Recovering Ex Ante Returns and Preferences for Occupations using Subjective Expectations Data

Preliminary

Peter Arcidiacono [*] V. Joseph Hotz[†] Arnaud Maurel[‡]
Teresa Romano [§]

March 19, 2014

Abstract

We use data on subjective expectations of outcomes from counterfactual choices to recover *ex ante* treatment effects as well as the non-pecuniary benefits associated with different treatments. The particular treatments we consider are the choice of occupation. By asking individuals about potential earnings associated with counterfactual choices of college majors and occupations, we can recover the full distribution of the *ex ante* monetary returns to particular occupations, and how these returns vary across majors. In particular, the elicited choice probabilities allow us to quantify the importance of sorting on *ex ante* monetary benefits when choosing an occupation. By linking subjective expectations to a model of occupational choice, we can then examine how individuals tradeoff their preferences for particular occupations with the corresponding monetary rewards. While sorting across occupations is partly driven by the *ex ante* monetary returns, sizable differences in expected earnings across occupations remain after controlling

[*]Duke University and NBER. We thank seminar participants at the London School of Economics, Stanford University and Washington University in St. Louis, as well as participants at the NBER Labor Studies 2014 Spring Meeting for helpful comments.

[†]Duke University, NBER and IZA.

[‡]Duke University, NBER and IZA.

[§]Duke University.

for selection on monetary returns, which in turn points to the existence of substantial compensating differentials.

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, for example Manski (1993) and Dominitz & Manski (1996,1997), subjective expectations are now being used in the estimation of structural dynamic models (see, e.g., Delavande, 2008, van der Klaauw & Wolpin, 2008, 2012).[1] Collecting data on subjective expectations makes it possible to estimate forward-looking models without making strong assumptions about how individuals form their beliefs about potential outcomes along different choice paths.

Relatively new to the literature is (i) the elicitation of the probabilities of taking particular courses of actions in the future and (ii) expectations about potential future outcomes corresponding to these counterfactual choices, which covers beliefs off the individual's actual choice path. In this paper, we use data on future choice probabilities as well as subjective expectations about outcomes both on and off the individual's choice path to recover the expected benefits as well as subjective costs associated with different treatments, and tell apart the role played by these two components in selection into treatment. Even though the proposed approach can be applied to potential outcome models in general, in this paper we consider the particular context of monetary returns to different occupations (for different college majors), as well as how individuals trade off the *ex ante* monetary returns with their non-pecuniary preferences for particular college majors and occupations. As recently emphasized in the literature (see, e.g., Cunha et al., 2005, Cunha & Heckman and Heckman & Navarro, 2007), *ex ante* (in comparison with *ex post*) monetary returns are of key interest since they correspond to what agents act on. The approach that we develop in this paper makes it possible to identify those *ex ante* returns, along with the non-pecuniary factors affecting the choice of occupation, while being agnostic on the information set of the agents and in the absence of exclusion restrictions between monetary returns and non-pecuniary factors.[2]

[1]See Manski (2004) for a survey of the literature. See Pantano & Zheng (2010) on using subjective expectations data to recover unobserved heterogeneity.

[2]In a recent work, D'Haultfoeuille & Maurel (2013) investigate the relative importance of *ex ante* monetary returns versus non-pecuniary factors in the decision to attend college. Their approach, which

Overall, there are large differences in the earnings of college graduates both across majors and occupations. For instance, data from the American Community Survey (2009-2010) reveal that those who majored in engineering earn as much as 77% more than those who majored in the humanities. To the extent that a sizable fraction of college graduates work in an occupation which does not match their major, those earnings differentials across majors mask the existence of substantial within-major dispersion (see, e.g., Kinsler & Pavan, 2012). However these earnings differentials are computed from those who chose particular college majors and occupations, thus telling us very little about what the individual would expect to earn had the individual pursued a counterfactual occupation or graduated with a different major. It follows that this type of observational data, which has been used in most the literature on college major and occupational choices (see Altonji et al., 2012, for a recent review), is also uninformative by itself as to how much individuals would need to be compensated for pursuing a different career path.

In this paper, we use elicited beliefs from male undergraduates at Duke University to quantify the importance of sorting across occupations on *ex ante* monetary returns versus preferences. This unique dataset contains student expectations regarding the probability of working in different occupations as well as their expected income in each of the occupations where the period of reference is ten years after they graduate.[3] These occupation probabilities and expected incomes were asked not only for the major the individual chose but also for counterfactual majors, making it possible to disentangle both the monetary returns from different majors in different occupations as well as how attractive working in particular occupations is with different majors. By doing so, we add to a growing set of papers using subjective expectations data to tell apart the role played by monetary returns versus non-pecuniary preferences in college major and

can be used in the absence of subjective expectation data, requires imposing stronger restrictions on the non-pecuniary factors. See also Eisenhauer et al. (2012), who use exclusion restrictions between monetary returns and non-pecuniary factors to tell apart those two components.

[3]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) and van der Klaauw (2012).

occupational choices (Betts, 1996, Zafar, 2011,2013, Arcidiacono et al., 2012, Wiswall & Zafar, 2012, Stinebrickner & Stinebrickner, 2013, and Osman, 2013).

The data allow us to identify both the *ex ante* treatment effects of particular occupations on earnings, for any given college major, as well as the *ex ante* treatment effects of particular majors on the probabilities of working in any given occupation. Even though we do not observe the actual occupations chosen by the individuals, we show that subjective expectation data on occupational choice probabilities can be used to recover the *ex ante* treatment effects of a given occupation (relative to a reference occupation) for the subpopulation of individuals who will end up working in that occupation (*ex ante* treatment effect on the treated). Taking the initial major as given, the *ex ante* treatment effect on the treated for a given occupation j is simply computed by weighting the reported earnings differences between occupation j and the reference by the probability the individual reports that he will work in occupation j (over the average declared probability of working in occupation j). *Ex ante* treatment effect on the untreated are obtained similarly, by using the declared probability that the individual will not work in occupation j. Importantly, our data allows us to go beyond these average effects and investigate the heterogeneity across individuals by estimating the full distributions of the *ex ante* treatment effects of working in any given occupation j relative to education, given the initial college major choice. Data on counterfactual occupational choice probabilities also allows us to recover the distribution of the *ex ante* treatment effects on the treated and untreated subpopulations.

The results reveal substantial differences in terms of expected earnings across majors as well as occupations. Treating the education occupation as the baseline, the *ex ante* treatment on the treated ranges from 25% higher earnings (government) to 89% higher earnings (health) ten years after graduation. Consistent with sorting across occupations being partly driven by expected monetary returns, the *ex ante* returns are generally higher for the treated than for the untreated, suggesting positive selection into occupations. Consistent with the existence of occupation-specific human capital accumulated within each major, we also document the existence of a substantial degree of heterogeneity in the *ex ante* returns for each occupation, depending on college major. For example, public policy majors who anticipate entering a health career expect a 38%

premium (relative to a career in education), while natural sciences majors expect a 117% premium for a health career.

We next link the subjective expectations data to a model of occupational choice where individuals are uncertain over their preferences for particular occupations in the future. This simple framework allows us to link the subjective data on expected earnings and choice probabilities with the non-pecuniary preferences. Specifically, under standard assumptions on unobserved preferences, those terms will have continuous support implying that perceived occupation probabilities should be bounded away from zero and one. However, in our data, some individuals do report zero probabilities of pursuing a particular occupation given a particular major. We reconcile our framework with the data by modeling the resolution of preference uncertainty as costly. Namely, we assume that individuals will only pay the cost to find out additional information about a given occupation if their expected benefits of doing so are sufficiently high. In estimation, we then follow Hotz & Miller (1993) and Berry (1994) and invert the perceived choice probabilities, taking into account the selection introduced by costly information acquisition, to recover preferences over occupation-major combinations.

