especially true for those see a high probability of choosing a business occupation in Phase 3. But as the *ex ante* treatment effects evolve so do the probabilities of choosing particular occupations, again consistent with a Roy-type model. Namely, occupations that see increases in *ex ante* treatment effects over time, such as the business occupation, also see increases in the probabilities of choosing those occupations. The reverse holds for occupations such as law where both the *ex ante* treatment effects and the probability of choosing that occupation fall over time.

Finally, we explore what students' subjective beliefs and actual occupational choices reveal about the importance of expected earnings and non-pecuniary factors in their choices of occupation. We find a positive, statistically, as well as economically, significant effect of earnings beliefs on occupational choices. This finding holds true for various specifications, and in particular whether the outcome measure is beliefs about occupational choice at Phase 1 or Phase 3, or the occupational choice at Phase 2, and whether the earnings measure is taken from Phase 1 or Phase 3 beliefs. This finding is also robust to the inclusion of individual preferences for each pair of occupation and major. The corresponding estimates of the earnings elasticities are qualitatively similar across specifications. 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. As such, these findings complement previous evidence using data on observed choices and realized income that pure income-maximizing models fail to predict individual choices (see, e.g., Heckman & Sedlacek, 1990 in the context of sectoral choices). Importantly, the use of our subjective beliefs data makes it possible to reach such a conclusion while remaining agnostic on the information set of the agents.

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, before showing in Section 3 that subjective beliefs are predictive of actual labor market outcomes. Section 4 shows how to obtain the means and distributions of *ex ante* treatment effects given the data, and then discuss the estimation results. 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. Section 7 concludes. Additional details on the data and supplementary estimation results 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.[4] 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.[5] 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.[6]

Students were also asked their expectations about their likelihood of choosing future careers, and how much they expected to earn in them. Namely, for each of six majors groups, 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,

[4]Arcidiacono et al. (2012) also use the Phase 1 DuCMES data employed in this paper. We refer the reader to that paper for a more comprehensive overview of the data.

[5]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).

[6]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.

Business, Government/Non-Profit, Education and Law.[7] It is important to note that, for all students in the sample, these probabilities and expected earnings were elicited for all possible occupation-major combinations, i.e. both for the chosen (or intended) majors and the counterfactual majors.

Specifically, to elicit career probabilities, students were asked:

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

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

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

Descriptive statistics for the Phase 1 data are shown in Appendix Tables A.1 through A.2. Table A.2 shows that there are substantial earnings variation across major/occupation pairs in ways that one would expect. For example, higher premiums in business are found if one majors in economics; higher premiums in health are found if one majors in the natural sciences. Table A.3 shows that higher earnings in a major/occupation pair are associated with higher probabilities of choosing an occupation conditional on major.

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

[7] 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.

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 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.[8] The respondents were contacted via email, *LinkedIn* message and/or text message.[9] 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.7 in Appendix A.2, 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,

[8] 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`.

[9] 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.

the comparison presented in Table A.7 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.[10]

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.

3 Predictive validity

3.1 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 1 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%.

[10] 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.

Although the beliefs are, on average, off for some of the occupations, the fourth and fifth columns of Table 1 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.[11]

Column (3) of Table 1 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, 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 1 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

[11]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 1: Chosen Occupations and Elicited Beliefs about Occupations

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

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.

3.2 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 2 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

Table 2: 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.

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 2 are driven by the fact that the estimation sample includes seniors for whom many already had jobs lined up upon graduation. In Columns (4)-(6) of Table 2, 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.

3.3 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.[12]

In Table 3 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 2). 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.[13] 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.

4 *Ex ante* treatment effects

In this section, we outline how the different types of *ex ante* treatment effects we are interested in can be measured, and show the corresponding effects in our data. We begin by considering standard treatment effect measures such as the average treatment effect, the average treatment on the treated, and the average treatment on the untreated. We then show how our beliefs data can be used to calculate the full distributions of the various treatment effects, and report examples from certain occupations.

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

[13]It is interesting to compare these results with Wiswall & Zafar (2016), who estimate in a different context the association between log realized earnings and log expected earnings. Among males, they find a positive but insignificant relationship between these two quantities, with an estimated elasticity of 0.167 without controlling for majors or occupations. They show that beliefs are much more predictive of the actual earnings among females.

