

Wiswall, Matthew; Zafar, Basit

**Working Paper**

## Determinants of college major choice: Identification using an information experiment

Staff Report, No. 500

**Provided in Cooperation with:**  
Federal Reserve Bank of New York

*Suggested Citation:* Wiswall, Matthew; Zafar, Basit (2011) : Determinants of college major choice: Identification using an information experiment, Staff Report, No. 500, Federal Reserve Bank of New York, New York, NY

This Version is available at:  
<https://hdl.handle.net/10419/60833>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

Federal Reserve Bank of New York  
Staff Reports

Determinants of College Major Choice:  
Identification Using an Information Experiment

Matthew Wiswall  
Basit Zafar

Staff Report no. 500  
June 2011

This paper presents preliminary findings and is being distributed to economists and other interested readers solely to stimulate discussion and elicit comments. The views expressed in this paper are those of the authors and are not necessarily reflective of views at the Federal Reserve Bank of New York or the Federal Reserve System. Any errors or omissions are the responsibility of the authors.

## **Determinants of College Major Choice: Identification Using an Information Experiment**

Matthew Wiswall and Basit Zafar

*Federal Reserve Bank of New York Staff Reports*, no. 500

June 2011

JEL classification: D81, D84, I21, I23, J10

### **Abstract**

This paper studies the determinants of college major choice using a unique “information” experiment embedded in a survey. We first ask respondents their *self*-beliefs—beliefs about their own expected earnings and other major-specific outcomes conditional on various majors, their *population* beliefs—beliefs about the population distribution of these characteristics, as well as their subjective beliefs that they will graduate with each major. After eliciting these baseline beliefs, we provide students with information on the true population distribution of these characteristics, and observe how this new information causes respondents to update their beliefs. Our experimental design creates unique panel data. We first show that respondents make substantial errors in population beliefs, and logically revise their self-beliefs in response to the information. Subjective beliefs about future major choice are positively and strongly associated with beliefs about self-earnings, ability, and spouse’s earnings. However, cross-sectional estimates are severely biased upward because of the positive correlation of tastes with earnings and ability. The experimental variation in beliefs allows us to identify a rich model of college major choice, with which we estimate the relative importance of earnings and earnings uncertainty on the choice of college major versus other factors, such as ability to complete coursework, spouse’s characteristics, and tastes for majors. While earnings are a significant determinant of major choice, tastes are the dominant factor in the choice of field of study. We also investigate why males and females choose different college majors.

Keywords: college majors, information, uncertainty, risk, subjective expectations, marriage market returns, gender differences

---

Wiswall: New York University, Department of Economics (e-mail: [mwiswall@nyu.edu](mailto:mwiswall@nyu.edu)).  
Zafar: Federal Reserve Bank of New York (e-mail: [basit.zafar@ny.frb.org](mailto:basit.zafar@ny.frb.org)). The authors thank the NYU Center for Experimental Social Sciences for providing assistance conducting the information survey and experiment and participants at the Federal Reserve Bank of New York seminar series and the NYU Experimental Economics Working Group. Da Lin and Scott Nelson provided outstanding research assistance. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

# 1 Introduction

Understanding the determinants of occupational choices is a classic question in the social sciences: How much do occupational choices depend on expected future earnings versus tastes for various non-pecuniary aspects of an occupation? Among college graduates, occupational choices are strongly associated with college major choices as the choice of major—whether in humanities, business, science or engineering fields—represents a substantial investment in occupation-specific human capital. Underscoring the importance of college major choices, a number of studies have documented that choice of post-secondary field is a key determinant of future earnings, and that college major composition can help explain long-term changes in inequality and earnings differences across racial groups and between men and women (Grogger and Eide, 1994; Garman and Loury, 1995; Brown and Corcoran, 1997; Weinberger, 1998; Arcidiacono, 2004; Wiswall, 2006).