The coefficient on our income measure then allows us to calculate compensating differentials for particular occupations, and how these compensating differentials vary for those who pursue different majors. Overall, our results are consistent with the existence of fairly large compensating differentials across occupations, which vary substantially across majors. For instance, while public policy majors would have to receive a premium of 137.8% to pursue a career in education rather than in government, the opposite is true for those with a major in the humanities, who would have to receive a premium of 73.7% to pursue a governmental career. Aside from the complementarities of preferences between different majors and occupations, we show that the large compensating differentials associated with major-occupation pairs may also be partly explained by search frictions, whereby job offer arrival rates for each occupation vary across college majors. In any case, our results provide clear evidence that majors have a substantial influence on occupations well beyond their impact on earnings.

The rest of the paper proceeds as follows. In section 2 we discuss the survey data used in the paper. Section 3 shows how to obtain *ex ante* treatment effects given the survey data with section 4 giving the estimated treatment effects. We then link the subjective occupational choice probabilities and expected incomes with a model of occupational choice in section 5. Estimates of the model and the corresponding implications in terms of compensating differentials and search frictions are presented in section 6. Finally, we conclude the paper in section 7.

2 Data

We use data collected on a sample of male undergraduate students at Duke University between February and April 2009. Gender was the only restriction on sample recruitment; students from any major, class, or race were eligible to participate in the survey. Sample members were recruited by posting flyers about our study around the Duke campus. Surveys were administered on computers in a designated room in Duke's Student Union.[4] All 173 students who completed the survey were paid $20.[5]

This is the same data as the one used in Arcidiacono et al. (2012). That paper provided many descriptive statistics on how majors, occupations, and earnings were related and we refer the reader to that paper for an overview of the data. We report in Table 1 a descriptive overview of our sample, compared with the overall male undergraduate population at Duke. One can see from Table 1 that our sample corresponds fairly closely to the Duke male undergraduate student body, even though it includes slightly more Asians and fewer Latinos and Blacks. It also appears that a higher percentage of our sample receives some financial aid than is the case in the Duke student body, although the 22.0% figure for the student body is based on aid provided by Duke, whereas the higher percentage of students receiving financial aid (40.5%) is likely due to the fact that our survey asked about receipt of financial aid, regardless of source. Finally, our sample is slightly tilted towards upper-classmen.

Distinctive to this paper is our focus on occupations as choices, as the previous paper treated occupations as lotteries. Evidence that individuals are viewing occupations as

[4]The questionnaire which was used in the survey is discussed further in Kang (2009).

[5]We drop from our analysis five individuals who reported that they would choose a career with certainty for each major, resulting in a final sample of 168 students.

Table 1: Sample Descriptive Statistics

	Sample	Duke Male Student Body
Current/Intended Major:		
Science	18.5%	14.8%
Humanities	9.5%	9.4%
Engineering	19.1%	20.7%
Social Science	18.5%	18.8%
Economics	20.2%	18.0%
Public Policy	14.3%	18.0%
Class/Year at Duke:		
Freshman	20.2%	
Sophomore	20.2%	
Junior	27.4%	
Senior	32.1%	
Characteristics of Students:		
White	66.7%	66.0%
Asian	19.6%	16.6%
Latino	4.8%	8.3%
Black	4.2%	5.9%
Other	4.8%	3.0%
U.S. Citizen	95.2%	94.1%
Receives Financial Aid	40.5%	22.0%
Sample Size	168	

choices can be found in Table 2. Table 2 reports, for each college major, the expected earnings computed using the subjective probabilities of entering each career, as well as the expected earnings under the counterfactual assumption that careers were randomly assigned.[6] For the random assignment case, we use the population probabilities of

[6]In our sample, only 1.57% 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

choosing each career for those in the same major. For all majors, students expected earnings are higher given their reported probabilities of sorting into careers relative to if they were randomly assigned. This pattern points to the existence of sizable gains to sorting, consistent with the individuals pursuing their comparative advantage when choosing a career.[7]

Table 2: Expected Earnings for Careers (Annual Earnings, in dollars)

Major	Reported Probabilities	Random Assignment	Difference
Natural Science	169,385	144,710	24,675
Public Policy	180,350	154,823	25,527
Humanities	115,786	106,325	9,461
Economics	160,488	133,363	27,125
Engineering	125,578	115,413	10,165
Social Sciences	125,578	111,214	14,364

Table 3 reports the average subjective probabilities of working in each occupation, conditional on each major. While the subjective probabilities of entering each career vary substantially across majors, it is worth noting that none of the majors are concentrated into only one (or two) occupations. Besides, even for majors which appear to be more tied to a specific occupation, such as Business career for Economics majors, subjective probabilities exhibit a fairly large dispersion across individuals (see Figure 1). Overall, this stresses the importance of treating occupations as resulting from choices,

of log-earnings on major and occupation indicators, interaction between major and occupation, 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. Besides, one individual in our sample declared that he expected to earn $1,000 for some occupation-major combinations. We assume that this individual declared monthly rather than yearly incomes, and rescale his expected income accordingly.

[7] Since we are using subjective data, one might be concerned that these gains to sorting are partly driven by ex post rationalization. However, in the paper we focus on the question of sorting across occupations, which have not been effectively chosen by the students at the time of the survey. Furthermore, in our sample, 90% of the declared probabilities of working in a given occupation are smaller than 40%, so that ex post rationalization is unlikely to affect our results.

even after conditioning on college major. Finally, Table 4 reports the prevalence of zero probability reports, for each major and occupation.[8] While some combinations display a large share of zero subjective probabilities, such as, e.g., Economics or Public Policy and Science, as well as Engineering and Law, there is always, for any given major and occupation, a substantial fraction of students who do not rule out the prospect of pursuing a career in that occupation, and consequently report a non zero probability.

Table 3: Probability of different occupations conditional on major

	Probability of Occupation in:					
Major	Science	Health	Business	Government	Education	Law
Science	0.345	0.323	0.124	0.072	0.071	0.066
Humanities	0.076	0.121	0.230	0.149	0.231	0.194
Engineering	0.399	0.200	0.191	0.076	0.069	0.065
Social Sciences	0.095	0.145	0.246	0.192	0.131	0.191
Economics	0.058	0.078	0.512	0.159	0.064	0.129
Public Policy	0.055	0.116	0.229	0.320	0.077	0.203

The previous work with this data showed that the expectations over first year salaries matched well with data from Duke's career office. Since it is important that these expectations reflect actual underlying beliefs for the rest of the analysis, we attempt to assess how "reasonable" they are by comparing them with data form the American Community Survey (ACS). These comparisons will allow to see where Duke students believe they rank relative to the population of college graduates in particular major-occupation combinations.

We utilize data from the 2009-2011 ACS which contains data on wages, college major and current occupation. We limit the ACS sample to males between the ages of 29 and 35[9] with a reported major field for their college degree. Majors in the ACS

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

[9]The Duke respondents, on average, would be of age 32 ten years after graduation.

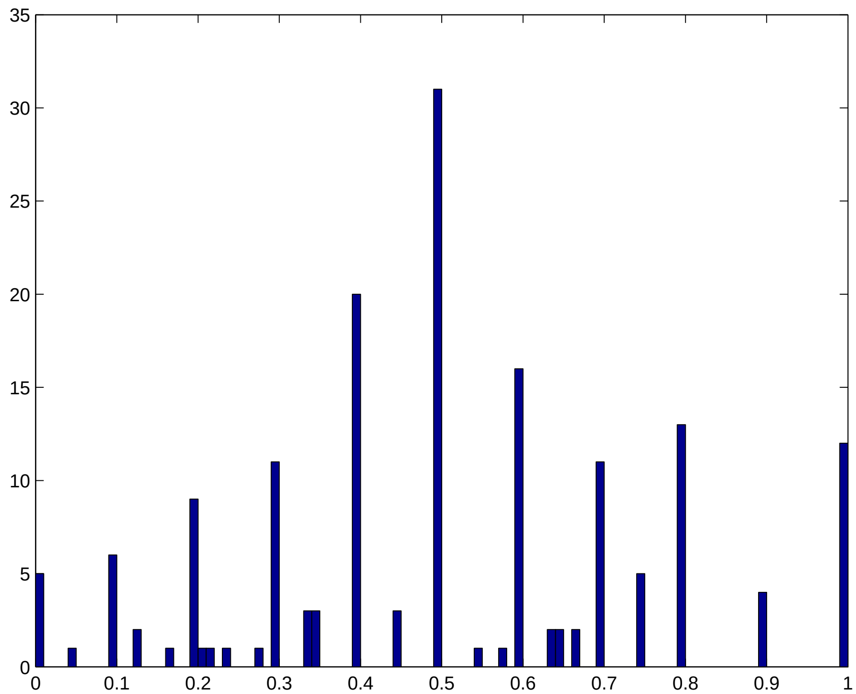

Figure 1: Distribution of subjective probabilities (economics major, business occupation)