Table 3: 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.

4.1 *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 $k = 1$.[14] 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.

Assuming that occupation-specific expected earnings are accurately measured through the beliefs that were elicited at Phase 1, the *ex ante* treatment effects are identified directly from the data. The average *ex ante* treatment effect of occupation k, denoted by $ATE(k)$, is then defined by:

$$ATE(k) := E\left(\Delta Y_{ik}\right). \tag{4.1}$$

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*

[14] We choose to use Education as a baseline because the earnings in this occupation do not vary much across college majors (see Table A.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 & Zafar (2016) 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.

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. This population parameter is identified from the data provided that any measurement errors affecting the elicited beliefs have the same mean across occupations. Under this assumption, $ATE(k)$ is consistently estimated using its sample analog:

$$\widehat{ATE}(k) = N^{-1} \sum_i \Delta Y_{ik}, \tag{4.2}$$

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:

$$TT(k) := E\left(\omega_{ik}^{TT} \Delta Y_{ik}\right), \tag{4.3}$$

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 (4.1) if individuals expect to sort into occupations with higher expected returns. This parameter is identified from the data on earnings beliefs and subjective probabilities of choosing occupation k. A consistent estimator of $TT(k)$ is then given by:

$$\widehat{TT}(k) = N^{-1} \sum_i \widehat{\omega}_{ik}^{TT} \Delta Y_{ik}, \tag{4.4}$$

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

It follows that the *ex ante* treatment effect on the treated is identified if agents have rational expectations about their future occupational choices. Indeed, under this assumption, Equation (4.3) still holds after replacing the weights ω_{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:

$$TT(k) = E\left(\Delta Y_{ik} | d_{ik} = 1\right), \tag{4.5}$$

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

Table 4: *Ex Ante* Treatment Effects by Occupation (Earnings in 2009 dollars)

	Phase 1			Phase 3		
Occupation	TT	TUT	ATE	TT	TUT	ATE
Science	29,820	20,674	22,320	61,879	49,942	51,968
	(4,786)	(3,246)	(3,121)	(14,337)	(6,070)	(5,786)
Health	117,700	57,808	68,065	119,588	35,131	54,224
	(18,802)	(6,879)	(8,575)	(26,631)	(5,352)	(8,897)
Business	104,224	84,201	89,533	220,938	82,518	139,815
	(14,664)	(8,052)	(8,480)	(28,211)	(10,380)	(17,843)
Government	26,733	25,753	25,875	18,008	9,524	10,046
	(7,162)	(3,918)	(3,970)	(3,932)	(1,979)	(1,927)
Law	110,423	84,343	88,750	54,175	66,990	65,995
	(20,033)	(10,595)	(11,280)	(16,723)	(8,168)	(7,763)

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}).

effect on the untreated, $TUT(k)$, is estimated similarly, after replacing p_{ik} by $1 - p_{ik}$.[15]

Table 4 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.[16]

Consistent with positive selection on expected gains across occupations, the estimated TUT's in Table 4 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*

[15] 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.

[16] Table A.6 in Appendix A.2 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.

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* earnings for occupation k to be $SE(k) := E\left(\omega_{ik}^{TT} Y_{ik}\right)$ for which a consistent estimator is:

$$\widehat{SE}(k) = N^{-1} \sum_{i} \widehat{\omega}_{ik}^{TT} Y_{ik}. \tag{4.6}$$

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:

$$\Delta\widehat{SE}(k) \quad = \quad \widehat{SE}(k) - \widehat{SE}(1) \tag{4.7}$$

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.We can then compare these estimates with the average *ex ante* treatment effects to quantify how much of the selected *ex ante* earnings premium $\Delta\widehat{SE}(k)$ is due to selection.

Panel A of Table 5 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 4 in that the gap between TT and TUT is especially large in Health and small in Business.

Panel B of Table 5 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.

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

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

$$f_{TE,k}^{Treated}(u) = \omega_{ik}^{TT}(u) \cdot f_{TE,k}(u), \tag{4.8}$$

Table 5: 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 ($N = 113$).

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.[17] The following plug-in estimator:

$$\widehat{f}_{TE,k}^{Treated}(u) = \widehat{\omega}_{ik}^{TUT}(u) \cdot \widehat{f}_{TE,k}(u), \tag{4.9}$$

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 (4.9).