This paper studies the determinants of college major choices using a unique survey and experimental design. We conduct an experiment on undergraduate college students of New York University (NYU), where in successive rounds we ask respondents their *self* beliefs about their own expected earnings and other major-specific aspects were they to major in different majors, their beliefs about the population distribution of these outcomes, and the subjective belief that they will graduate with each major. After the initial round in which the baseline beliefs are elicited, we provide students with accurate information on the population characteristics and observe how this new information causes respondents to update their self beliefs and their subjective probabilities of graduating with each particular major. Our experimental design creates unique panel data for major choice, which is otherwise a one-time decision. By comparing the experimental *changes* in subjective probabilities of majoring in each field with the *changes* in subjective expectations about earnings and other characteristics of the major, we can measure the relative importance of each of these various characteristics in the choice of major, without bias stemming from the correlation of fixed preferences with characteristics. Underscoring the importance of this bias, we compare cross-sectional OLS estimates of major choice to expectations about earnings with our panel fixed effects estimates, and find that the OLS estimates are severely biased upward due to positive correlation of tastes with earnings expectations and perceived ability.

Our approach is motivated by previous research which has found that many college students have biased beliefs about the population distribution of earnings among current graduates (Betts, 1996), and that students tend to be misinformed about returns to schooling (Jensen, 2010; Nguyen, 2010). We test whether students update their beliefs if given accurate information on the current population earnings, and find heterogenous errors in population beliefs, and substantial and logical updating in response to our information treatment. We show how the ex-

perimental variation alone identifies a rich model of college major choice, and we use this model to understand the importance of earnings and earnings uncertainty on the choice of college major relative to other factors such as ability to complete coursework, spousal characteristics, and tastes for majors.

The standard economic literature on decisions made under uncertainty, such as occupational choice, generally assumes that individuals, after comparing the expected outcomes from various choices, choose the option that maximizes their expected utility. Given the choice data, the goal is to infer the parameters of the utility function. Because one does not typically observe expectations about future choice-specific outcomes, such as the student's expectations of earnings and ability in a major, assumptions have to be made on expectations to infer the decision rule. This approach requires a mapping between objective measures (such as realized earnings) and beliefs about them. Moreover, assumptions also have to be invoked about expectations for counterfactual majors, i.e., majors not chosen by the student. Several studies of college major choice use this approach (Freeman, 1971; Bamberger, 1986; Berger, 1988; Montmarquette, Cannings, and Mahseredjian, 2002; Arcidiacono, 2004; Beffy, Denis, and Maurel, 2011). These studies assume expectations are either myopic or rational, and that the expectations formation process is homogenous conditional on observables. This approach is problematic because observed choices might be consistent with several combinations of expectations and preferences, and the list of underlying assumptions may not be valid (Manski, 1993).

A recent literature has evolved which collects and uses subjective expectations data to understand decision-making under uncertainty (see Manski, 2004, for a survey of this literature). In the context of schooling choices, Zafar (2009, 2011), Giustinelli (2010), Arcidiacono, Hotz, and Kang (2011), Kaufmann (2011), and Stinebrickner and Stinebrickner (2010, 2011) incorporate subjective expectations into models of choice behavior. These studies collect data on expectations for the chosen alternative as well as counterfactual alternatives, thereby eliminating the need to make assumptions regarding expectations. However, as we show in Section 3, one cannot separately identify the tastes for each major from other aspects of the choice (earnings, ability, etc.) without imposing further modeling restrictions. Experimental variation in beliefs allows us to accomplish that. More precisely, at the baseline, we collect self beliefs and beliefs about the population distribution of some college major characteristics, as well as probabilistic choices of major. We then provide students with accurate fact-based information on population characteristics. If students are mis-informed about population characteristics and perceive some link between population and self beliefs, this information should cause them to revise their beliefs and choices. There are in fact substantial errors in population beliefs, with students, on average, under-estimating the population earnings in most majors. For example, male and female respondents underestimate the male population full-time earnings in

Engineering/Computer Science by around 14%. We next find that students logically revise their self beliefs about own and spouse earnings and own ability in response to the information we provide. The response, however, is inelastic: For a 1 percent error, students revise their self earnings by 0.196 percent, suggesting that self beliefs are not entirely linked to the type of public information that we provide.