Table 4: Zero probability reports

	Occupation					
Major	Science	Health	Business	Gov.	Ed.	Law
Science	3.57%	7.14%	27.98%	35.50%	39.29%	43.45%
Humanities	48.81%	34.52%	13.69%	18.45%	17.26%	16.67%
Engineering	7.14%	22.02%	20.83%	45.24%	47.02%	50.60%
Social Sciences	45.24%	30.95%	10.12%	13.10%	25.00%	17.86%
Economics	53.57%	49.40%	2.38%	17.26%	45.24%	28.57%
Public Policy	55.36%	36.31%	13.10%	3.57%	38.69%	11.90%

were categorized similarly to the Duke data. Several majors in the ACS are not offered at Duke; to the extent they clearly fell into a major category, they were included.[10] To construct occupations, matches between the occupations categories in the ACS and the career groupings in the Duke data were constructed.[11]

To compare the ACS to the Duke expected earnings, the following regression is estimated:

$$\ln(w_{ij}) = \alpha_j + \beta age_i + \nu_{ij} \tag{2.1}$$

where w_{ij} is the wage of person i with major j and α_j is a vector of dummy variables for each major j. This regression was estimated separately for each occupation. The regression results were then used to compute the average log wage at age 32 for each occupation conditional on major. The variance of the distribution of log wages was calculated from the regression residuals, enabling the comparison of the ACS income and Duke expected income distributions.

Table 5 gives the percentile of the ACS distribution of the median Duke student conditional on chosen major for each occupation. The percentiles tend to be above 50%

[10]Most of the excluded majors were health services majors or vocational majors such as construction services.

[11]Science, computing, and engineering classifications were coded as science and technology careers; medicine was coded as health careers; business and finance was coded as business career; education was coded as education careers; legal was coded as law careers. Workers classified as nonprofit works or local, state or federal employees were coded as government/nonprofit.

but below 90%, with most entries in the seventies and eighties. These predictions seem reasonable given that Duke is a highly selective institution, generally ranked in the top 10 according to U.S. News & World Report.

Table 5: Percentile of the ACS for the median Duke student conditional on chosen major

	Occupation					
Major	Science	Health	Business	Government	Education	Law
Sciences	87.61%	91.33%	93.33%	90.06%	87.07%	79.26%
Humanities	80.06%	90.31%	82.44%	83.58%	78.79%	68.25%
Engineering	58.08%	91.82%	68.29%	69.01%	76.03%	55.76%
Social Sciences	91.24%	94.82%	98.37%	86.68%	78.79%	56.29%
Economics	72.45%	93.68%	73.33%	70.23%	52.05%	79.46%
Public Policy	73.86%	85.52%	87.01%	70.58%	73.94%	78.40%

3 Ex ante treatment effects

In this section we outline the different types of *ex ante* treatment effect parameters we are interested in, and discuss how each of these parameters can be estimated using our subjective expectations data. We further discuss the estimation of the distributions of the *ex ante* treatment effects within the overall population, as well as the treated and untreated populations.

3.1 Occupation ex ante treatment effects

We define the *ex ante* treatment effects for particular occupations relative to education, which is chosen as a baseline occupation.[12] We label the education occupation as $k = 1$. We calculate the *ex ante* treatment on the treated for any given occupation

[12]We choose to use education as a baseline because the earnings in this occupation are not tied to choice of major.

$k \in \{2,4,4,5,6\}$, denoted by $TT(k)$, by weighting the differences in the reported earnings between occupation k and the baseline by the probability the individual reports that he will work in occupation k 10 years after graduation (over the average declared probability of working in occupation k). Namely:

$$TT(k) = \frac{\sum_i \sum_j' I(d_i = j')p_{ij'k}\left[w_{ij'k} - w_{ij'1}\right]}{\sum_i \sum_{j'} I(d_i = j')p_{ij'k}} \tag{3.1}$$

where $p_{ij'k}$ is the probability declared by individual i of choosing occupation k given major j', $I(d_i = j')$ is an indicator for whether i chose major j', and $w_{ij'k}$ the earnings expected by individual i in occupation k given major j'.

Similarly, we compute the *ex ante* treatment on the untreated for occupation k as:

$$TUT(k) = \frac{\sum_i \sum_j' I(d_i = j')(1 - p_{ij'k})\left[w_{ij'k} - w_{ij'1}\right]}{\sum_i \sum_{j'} I(d_i = j')(1 - p_{ij'k})} \tag{3.2}$$

Finally, the average *ex ante* treatment effect is given by:

$$ATE(k) = \frac{\sum_i \sum_{j'} I(d_i = j')\left[w_{ij'k} - w_{ij'1}\right]}{N} \tag{3.3}$$

where N is the sample size.

Note that these *ex ante* treatment effect parameters are computed based on chosen majors. We discuss in Subsection 3.2 the estimation of occupation *ex ante* treatment effects conditional on counterfactual majors.

3.2 Heterogeneity in ex ante treatment effects by chosen major

We can also calculate the occupation *ex ante* treatment effect parameters for those choosing particular majors. Namely, conditional on a chosen major j, we compute the *ex ante* treatment on the treated, treatment on the untreated and average *ex ante* treatment effect as follows:

$$TT(k|j) = \frac{\sum_i I(d_i = j)p_{ijk}\left[w_{ijk} - w_{ij1}\right]}{\sum_i I(d_i = j)p_{ijk}} \tag{3.4}$$

$$TUT(k|j) = \frac{\sum_i I(d_i = j)(1 - p_{ijk})\left[w_{ijk} - w_{ij1}\right]}{\sum_i I(d_i = j)(1 - p_{ijk})} \tag{3.5}$$

$$ATE(k|j) = \frac{\sum_i I(d_i = j)\left[w_{ijk} - w_{ij1}\right]}{\sum_i I(d_i = j)} \tag{3.6}$$

Given that we also elicit the subjective expectations for counterfactual majors, we can compute similarly (after replacing $I(d_i = j)$ by $I(d_i \neq j)$) the *ex ante* treatment effect parameters for those *not* choosing particular majors.

3.3 Distributions of ex ante treatment effects

Our data allows us to go beyond the average effects and estimate the distributions of the *ex ante* treatment effects of working in any given occupation k relative to education, given the initial college major choice. We can estimate those distributions for three different subgroups of interest, namely (i) the overall population, (ii) the treated subpopulation, and (iii) the untreated subpopulation.

First, the density of the distribution of the *ex ante* treatment effects on the overall population can be simply estimated with a kernel density estimator, using the fact that we directly observe the *ex ante* treatment effect for any individual in our sample. We denote the corresponding density by $f_{TE,k}(.)$ (and its estimator $\widehat{f_{TE,k}}(.)$).

Second, it follows from Bayes' rule that we can estimate the density of the distribution of the *ex ante* treatment effects on the treated subpopulation (denoted by $f_{TE,k}^{Treated}(.)$), as follows, for any scalar u:

$$\widehat{f_{TE,k}^{Treated}}(u) \quad = \quad \frac{\widehat{f_{TE,k}}(u) \times E(\sum_{j'} \widehat{I(d_i = j')} p_{ij'k}|TE = u)}{1/N \times \sum_i \sum_{j'} I(d_i = j') p_{ij'k}} \tag{3.7}$$

The conditional expectation term above can be simply estimated using a Nadaraya-Watson nonparametric regression estimator. Finally, the distribution of the *ex ante* treatment effects on the untreated can be estimated by replacing $p_{ij'k}$ by $1 - p_{ij'k}$ in the formula above.