Figures 1, 2, and 3 plot the densities of the *ex ante* treatment on the treated and treatment on the untreated for Government, Health, and Business occupations, respectively.[18] (The distributions of the *ex ante* treatment effects for Science and Law are displayed in Figures A.1 and A.2 in Appendix A.3.) 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.[19] 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.

[17] 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.

[18] 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.

[19] 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).

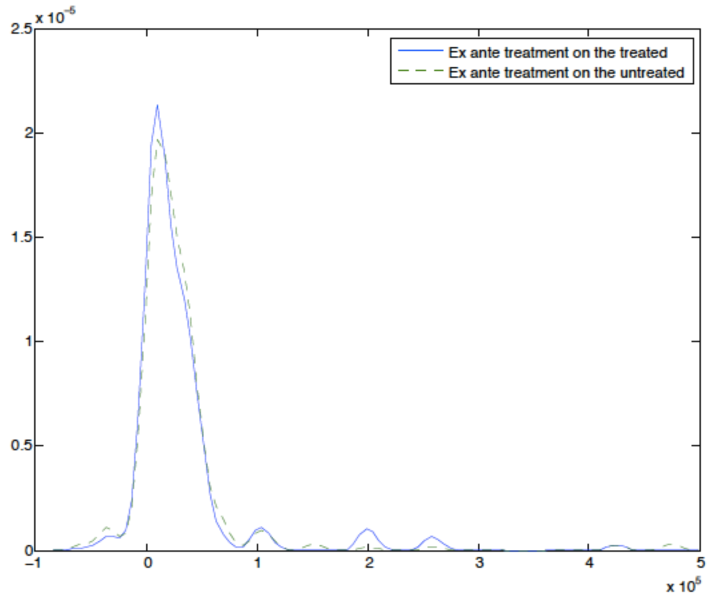

Figure 1: Densities of *Ex Ante* Treatment Effects on the Treated & Untreated: Government

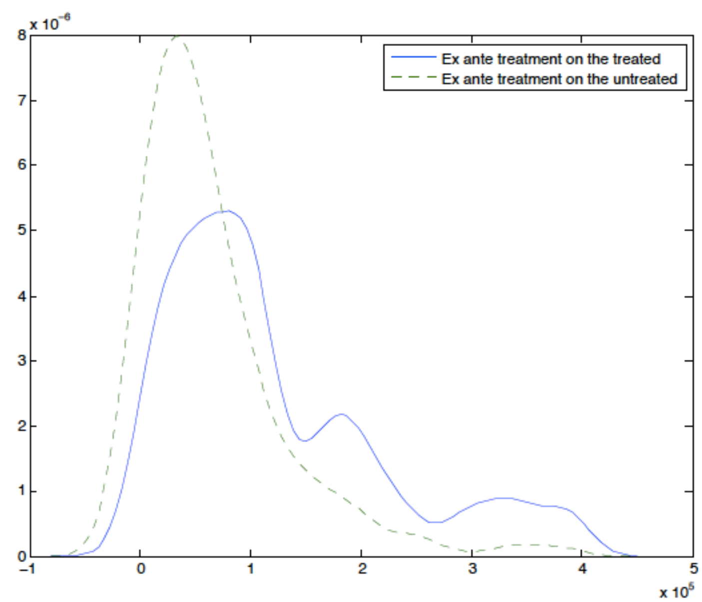

Figure 2: Densities of *Ex Ante* Treatment Effects on the Treated & Untreated: Health