Another novel feature of our framework is that we collect data to investigate whether marriage market returns are a determinant of field of study. This question is motivated by recent theoretical models that have emphasized that investment in education generates returns in the marriage market (Iyigun and Walsh, 2007; Chiappori, Iyigun, and Weiss, 2009). These models are based on the idea that changes in marriage market conditions (such as sex ratios and degree of assortative mating in age and education) have an effect on the outside option of each spouse, which in turn alters bargaining weights and leads to changes in the way the household surplus is shared. If individuals are forward-looking and anticipate these conditions, this should be reflected in their expectations. Since such data are typically not available, empirical evidence of the effect of marriage market considerations on educational choices is scant, and is inferred indirectly (Ge, 2010; Lafortune, 2010; Attanasio and Kaufmann, 2011). In this paper, we are able to provide direct evidence on whether marriage market returns are a determinant of field of study. We collect data on students' beliefs about the probability of marriage, spouse's earnings, and spouse's labor supply conditional on own field of study.

Our reduced-form estimates using baseline (cross-sectional) data show that beliefs about future relative major choices are positively and strongly associated with beliefs about future self earnings, ability, and spouse's earnings. For example, a 1 percent increase in beliefs about self earnings in a major (relative to humanities/arts) increases the log odds of majoring in that field (relative to humanities/arts) by about 2 percent. Spousal earnings have a considerably lower effect on major choice, with the effect being smaller for female respondents. On the other hand, using the revisions in beliefs and choices, we show that in fact the estimates using cross-sectional data are biased upwards because of the positive correlation between the unobserved individual-specific taste component and beliefs about ability and earnings. For example, the choice elasticity with respect to beliefs about earnings is an order of magnitude lower (about 0.28 percent) using revisions in beliefs and choices, as part of an individual fixed effect analysis.

We next estimate a structural life-cycle model utility model of college major choice. Unlike the existing literature on educational choices that only elicits beliefs of expected future earnings, we collect data on the underlying earnings distribution, and also investigate the role that risk plays in college major choice (Attanasio and Kaufmann, 2011, is an exception). Our parameter estimates imply a relative risk aversion coefficient of around 5, similar to that found by Nielsen and Vissing-Jorgensen (2006) in a Danish dataset on labor incomes and educational

choices. Moreover, our estimate of relative risk aversion is higher for females, which is consistent with experimental studies of gender differences in risk preferences (Eckel and Grossman, 2008; Croson and Gneezy, 2009). Imposing risk neutrality in our model—a common assumption in existing studies of college major choice—shows that we would substantially over-estimate (under-estimate) the probability of majoring in high (low) earnings fields.

Our model estimates indicate that earnings are a significant determinant of major choice. However, the taste component at the time of choosing a college major is the dominant factor in the choice of field of study, a finding similar to that of Arcidiacono (2004) and Beffy et al. (2011). With respect to the marriage market returns to major choice, we find that they have a small positive impact on choosing high-earnings majors, but a substantial negative impact on choosing the "not graduate" category.

This paper also contributes to the literature on gender differences in schooling choices. Males and females are known to choose very different college majors (Turner and Bowen, 1999; Dey and Hill, 2007; Gemici and Wiswall, 2011). Niederle and Vesterlund (2007) speculate that women being less over-confident than men may be one possible explanation for this. Zafar (2010), in his sample of Northwestern University undergraduates, finds that gender differences in tastes (and not ability) are the main source of these differences. In our sample, we find that women, on average, do have lower beliefs of ability in all fields relative to men. In order to shed light on gender differences in major choice, we obtain gender-specific estimates of the structural model. The model estimates show that earnings differences across majors are a substantially smaller factor in college major choice for women than men, and that ability differences matter substantially more for women. The taste component is, however, dominant for both males and females.

While our experimental variation generates a panel that may look similar to other datasets with longitudinal information on beliefs (see Stinebrickner and Stinebrickner, 2010, 2011; Zafar, 2011, in the context of college major choice), there is an important distinction: Beliefs in our survey are separated by only a few minutes, while in conventional panels, the gap is typically of several months. We can then credibly claim that the utility function, most notably the individual and major specific taste parameters, are truly time invariant in our context—the key assumption to identifying the tastes non-parametrically—and that our experimentally derived panel data satisfies the standard fixed effects assumptions. Estimating the taste parameters non-parametrically, we find that i) the distribution of tastes is bimodal, ii) average tastes of females are negative for all majors (relative to humanities/arts), and iii) male students have a strong relative taste for economics/business majors. Moreover, the fit of the estimated structural model using the experimental variation in beliefs is substantially better than when we estimate the model using cross-sectional data and impose a parametric assumption on the

taste parameter, as in Arcidiacono et al. (2011).