4 Results: Ex ante treatment effects

4.1 Occupation treatment effects

Table 6 provides estimates of the three *ex ante* treatment effect parameters of occupations on earnings 10 years after graduation which correspond to the formulas (3.1)-(3.3) in Section 3. Relative to the baseline career of education, the average *ex ante* treatment effects range from $22,542 for science (30.6% of the mean expected earnings in education) to as much as $90,066 in law (122.3% of the mean expected earnings in education). Health, business and law careers all have very large earnings premia of over 94%, while those entering a science or government occupation expect a much smaller premium of 30.6% to 35.7% ten years after graduation. Consistent with sorting across occupations being partly based on comparative advantages, the *ex ante* treatment on the untreated effects show that, for all occupations, the untreated anticipate lower premia than the treated. The difference is particularly large for health occupations, which is almost two times smaller for those not anticipating a career in health compared to those who plan to enter a health related occupation. Interestingly, these sorting effects are much weaker for science, where the untreated anticipate to earn 70% as much than the treated, and negligible for government.

But, as stated in section 3.3, substantially more information is available in the data than just average effects. Namely, we can plot the full distributions of the treatment on the treated and the treatment on the untreated. Figures 4.2, 3, and 4 plot the full distributions for government, health, and business occupations respectively.

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. For business, the bottom end of the distribution suggests significant selection. But at the top end, the treated and untreated distributions are quite close. This suggests that, in this part of the distribution, there is a significant group of individuals who would do quite well in business–as well as the best group of the group treated–but whose preferences or expected earnings in other occupations lead them away from business. These results highlight the importance of moving beyond the average effects and looking at the full distribution of the *ex ante* returns.

4.2 Occupation treatment effects conditional on major

Table 7 shows substantial heterogeneity in the expected earnings premium for a given occupation by the student's college major. Notably, natural science majors expect on average a $136,450 premium for a health career relative to education, which is more than five times larger than the $24,670 premium expected by public policy majors who anticipate to enter this type of career. Examining some of the other average *ex ante* treatment effects, economics majors have the highest premium for business occupations, while engineering and natural science majors have the highest premia for science careers. Overall, these patterns are consistent with certain majors being closely tied to specific occupations. In particular, the major-occupation pairs that are typically thought of as being closely related to one another, such as economics and business, science and health, and engineering or natural science and science careers, still have the highest premia. Those patterns are both consistent with the accumulation of occupation-specific human capital within each major, and with selection effects, where, for instance, individuals who expect to be more productive in health are more likely to choose a science major.

Ex ante treatment on the untreated effects by student's major are still generally lower than the treatment on the treated effects. There are however, several exceptions; for instance, science careers have higher effects on the untreated in social sciences majors, while government careers have a higher effect on the untreated in the humanities and social sciences. The difference between the *ex ante* treatment on the treated effects and treatment on the untreated effects quantifies the importance of selection on the expected differences in occupation-major premia. The difference is positive but quantitatively small, for the majority of occupation-major pairs. However, selection into law by social sciences majors explains more than 40% of the major-occupation premium. Besides, while selection effects into government are, on average across all majors, virtually nonexistent, selection turns out to explain a large share of the earnings premia for science majors (around 50%).

Finally, Table 8 provides estimates of the three *ex ante* treatment effects by counterfactual non-chosen major. The treatment on the treated effects are again generally larger than the treatment on the untreated, with a few exceptions: engineering and economics majors with science careers, government occupations with economics and public policy majors, and law with humanities and public policy. It is worth noting that these *ex ante* treatment effect parameters also exhibit a substantial degree of heterogeneity across majors. Notably, expected premia for business (relative to education) are higher for economics major, while returns to science are higher for engineering and natural science majors. The fact that these types of complementarities between majors and occupations still hold when focusing on the majors which were *not* chosen by the individuals points to the accumulation of occupation-specific human capital within majors.[13]

Table 6: Ex Ante Treatment Effects (Annual Earnings, in dollars)

Occupation	TT	TUT	ATE	ATE share of education income
Science	30,040	20,903	22,542	30.6%
Health	117,770	59,241	69,556	94.4%
Business	101,720	83,740	88,562	120.2%
Government	26,740	26,214	26,282	35.7%
Law	116,590	85,159	90,066	122.3%

5 Linking subjective expectations to utilities

We model the choice of occupation as taking place in three stages. First, an individual enrolls in a given college major. Second, upon graduating from college, the individual can make costly decisions to acquire more information about the value of a set of particular occupations (conditional on the major chosen in the first stage). Finally, the

[13]See also Kinsler & Pavan (2012), who provide evidence from the Baccalaureate and Beyond Longitudinal Study that the human capital accumulated while in college is partly major-specific, which translates into higher wages when working in an occupation related to one's field of study.

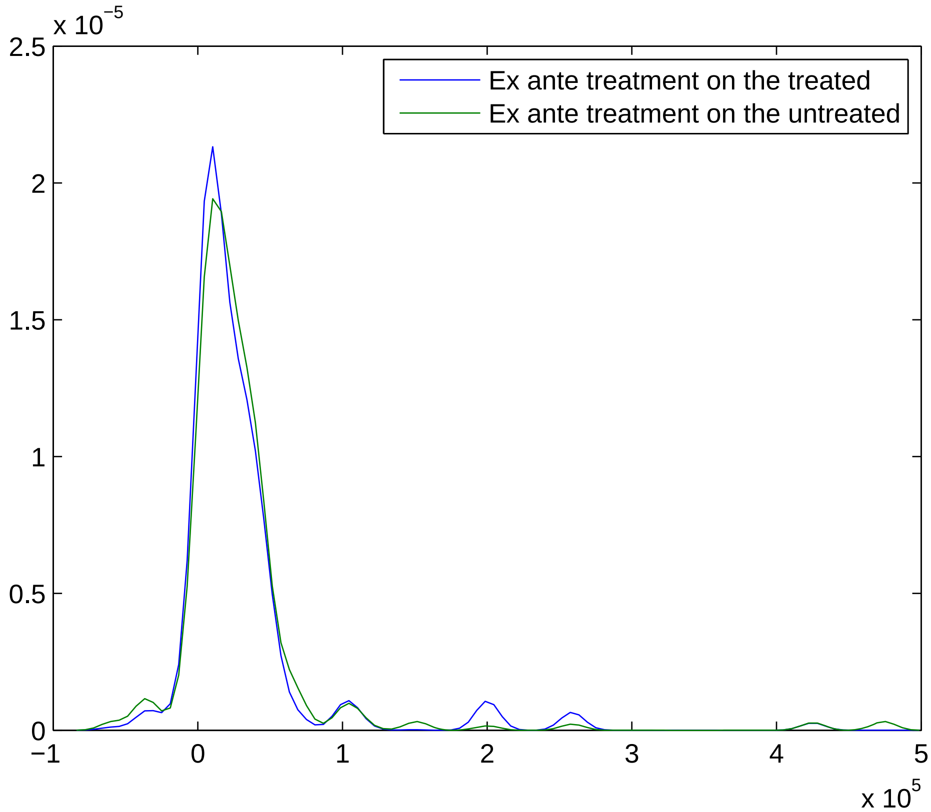

Figure 2: Distribution of ex ante treatment effects: government (annual earnings, in dollars)