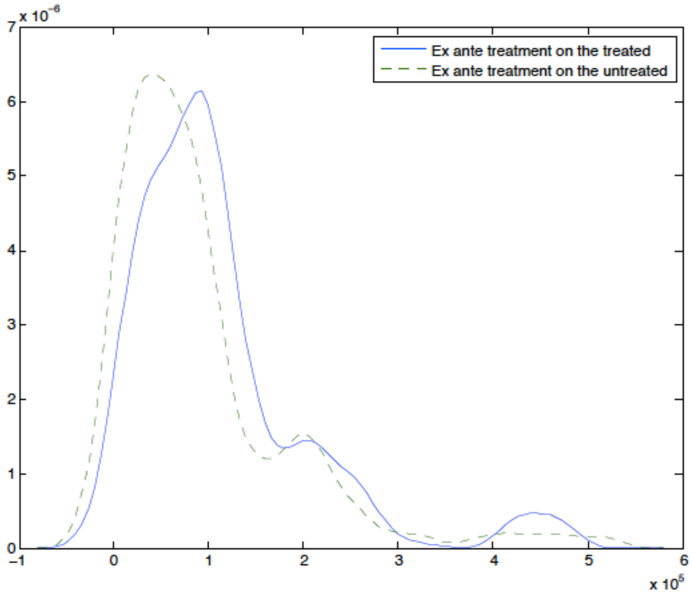

Figure 3: Densities of *Ex Ante* Treatment Effects on the Treated & Untreated: Business

5 Occupational choice and sorting on expected earnings

The findings in the preceding sections 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 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 $K = 6$ mutually exclusive occupations. Let $d_{ijk} = 1$ if occupation k is chosen, and zero otherwise. The values for d_{ijk} 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 ε_{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, α_k, and normalize α_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, and estimate a conditional logit where we substitute in for y_{ijk} using the Phase 1 beliefs on earnings. Namely, we set $y_{ijk} = \ln [Y_i(j, k, 1)]$.[20] Estimates for the β parameter in (5.2) are displayed in Column (1) of Table 6, both for the full sample and the sample excluding seniors.[21] 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 ε_{ijk}'s, is known by individual i at that time, but some is not known to either the researcher or individual i back at Phase 1. In this setting, it is the lack of knowledge about this latter part of ε_{ijk} that makes individuals uncertain about which occupation is best for them, and why they do not just report ones and zeros for the $[p_i(j, k, 1)]$'s.

To acknowledge this source of individuals' uncertainty, let

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

[20] 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 also 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.

[21] 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 indeed be absorbed by the occupation dummies α_k, and the conditional logit would therefore still consistently estimate the earnings coefficient β in this case.

where the ϕ_{ijk}'s are known to i at Phase 1, but the ζ_{ijk}'s are unknown. Assuming that ζ_{ijk} has a standard Type 1 extreme value distribution and that students form rational expectations over their future choice of occupation, we can invert the self-reported probabilities to obtain:[22]

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

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

We first estimate β 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 et al. (2010), and then estimate the flow utility parameters using a least absolute deviation (LAD) estimator.[23] Results of the LAD estimation of (5.4) are given in Column (2) of Table 6. The estimate of β 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 us to control for major-occupation dummies even with our sample size. That is, we can replace the α_k's with α_{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 6. 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 fixed effects in addition to the major-occupation interactions.[24] The

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

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

[24]See Wiswall & Zafar (2017) who provide evidence from NYU students that preferences for non-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 elic-

Table 6: 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.

results for this final and preferred specification, where the individual-occupation fixed effects are eliminated by within-transformation, are reported in Column (4) of Table 6, again using occupational choice probabilities and expected earnings elicited in Phase 1. (The standard errors were clustered at the individual-occupation level.) Note that the multiple observations per respondent help to produce more precise estimates of β. And, while the estimate of β 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 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

iting 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 are robust to such unobserved heterogeneity (see Table 8).

6. 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 β ($\widehat{\beta} = 0.953$ 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 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 et al., 2012; Long et al., 2015; Wiswall & Zafar, 2015; and Altonji et al., 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 7, we repeat the *ex ante* occupational choice analyses found in Table 6 using these more recent beliefs. Specifically, 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 [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 6 provides evidence that beliefs about future choice

Table 7: 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.

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 8 reports the LAD estimation results which correspond to the specification in (5.4), but here also 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 preferences for occupation- and 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 7 (Column 3, 0.914) and in Table 6 (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 together, these results provide yet further evidence that individuals sort across occupations based on their expected earnings.

Table 8: Changes in subjective probabilities of choosing occupations

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

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. Understanding the role played by pecuniary versus non-pecuniary components in sorting across occupations has long been a been a question of interest in labor economics (see, in particular, Heckman & Sedlacek, 1990 in the context of sectoral choices). We offer a new perspective on this question by using 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 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.[25]

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.

In Table 9 we display estimates of features of the distribution of the *ex ante* earnings losses due to individuals making their occupational choices based on factors other than expected

[25]Related work by D'Haultfoeuille & 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 measures of subjective expectations about future returns, it does rely on stronger assumptions concerning the non-pecuniary factors. See also Eisenhauer et al. (2015), who use exclusion restrictions between monetary returns and non-pecuniary factors to separately identify these two components in the absence of subjective expectations.

earnings. Column (1) of the table first 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, $\overline{Y}_i(j^c, 3)$, which are computed by weighting the elicited expected earnings in each occupation 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 differences, or gaps, 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* earnings loss associated with not choosing the highest paying occupation.

Panel A of Table 9 shows the distributions of $Y_i^{max}(j^c, 3)$, $\overline{Y}_i(j^c, 3)$, and $G_i(j^c, 3)$ for the full sample of respondents to the Phase 3 survey. The average gap of slightly less than \$30,000 represents about fourteen percent of the maximum earnings individuals expect to receive. Note that these estimates are lower bounds on *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.[26] 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 $\overline{Y}_i(j^c, 3)$, reflecting the fact that rank invariance does not hold for these two distributions.

Panel B of Table 9 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. for whom $Y_i^{max}(j^c, 3) > \overline{Y}_i(j^c, 3)$. Note that these individuals as a whole tend to have lower maximum earnings than those who are certain of choosing their 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 annual earnings ten years after college as a result of not choosing their

[26] 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 9: 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

(*ex ante*) income-maximizing occupation, or a little over 21% of their maximum expected earnings. The distribution is skewed, however, with a smaller but still quite substantial median loss of about $24,000. Overall, these results are consistent with non-monetary factors playing an important 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 expected monetary gains and non-pecuniary factors in the choice of occupation.

The distributions of the *ex ante* returns for particular occupations are 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. 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, and consistent with a generalized Roy model of occupational choice, non-pecuniary components also play an important role, with individuals expecting to give up sizable amounts of money as a result of not choosing the income-maximizing 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, J., Arcidiacono, P. & Maurel, A. (2016), The analysis of field choice in college and graduate school: Determinants and wage effects, *in* E. Hanushek, S. Machin & L. Wößmann, eds, 'Handbook of the Economics of Education', Vol. 5, Elsevier, pp. 305–396.

Antonovics, K. & Golan, L. (2012), 'Experimentation and job choice', *Journal of Labor Economics* **30**, 333–366.

Arcidiacono, P. (2004), 'Ability sorting and the returns to college major', *Journal of Econometrics* **121**, 343–375.

Arcidiacono, P. (2005), 'Affirmative action in higher education: How do admission and financial aid rules affect future earnings?', *Econometrica* **73**, 1477–1524.

Arcidiacono, P., Aucejo, E., Fang, H. & Spenner, K. (2011), 'Does affirmative action lead to mismatch? A new test and evidence', *Quantitative Economics* **2**, 303–333.

Arcidiacono, P., Hotz, V. J. & Kang, S. (2012), 'Modeling college major choices using elicited measures of expectations and counterfactuals', *Journal of Econometrics* **166**, 3–16.

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

Beffy, M., Fougere, D. & Maurel, A. (2012), 'Choosing the field of studies in postsecondary education: Do expected earnings matter?', *Review of Economics and Statistics* **94**, 334–347.

Blass, A. A., Lach, S. & Manski, C. F. (2010), 'Using elicited choice probabilities to estimate random utility models: Preferences for electricity reliability', *International Economic Review* **51**, 421–440.

Cardell, N. S. (1997), 'Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity', *Econometric Theory* **13**, 185–213.

Carneiro, P. M., Hansen, K. T. & Heckman, J. J. (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, F. & Heckman, J. J. (2007), 'Identifying and estimating the distributions of *Ex Post* and *Ex Ante* returns to schooling', *Labour Economics* **14**, 870–893.

Cunha, F. & Heckman, J. J. (2008), 'A new framework for the analysis of inequality', *Macroeconomic Dynamics* **12**, 315–354.

Cunha, F., Heckman, J. J. & Navarro, S. (2005), 'Separating uncertainty from heterogeneity in life cycle earnings', *Oxford Economic Papers* **57**, 191–261.

Delavande, A. (2008), 'Pill, patch, or shot? subjective expectations and birth control choice', *International Economic Review* **49**, 999–1042.