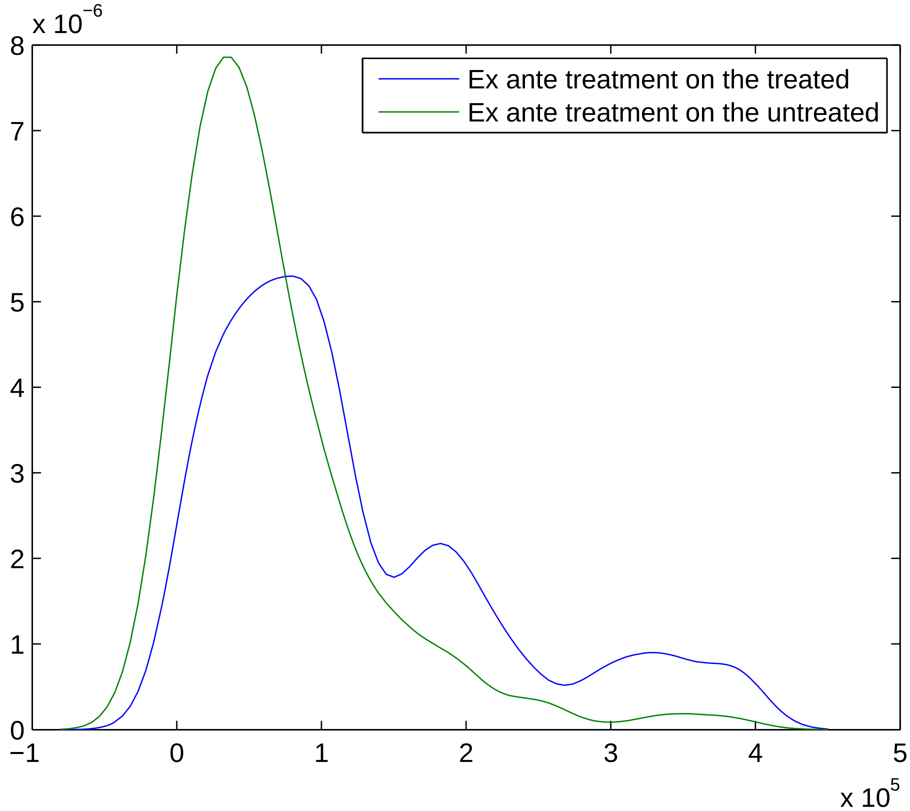

Figure 3: Distribution of ex ante treatment effects: health (annual earnings, in dollars)

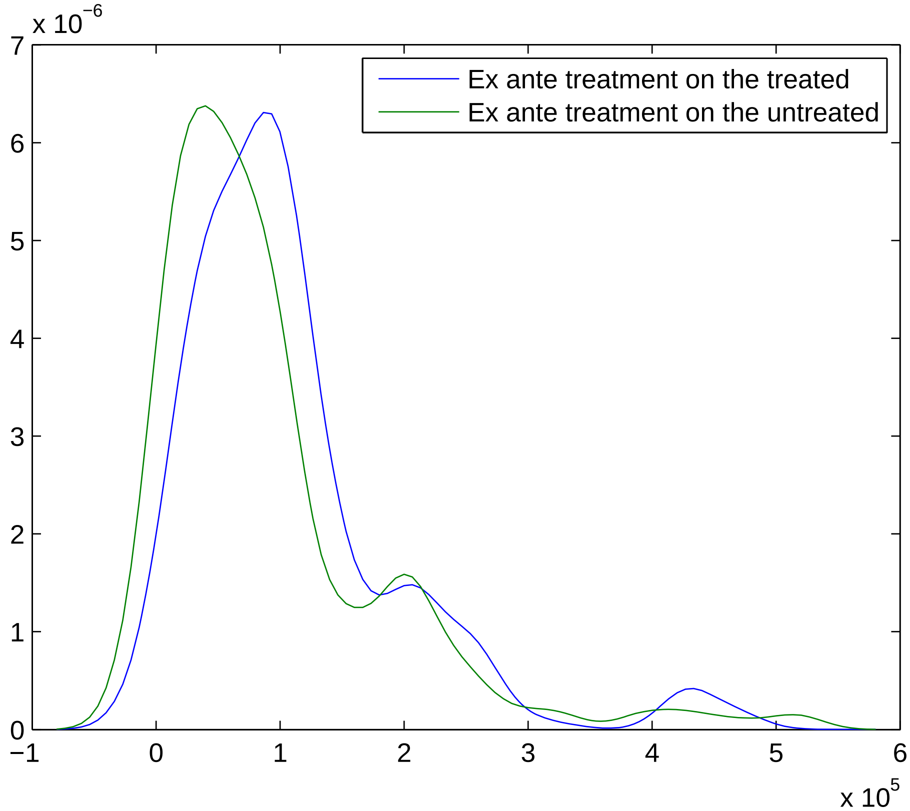

Figure 4: Distribution of ex ante treatment effects: business (annual earnings, in dollars)

Table 7: Heterogeneous Ex Ante Treatment Effects by Chosen Major (Annual Earnings, in dollars)

		Economics	Engineering	Humanities	Natural Sciences	Public Policy	Social Sciences
Science	TT	18,590	39,330	17,320	28,840	25,840	14,630
	TUT	17,680	27,420	6,460	36,040	17,030	19,600
	ATE	17,750	31,980	7,040	33,710	17,280	18,970
Health	TT	89,740	84,090	53,970	82,780	38,620	69,140
	TUT	60,440	57,480	59,170	106,830	23,740	55,750
	ATE	62,940	62,450	58,380	136,450	24,670	57,770
Business	TT	120,430	71,990	66,120	112,070	94,240	92,630
	TUT	120,450	70,810	56,640	107,140	67,580	75,490
	ATE	120,440	71,070	57,880	107,580	74,580	79,290
Government	TT	26,740	11,310	16,250	66,660	31,200	16,750
	TUT	25,770	11,820	23,880	33,670	25,440	36,310
	ATE	25,880	11,790	22,810	35,320	27,290	33,650
Law	TT	91,590	57,730	94,930	116,580	174,810	114,270
	TUT	93,630	67,550	62,090	88,930	138,780	63,000
	ATE	93,380	66,720	70,690	90,160	148,330	75,320

individual receives more information about each of those occupations before making a one-time decision regarding his occupation.

5.1 Choice of occupation

We begin by examining the last decision, namely the choice of occupation conditional on major and paying the information cost for a subset of the occupations.[14] Let v_{ijk} denote the expected present value of lifetime utility for individual i from choosing occupation k conditional on major j, *before* the realization of the information shock. Individuals form their subjective expectations regarding the probabilities of entering

[14]In practice, the information cost can be thought of as a cost of application (per occupation).

Table 8: Heterogeneous Ex Ante Treatment Effects by Counterfactual Major (Annual Earnings, in dollars)

		Economics	Engineering	Humanities	Natural Sciences	Public Policy	Social Sciences
Science	TT	6,660	42,960	19,250	35,860	18,560	12,070
	TUT	10,344	47,353	10,378	32,540	17,999	13,708
	ATE	10,140	45,580	11,090	33,700	18,030	13,560
Health	TT	63,330	108,620	88,130	87,000	73,350	74,080
	TUT	49,949	77,177	51,290	78,234	57,628	51,953
	ATE	51,000	83,570	55,650	80,930	59,620	55,170
Business	TT	130,560	88,140	67,360	62,780	100,270	94,280
	TUT	98,567	80,886	58,376	56,942	84,208	63,424
	ATE	114,390	82,220	60,540	57,710	87,800	71,200
Government	TT	20,160	28,550	24,370	24,900	28,760	34,370
	TUT	24,453	25,163	19,383	18,862	35,450	20,232
	ATE	23,720	25,440	20,140	19,340	33,310	23,120
Law	TT	88,460	111,330	78,380	80,020	78,150	78,460
	TUT	78,666	98,807	78,593	68,761	87,613	82,883
	ATE	79,960	99,600	78,550	69,580	85,790	82,080

different careers based on these *ex ante* value functions. The new information consists of a vector of shocks ϵ_{ijk} that vary at the individual-major-occupation level. For any given major j, we assume that the ϵ_{ijk}'s are independent draws from a Type 1 extreme value distribution. After making an initial major choice and graduating from college, these shocks are realized and the individual then proceeds to choose an occupation. An individual who chose major j then chooses his occupation k^* according to:

$$k^* = \arg\max_{k \in K^*_{ij}} (v_{ijk} + \epsilon_{ijk}) \tag{5.1}$$

where K^*_{ij} is the set of occupations where the individual has paid for the new information conditional on an initial major j. We will discuss the decision to acquire more information about particular occupations in Subsection 5.3.

5.2 Linking subjective probabilities to occupation-major preferences

An individual's self-reports of the probabilities of choosing particular occupations can then be used to recover their expected utilities (up to a reference alternative). To see this, first consider the case where it is optimal for the individual to pay the informational cost for all occupations conditional on major j. With the Type 1 extreme value assumption on the ϵ_{ijk}'s, we can recover the difference in conditional value functions by inverting the choice probabilities following Hotz & Miller (1993) and Berry (1994):

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

We assume that the conditional value functions, for any given major j and occupation k, can be written as follows:

$$v_{ijk} = \alpha_{ik} + \delta_{jk} + \gamma_w \ln w_{ijk} + \eta_{ijk}$$

where α_{ik} is the preference i has for occupation k, δ_{jk} captures the average complementarity of preferences between major j and occupation k, w_{ijk} is the expected earnings measure for i under choices $\{j, k\}$, and η_{ijk} is an orthogonal preference term for occupation k given major j.[15] Similarly to Arcidiacono(2004,2005) in the context of college major choice, the value function is assumed to depend on future labor market outcomes through the logarithm of the expected earnings.[16]

Taking the difference with respect to the baseline occupation, it follows that the following equality holds:

$$\ln(p_{ijk}) - \ln(p_{ij1}) = (\alpha_{ik} - \alpha_{i1}) + (\delta_{jk} - \delta_{j1}) + \gamma_w(\ln w_{ijk} - \ln w_{ij1}) + \zeta_{ijk} \tag{5.3}$$

[15] In practice, monetary or psychic costs of schooling associated with the occupations which typically require an advanced degree, such as Law, would also be captured by these preference terms. It follows that our empirical strategy will presumably lead to underestimate the true preferences for those specific occupations.

[16] While forward-looking individuals should consider the present value of lifetime earnings associated with each occupation, in practice we only observe the expected earnings ten years after graduation. However, we show in Appendix A that, under some plausible assumptions on the discount factor, worklife duration and earnings growth, the earnings ten years out are, up to a constant, a reasonably good approximation of the present value of lifetime earnings.

where $\zeta_{ijk} \equiv \eta_{ijk} - \eta_{ij1}$.

5.3 Information costs

We now consider the information acquisition stage. Note that this stage arises because the subjective probability of some choices (conditional on a particular major) are zero. With the information having continuous support, a subjective probability of zero would not be possible if the information was costless. However, if the individual can choose whether or not to acquire the information, zero probabilities can result.

The decision to acquire information hinges on expectations of the maximal utility associated with different choice sets. Given the Type-1 Extreme Value assumptions regarding the distribution of the ϵ's, McFadden (1978) showed that the expected maximum utility for any choice set K, $V_{ij}^{(K)}$, can be written as:

$$V_{ij}^{(K)} = \ln\left[\sum_{k\in K}\exp(v_{ijk})\right] + \gamma$$

where γ is Euler's constant.

Without loss of generality, denote v_{ij1} as the payoff associated with the career that gives the highest utility prior to the new information, denote v_{ij2} as the utility associated with the next highest, etc. We denote the utility cost of obtaining information on a particular occupation-major pair as c. Individuals only obtain information if the expected gain is high enough to overcome the cost. Conditional on paying the information cost for the first $(k-1)$ occupations, information on career k (the kth highest payoff) is obtained when:[17]:

$$\begin{aligned} c &\leq \ln\left(\sum_{k'=1}^{k}(\exp(v_{ijk'}))\right) - \ln\left(\sum_{k'=1}^{k-1}(\exp(v_{ijk'}))\right) \\ &\leq \ln\left(\frac{\sum_{k'=1}^{k}(\exp(v_{ijk'}))}{\sum_{k'=1}^{k-1}(\exp(v_{ijk'}))}\right) = -\ln(1-p_{ijk}) \end{aligned} \tag{5.4}$$

We can then get an upper bound estimate of c from the lowest positive self-reported probability of choosing an occupation conditional on a major.

[17]Note that, at the individual level, it is always optimal to consider the occupations in this order.

5.4 Selection

Reports of zero probabilities can not be ignored in estimation because of the selection problem: those who report zero probabilities have particularly low values for those occupation-major pairs. Now suppose that k is not in the set K_{ij}^*. In this case the inequality in (5.4) is flipped:

$$c > -\ln(1 - p_{ijk}) \tag{5.5}$$

Note that the p_{ijk} term in (5.5) is *conditional* on k being in the choice set. Since k was not in the choice set (the information cost was not paid), we have no measure of p_{ijk}. However, we can substitute in for (5.5) with the relevant v_{ijk}'s where the choice set is now $K_{ij}^* \cup \{k\}$:

$$\begin{aligned} c &> -\ln\left(1 - \frac{\exp(v_{ijk})}{\exp(v_{ijk}) + \sum_{k' \in K_{ij}^*} \exp(v_{ijk'})}\right) && (5.6) \\ &> -\ln\left(1 - \frac{\exp(v_{ijk} - v_{ij1})}{\exp(v_{ijk} - v_{ij1}) + \sum_{k' \in K_{ij}^*} \exp(v_{ijk'} - v_{ij1})}\right) && (5.7) \end{aligned}$$

We then need to solve this equation for $v_{ijk} - v_{ij1}$ as the other differenced conditional value functions are known from (5.2). Solving,

$$\begin{aligned} \exp(-c) &< \left(\frac{\sum_{k' \in K_{ij}^*} \exp(v_{ijk'} - v_{ij1})}{\exp(v_{ijk} - v_{ij1}) + \sum_{k' \in K_{ij}^*} \exp(v_{ijk'} - v_{ij1})}\right) \\ \exp(v_{ijk} - v_{ij1}) &< \frac{(1 - \exp(-c)) \sum_{k' \in K_{ij}^*} \exp(v_{ijk'} - v_{ij1})}{\exp(-c)} \\ v_{ijk} - v_{ij1} &< \ln\left(\frac{(1 - \exp(-c)) \sum_{k' \in K_{ij}^*} \exp(v_{ijk'} - v_{ij1})}{\exp(-c)}\right) \equiv c_{ijk}^* \end{aligned}$$

Up to now we have not needed to make a distributional assumption on the ζ_{ijk}'s. With zero probabilities, this is no longer the case. We assume that ζ_{ijk} is distributed i.i.d. $N(0, \sigma)$, implying that the log likelihood contribution in the zero probability case is:

$$\ln(p_{ijk} = 0) = \ln \Phi\left(\frac{c_{ijk}^* + (\alpha_{i1} - \alpha_{ik}) + (\delta_{j1} - \delta_{jk}) + \gamma_w(Y_{ij1} - Y_{ijk})}{\sigma}\right) \tag{5.8}$$

where Φ is the standard normal cdf.

5.5 Heterogeneous information sets

It may be that students have better information about the labor market for some majors than others. In particular, it may be the case that individuals have better information about the labor market in their own major than in counterfactual majors. The model we have developed can be relaxed to allow for counterfactual majors to have higher variances associated with the information shocks.

Absent additional assumptions, discrete choice models are only identified relative to the variance scale parameter. Implicit in (5.3) is a normalization of the variance scale parameter to one. With the structure we have placed on (5.3), we can allow for the variance parameter to be different for counterfactual majors. We then specify (5.3) as (without loss of generality):

$$\ln(p_{ijk}) - \ln(p_{ij1}) = \frac{(\alpha_{ik} - \alpha_{i1}) + (\delta_{jk} - \delta_{j1}) + \gamma_w(Y_{ijk} - Y_{ij1}) + \zeta_{ijk}}{1 + \phi I(d_i = j)} \tag{5.9}$$

If ϕ is greater than zero, then students are less certain about outcomes in counterfactual majors than they are in their own majors.

5.6 Compensating differentials

Our specification of the payoffs for major-occupation bundles allows us to recover individual-level preferences for occupation k relative to occupation 1, $\alpha_{ik} - \alpha_{i1}$, as well as estimates of the average preferences for occupation k relative to occupation 1 conditional on major j, $\delta_{jk} - \delta_{j1}$. We can translate this into monetary units using the expected earnings coefficient γ, thus translating those parameters into (expected) compensating differentials for the different occupations (given each college major).