Delavande, A. & Zafar, B. (2018), University choice: The role of expected earnings, non-pecuniary outcomes and financial constraints.

D'Haultfoeuille, X. & Maurel, A. (2013), 'Inference on an extended roy model, with an application to schooling decisions in france', *Journal of Econometrics* **174**, 95–106.

Dominitz, J. & Manski, C. F. (1996), 'Eliciting student expectations of the returns to schooling', *Journal of Human Resources* **31**, 1–26.

Dominitz, J. & Manski, C. F. (1997), 'Using expectations data to study subjective income expectations', *Journal of the American Statistical Association* **92**, 855–867.

Eisenhauer, P., Heckman, J. J. & Vytlacil, E. (2015), 'The generalized Roy model and the cost-benefit analysis of social programs', *Journal of Political Economy* **123**, 413–443.

Heckman, J. J., Humphries, J. E. & Veramendi, G. (Forthcoming), 'Returns to education: The causal effects of education on earnings, health and smoking', *Journal of Political Economy* .

Heckman, J. J. & Sedlacek, G. L. (1990), 'Self-selection and the distribution of hourly wages', *Journal of Labor Economics* **8**, S329–S363.

Hurd, M. (2009), 'Subjective probabilities in household surveys', *Annual Review of Economics* **1**, 543–562.

Hurd, M. & McGarry, K. (1995), 'Evaluation of the subjective probabilities of survival in the hrs', *Journal of Human Resources* **30**, S268–S292.

Hurd, M. & McGarry, K. (2002), 'The predictive validity of subjective probabilities of survival', *Economic Journal* **112**, 966–985.

Kang, S. (2009), Eliciting expectations and counterfactuals about college major choice: Data documentation. Working paper.

Keane, M. & Wolpin, K. I. (1997), 'The career decisions of young men', *Journal of Political Economy* **105**, 473–522.

Kirkeboen, L. J., Leuven, E. & Mogstad, M. (2016), 'Field of study, earnings, and self-selection', *Quarterly Journal of Economics* **131**, 1057–1111.

Lee, K. (2015), 'The continuing decline in lsat takers'.
URL: *http://abovethelaw.com/2015/08/the-continuing-declining-in-lsat-takers/*

Long, M. C., Goldhaber, D. & Huntington-Klein, N. (2015), 'Do students' college major choices reflect changes in wages?', *Economics of Education Review* **49**, 1–14.

Manski, C. F. (1993), Adolescent econometricians: How do youths infer the returns to schooling?, *in* C. Clotfelter & M. Rothschild, eds, 'Studies of Supply and Demand in Higher Education', Chicago: University of Chicago Press, pp. 43–57.

Manski, C. F. (2004), 'Measuring expectations', *Econometrica* **72**, 1329–1376.

Miller, R. A. (1984), 'Job matching and occupational choice', *Journal of Political Economy* **92**, 1086–1120.

Siow, A. (1984), 'Occupational choice under uncertainty', *Econometrica* **52**, 631–645.

Stinebrickner, T. & Stinebrickner, R. (2014), 'A major in science? initial beliefs and final outcomes for college major and dropout', *Review of Economic Studies* **81**, 426–472.

Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press.

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

van der Klaauw, W. & Wolpin, K. I. (2008), 'Social security and the retirement and savings behavior of low-income households', *Journal of Econometrics* **145**, 21–42.

Wiswall, M. & Zafar, B. (2015), 'Determinants of college major choice: Identification using an information experiment', *Review of Economic Studies* **82**, 791–824.

Wiswall, M. & Zafar, B. (2016), Human capital investments and expectations about career and family. NBER Working Paper No. 22543.

Wiswall, M. & Zafar, B. (2017), 'Preference for the workplace, investment in human capital, and gender', *Quarterly Journal of Economics* **133**, 457–507.

Zafar, B. (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	***Public Policy***
Music	Environmental Science and Policy
Philosophy	Political Science
Religion	Public Policy Studies
Spanish	
Theater Studies	
Visual Arts	

A.2 Additional Tables

Table A.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.