Of key interest here is the average compensating differential for occupation k relative to occupation 1, conditional on major j, which is given by:

$$CD(k|j) = \frac{\delta_{j1} - \delta_{jk}}{\gamma_w} \tag{5.10}$$

Furthermore, using the estimates of the parameters $\alpha_{ik} - \alpha_{i1}$, we can also see how compensating differentials for each occupation vary across individuals. In particular, similarly to the *ex ante* treatment effects parameters that we have estimated (namely

ATE, TT and TUT), we can compute, for each occupation k, the average compensating differential, the average compensating differential conditional on choosing occupation k as well as the average compensating differential conditional on not choosing occupation k. For example, the additional compensating differential for occupation k relative to occupation 1 for those who chose major j is:

$$CD(k|d_j = 1) = \frac{\sum_i I(d_{ij'} = 1)\alpha_{ik}}{\gamma_w \sum_i I(d_{ij'} = 1)} - \frac{\sum_i I(d_{ij} = 1)\alpha_{ik}}{\gamma_w \sum_i I(d_{ij} = 1)} \tag{5.11}$$

6 Results: Compensating Differentials

Estimates of the earning parameter, γ, for different specification of the conditional valuation functions are given in Table 9. For our earnings measure, we use the log of expected earnings ten years after graduation. Hence, when discussing compensating differentials, they will be percentage increases in earnings ten years out. For each of the specifications, log expected earnings are statistically significant.

The final column allows the variance on the new information to be different for counterfactual majors. The coefficient estimate for ϕ was small and insignificant and we can not reject that it is zero. Note that this specification is adding the flexibility in the variance after controlling for individual occupation dummies. In estimates not reported here, the variance for counterfactual majors was higher and statistically significant if we allowed for different variances in models 1 and 2. Given these results, we focus on model 3 as our preferred specification.

To assess the extent to which expected earnings affects occupational choice, we can calculate the percentage change in the probability of choosing an occupation given a percentage change in earnings. At the intensive margin, the elasticity formula for our specification is (see Train, 2003):

$$\eta_{ijk} = (1 - Pr_{ijk})\gamma_w$$

For those on the intensive margin, the subjective probabilities of entering a given career conditional on a given major range from 0.003 to 0.962, yielding elasticities from zero to 0.64 for our preferred specification (Model 3). Taking the major from the data as given, we can estimate the population elasticity of occupation k using:

$$\hat{\eta_k} = \frac{\sum_i \sum_j I(j|i)(1 - Pr_{ijk})\gamma_w}{N}$$

These occupation-specific elasticities range from 0.49 (for Business) to 0.60 (for Education), resulting in a mean elasticity across all occupations equal to 0.55.

Table 9: Structural Model Estimates

	Model 1	Model 2	Model 3	Model 4
Log expected earnings 10 years out	1.252	0.617	0.664	0.668
	(0.027)	(0.020)	(0.017)	(0.017)
Occupation dummies	yes	no	no	no
Occupation-major dummies	no	yes	yes	yes
Individual occupation dummies	no	no	yes	yes
Better information in own major	no	no	no	yes
Log likelihood (000's)	-18.47	-14.03	-6.466	-6.466

6.1 Compensating differentials

We next report how compensating differentials for particular occupations vary among those who chose particular majors using equation (5.11). All of the heterogeneity in compensating differential is relative to the education occupation. Note that the average compensating differential in the population is not present here because it is captured by the δ_{jk}'s.

Table 10 gives the results with the units reported as percentage changes in expected earnings ten years out to make the average individual of a particular major indifferent between the two occupations, all else equal. Economics majors and public policy majors have strong preferences to avoid the education occupation relative to the average Duke student and strongly prefer business and government occupations relative to other majors. On the other hand, natural science majors, social science majors, and humanities majors prefer education over business.

Table 10: Heterogeneity in Compensating Differentials by Chosen Major Relative to Education

	Science	Health	Business	Government	Law
Natural Science	0.1%	4.6%	105.1%	48.7%	138.5%
Engineering	15.1%	64.1%	-21.6%	-10.9%	-15.3%
Economics	-83.2%	-117.7%	-130.5%	-62.2%	-56.6%
Public Policy	-31.1%	5.0%	-111.7%	-137.8%	-123.6%
Social Science	26.7%	13.6%	61.2%	56.6%	-26.9%
Humanities	100.0%	57.1%	110.3%	73.7%	57.0%

The estimates of the individual preferences for occupations also allow us to examine their correlation patterns. Table 11 gives the variance of the occupation-specific preferences while the off-diagonal elements give the correlation coefficients. Preferences for business and law tend to be negatively associated with preferences for education, resulting in particularly high correlation coefficients between business and law with each other as well as with health and, to a lesser extent, government.

Table 11: Variances and correlation coefficients for occupation-specific preferences

	Science	Health	Business	Government	Law
Science	1.837	0.215	0.149	-0.134	0.178
Health	0.215	2.55	0.607	-0.037	0.569
Business	0.149	0.607	2.235	0.373	0.681
Government	-0.134	-0.037	0.373	2.543	0.329
Law	0.178	0.569	0.681	0.329	3.258

6.2 Major-specific compensating differentials

We next examine how compensating differentials are affected by major, translating our estimates of the δ_{jk}'s into percentage increases in earnings. Table 12 reports average compensating differentials for particular occupation-major combinations, again relative to the education occupation. Although the signs are all intuitive, the magnitudes are such that there is likely more to the story than just compensating differentials. For example, an economics major makes working in business so attractive that on average individuals would need to make over three times as much in education (or making less than a third of what they would make in business) to be indifferent between the two occupations. Similarly, a science major makes working in a science occupation so attractive that on average individuals would need to make over two and a half times more in education to be indifferent between the two occupations.[18]

The final column of Model 3 reports what the compensating differentials would need to be if we did not account for differences in earnings. In this case, a coefficient on earnings is therefore not estimated and we use the coefficient from Model 3 with earnings to perform the calculations. Comparing the last two columns of Table 12 then allows use to see the role earnings play in mitigating compensating differentials. As expected, the compensating differentials in the last column are all higher (in absolute value) than those when earnings are accounted for, as expected earnings in education are substantially lower than in other occupations. Not accounting for those earnings differences would make it appear as though education was even more unattractive than it actually was.

6.3 Search frictions

What can explain the very large estimates of the compensating differentials? One explanation is that the average differences in compensating differential across majors is partly driven by search frictions. That is, being an economics major does not make

[18]It is interesting to note that these findings are in line with the literature on major choice, which tends to find that preferences play a key role in this decision (see, e.g., Arcidiacono, 2004, Beffy et al., 2012, and Wiswall & Zafar, 2012) .

business occupations more attractive beyond the salary gains but the arrival rate of offers in the business occupations is higher if the individual is an economics major.

To illustrate how search frictions will affect our estimates of compensating differentials, consider a simple case where there are two occupations, $k \in \{1, 2\}$. Suppose for major j individuals are given one offer in occupation 1 but two offers in occupation 2. The difference between the two offers in occupation 2 comes solely through the non-pecuniary shocks, not through income. If the non-pecuniary shocks are treated as just another extreme value shock, then the probability of choosing occupation 2 will be:

$$Pr(k = 2|j) = \frac{2 * \exp(v_{j2})}{\exp(v_{j1}) + 2 * \exp(v_{j2})} = \frac{\exp(v_{j2} + \ln(2))}{\exp(v_{j1}) + \exp(v_{j2} + \ln(2))} \tag{6.1}$$

Hence, if offer rates for various occupations differ by major, then this will manifest itself as a compensating differential.[19]

We cannot separate compensating differentials from offer rates, but we can say how big differences in offer rates would have to be to explain the average compensating differentials we find for particular major-occupation combinations. Denote λ_{jk} as the arrival rate of offers for occupation k conditional on major j. We assume that the offers unobserved component is Type 1 extreme value: there is no correlation of offers within occupation categories. Allowing for correlation in this component within an occupation category would result in increases in magnitudes of the differences in arrival rates in order to account for the estimated differences in compensating differentials. Hence, one can think of our approach as identifying the *minimum* amount of differences in occupation-major arrival rates that account for the observed compensating differentials. Our estimates of $(\delta_{jk} - \delta_{j1})$ can be transformed into differences in arrival rates using:

$$\delta_{jk} - \delta_{j1} = \ln(\lambda_{jk}) - \ln(\lambda_{j1}) \tag{6.2}$$

Since we can only identify five of the six arrival rates for each major, we normalize λ_{j1} to one. Solving for λ_{jk} then gives the number of offers in occupation k per offer in education.