Table A.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).[27] Note that each cell contains averages of the responses by each of the 173 students. Expected incomes exhibit sizable variation both across majors and occupations. For instance, majoring in the natural sciences or engineering is perceived to lead to higher earnings in Science and Health careers, while expected earnings in Business are, on average, higher for economics majors. Differences across occupations are even starker. In particular, average expected incomes are lowest for a career in Education and generally highest for a career in Law, with the exception of natural sciences and economics majors, for which expected incomes are highest for Health and Business occupations, respectively.

Turning to the choice of occupation, Table A.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. Overall, the likelihood of working in the various occupations appear to be selectively different across individuals, even after conditioning on a college major.[28]

Finally, Table A.4 in Appendix A.2 reports the prevalence of students reporting that the probability they would choose a particular occupation was zero for each major-occupation combination ($p_i(j, k, 1) = 0$).[29] While some combinations display a large share of zero

[27] 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.

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

[29] 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

Table A.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?"

Table A.2: Mean of Phase 1 expected incomes for different major/occupation combinations 10 years after graduation (Annual Incomes, in dollars)

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

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

Table A.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.

subjective probabilities, the shares are well below one, suggesting that particular majors do not rule out certain occupations for all individuals.

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 A.4: 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.5: *Ex Ante* Treatment Effects of Occupations by Counterfactual Major (Earnings in 2009 dollars)

		Counterfactual Major					
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.6: 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.7: 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.3 Additional Figures

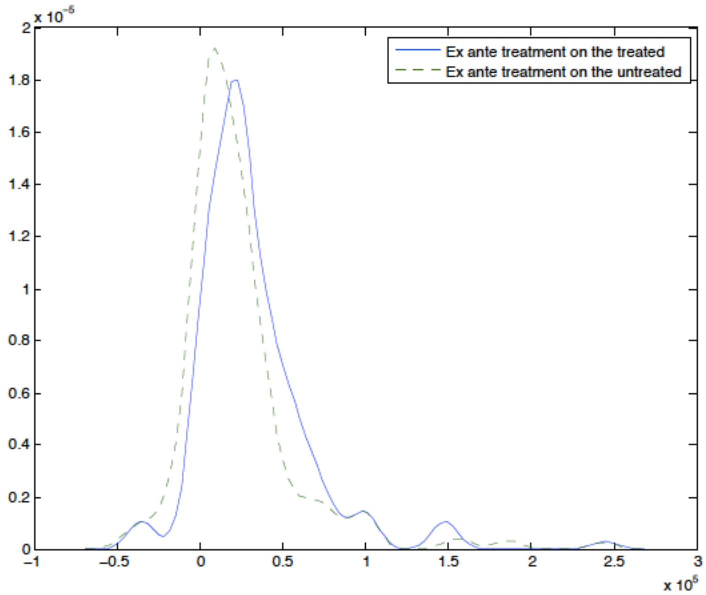

Figure A.1: Densities of *Ex Ante* Treatment Effects: Science

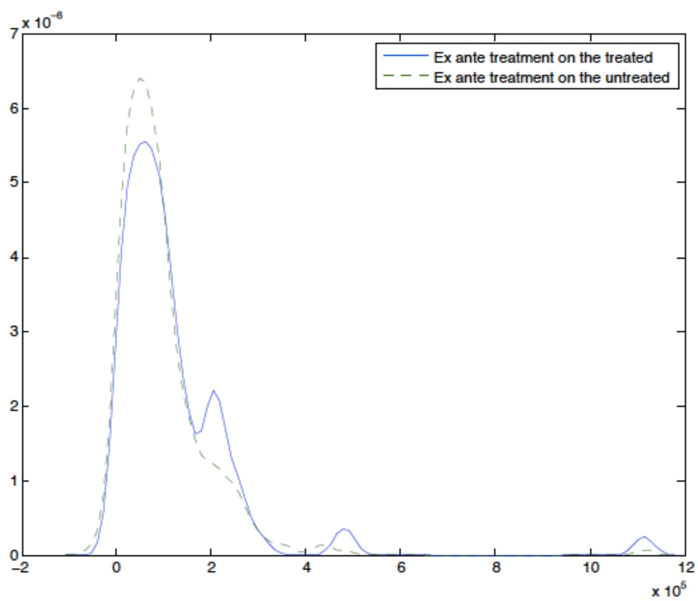

Figure A.2: Densities of *Ex Ante* Treatment Effects: Law