[19]Note that variance in earnings from which offers were drawn would also generate a similar result, but would require heterogeneity in the variance due to the major. Variance in offered wages would have to unreasonably different across majors to explain our results.

Results are presented in Table 13. In order for job offer rates to account for the estimated compensating differentials, natural science majors would have to receive at least 6.4 offers in science occupations and at least 1.5 offers in business for every one offer in education. In contrast, humanities majors would expect significantly fewer offers in the sciences, 0.5 offer for every offer in education, with roughly equal offers in business as in education. Majoring in economics would need to result in at least 10 offers in business for every offer in education to account for the compensating differential associated with the economics-business combination. These results, combined with those in Table 12, show that some combination of large differences in arrival rates occur due to one's major or one's major makes jobs in particular occupations much more enjoyable. In either event, majors have a substantial effect on the labor market outcomes beyond their impact on earnings.

7 Conclusion

This paper shows how subjective expectation data on counterfactual outcomes can be used to recover the *ex ante* treatment effects as well as the non-pecuniary benefits associated with different treatments. We consider the particular context of sorting across occupations, using elicited beliefs from a sample of male undergraduates at Duke University on the probability of working in different occupations as well as the expected income in each of those occupations (10 years after graduation). Importantly, these beliefs were asked not only for the college major the individual chose, but also for counterfactual majors, thus making it possible to examine the heterogeneity across majors of the *ex ante* returns to different occupations and the subjective probabilities of working in any given occupation. This individual variation across counterfactual majors is key to tell apart the role of *ex ante* returns and preferences in the context of sorting across occupations. While sorting across occupations is found to be partly driven by the *ex ante* monetary returns, large differences in expected income across occupations remain after controlling for selection on monetary returns, which in turn points to the existence of substantial compensating differentials for particular occupations.

References

Altonji, J., Blom, E. & Meghir, C. (2012), 'Heterogeneity in human capital investments: High school curriculum, college majors, and careers', *Annual Review of Economics* **4**, 185–223.

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

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.

Berry, S. (1994), 'Estimating discrete-choice models of product differentiation', *The RAND Journal of Economics* **25**, 242–262.

Betts, J. (1996), 'What do students know about wages? evidence from a survey of undergraduates', *Journal of Human Resources* **31**, 27–56.

Cunha, F. & Heckman, 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. & 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.

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. (1996), 'Eliciting student expectations of the returns to schooling', *Journal of Human Resources* **19**, 408–428.

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

Eisenhauer, P., Heckman, J. & Vytlacil, E. (2012), The generalized Roy model and the cost-benefit analysis of social programs. Working paper.

Heckman, J. J. & Navarro, S. (2007), 'Dynamic discrete choice and dynamic treatment effects', *Journal of Econometrics* **136**, 341–396.

Hotz, V. & Miller, R. (1993), 'Conditional choice probabilities and the estimation of dynamic models', *Review of Economic Studies* **60**, 497–529.

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

Kinsler, J. & Pavan, R. (2012), The specifity of general human capital: evidence from college major choice. Working paper.

Manski, C. (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.

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

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

Osman, A. (2013), Occupational choice under credit and information constraints. Working paper.

Pantano, J. & Zheng, Y. (2010), Using subjective expectations data to allow for unobserved heterogeneity in hotz-miller estimation strategies. Working paper.

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

Stinebrickner, T. & Stinebrickner, R. (2013), 'A major in science? initial beliefs and final outcomes for college major and dropout', *Forthcoming in the Review of Economic Studies* .

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. (2008), 'Social security and the retirement and savings behavior of low-income households', *Journal of Econometrics* **145**, 21–42.

Wiswall, M. & Zafar, B. (2012), Determinants of college major choice: identification using an information experiment. Working paper.

Zafar, B. (2011), 'How do college students form expectations?', *Journal of Labor Economics* **29**, 301–348.

Zafar, B. (2013), 'College major choice and the gender gap', *Journal of Human Resources* **48**, 545–595.

A Appendix A

We provide below some sufficient conditions under which the earnings ten years after graduation can be used to approximate the present value of lifetime earnings, for any major and occupation.

Specifically, we let w_{10} (respectively w_t) denote the earnings ten years out (resp. t years out), β the annual discount factor and T the worklife duration. We define the approximation error as $\Delta \equiv \left| w_{10} - \frac{\sum_{t=1}^{T} \beta^t w_t}{\sum_{t=1}^{T} \beta^t} \right|$. Individual, major and occupation subscripts are omitted to save on notations. Assuming that earnings grow at a constant rate ρ ($w_{t+1} = w_t \exp(\rho)$), it follows that the approximation error can be written as:

$$\begin{aligned} \Delta &= \left(\frac{w_{10}}{\sum_{t=1}^{T} \beta^t} \right) \left(\sum_{t=1}^{T} \beta^t \left(\exp((t-10)\rho) - 1 \right) \right) \\ &= \left| w_{10} \left(\exp(-10\rho) \frac{\sum_{t=1}^{T} \beta^t \exp(\rho t)}{\sum_{t=1}^{T} \beta^t} - 1 \right) \right| \end{aligned}$$

Setting $\beta = 0.9$, $\rho = 3\%$ and $T = 40$ years yields $\frac{\Delta}{w_{10}} \simeq 0.015$. It follows that, under those assumptions, the earnings ten years after graduation are reasonably close to the present value of lifetime earnings, weighted by the sum $\sum_{t=1}^{T} \beta^t$. The latter term does not vary across occupations and therefore drops out when taking the difference with respect to the baseline occupation (see Equation 5.3 in the main text).

Table 12: Average Compensating Differentials by Major-Occupation Pairs Relative to Education

Major	Occupation	Model 2	Model 3	Model 3 w/o Earnings
	Science	-258%	-279%	-325%
Natural	Health	-197%	-215%	-296%
Science	Business	-35%	-66%	-129%
	Government	11%	-33%	-61%
	Law	78%	56%	-2%
	Science	-298%	-306%	-358%
	Health	-130%	-136%	-209%
Engineering	Business	-152%	-151%	-221%
	Government	-15%	-52%	-77%
	Law	67%	44%	-11%
	Science	-95%	-102%	-132%
	Health	-70%	-48%	-107%
Economics	Business	-372%	-348%	-444%
	Government	-162%	-205%	-234%
	Law	-44%	-76%	-139%
	Science	-75%	-65%	-98%
Public	Health	-140%	-90%	-148%
Policy	Business	-247%	-197%	-271%
	Government	-284%	-280%	-321%
	Law	-111%	-129%	-198%
	Science	-51%	-30%	-59%
Social	Health	-43%	-6%	-62%
Science	Business	-166%	-118%	-184%
	Government	-95%	-105%	-136%
	Law	-5%	-17%	-81%
	Science	131%	117%	93%
	Health	121%	138%	83%
Humanities	Business	-29%	-11%	-66%
	Government	88%	48%	24%
	Law	182%	143%	87%

Table 13: Number of Offers per Offer in Education Necessary to Account for Average Major-Occupation Compensating Differentials

	Occupation				
Major:	Science	Health	Business	Government	Law
Natural Science	6.38	4.17	1.55	1.24	0.69
Engineering	7.63	2.47	2.73	1.41	0.75
Economics	1.97	1.38	10.08	3.90	1.66
Public Policy	1.54	1.82	3.70	6.42	2.36
Social Science	1.22	1.04	2.19	2.01	1.12
Humanities	0.46	0.40	1.08	0.73	0.39

Note: Calculations from estimates of Model 3