This paper is organized as follows. Section 2 outlines the model of college major choice. In Section 3, we explore identification of the model using: i) commonly used revealed choice data, ii) cross-sectional beliefs, and iii) panel data on beliefs. The data collection methodology is outlined in Section 4. We examine heterogeneity in beliefs about earnings and revisions in self beliefs following the information treatment in Section 5. Section 6 reports reduced-form regressions on the relationship between beliefs about major choice and beliefs about elements of future post-graduation utility, while Section 7 reports estimates from a structural life-cycle utility model of major choice. Finally, Section 8 concludes.

## 2 Model

In this section we specify the model of college major choice. The next section shows how we use the information experiment to identify the model.

Individuals choose one of  $K$  majors:  $k = 1, \dots, K$ .<sup>1</sup> At the initial period  $t = -1$ , individuals are enrolled in college and have not chosen a particular college major. At the beginning of period  $t = 0$ , the individual makes a college major choice and graduates from college. From period  $t = 1$  onward, the college graduate makes all remaining choices, including choices regarding labor supply and marriage.<sup>2</sup>

We do not explicitly model any of the choices during or after college (i.e., choice to take particular courses during college, or any of the post-graduation choices). Instead we specify a preference ordering over the particular college majors. At the period  $t = -1$  (prior to choice of major), expected utility for each college major is given by

$$V_{-1,k} = \gamma_k + v(a_k) + EV_{0,k}, \tag{1}$$

where the  $\gamma_1, \gamma_2, \dots, \gamma_K$  components represent the preferences or tastes for each college major  $k$  at the initial pre-graduation stage. We define "tastes" at the point when students are in college. These could be tastes for major-specific outcomes realized in college, such as the enjoyability of coursework, or major-specific post-graduation outcomes, such as non-pecuniary aspects of jobs.  $v(a_k)$  is the mapping of a student's ability in each major,  $a_1, \dots, a_K$  with  $a_k \geq 0$  for all  $k$ , to pre-graduation utility from each major. We assume  $\partial v(a_k)/\partial a_k \geq 0$ , reflecting that higher ability in a particular major improves performance in each major's coursework and reduces

---

<sup>1</sup>As described below in the Data section, in order to model the complete potential choice set, one of the "majors" is a "no graduation" (college drop-out) choice.

<sup>2</sup>To make clear how this timing convention is reflected in our survey design, note that we survey college students (1st-3rd year students) at period  $t = -1$ , prior to college graduation. We do not survey 4th year and later students because they may have already chosen a particular college major.



the effort cost of completing a major. Ability in coursework and ability in the labor market can be closely correlated, but we do not explicitly model this interaction since our data allows us to measure expected earnings in each field and beliefs about ability in each field directly.<sup>3</sup> Expectations are formed according to the beliefs in period  $t = -1$ .<sup>4</sup>

At period  $t = 0$ , the student realizes some preference shock and then chooses her college major. Expected utility at the time of graduation for each major  $k$  is given by

$$V_{0,k} = \eta_k + \beta EV_{1,k}, \quad (2)$$

where  $\eta_1, \eta_2, \dots, \eta_K$  are the period  $t = 0$  preference shocks that reflect any change in preferences that occur between the initial pre-major choice period  $t = -1$  and the period when the college major is chosen. In the Blass, Lach, and Manski (2010) taxonomy,  $\eta_k$  is "resolvable" uncertainty—uncertainty that is resolved at the point at which the choice of major is made.

After college graduation, the expected discounted sum of future post-graduation utility from each major  $k$  is given by

$$EV_{1,k} = \sum_{t=1}^T \beta^{t-1} \int u(X) dG(X|k, t), \quad (3)$$

where  $u(X)$  is the utility function that provides the mapping from the finite vector of events  $X$  to utility.  $X$  can include a wide range of events (e.g. earnings, labor supply, marriage, spousal earnings, and so on).  $G(X|k, t)$  represents the beliefs about the distribution of future events in period  $t$ , conditional on choice of major  $k$ . The distributions of future events  $G(X|k, t)$  represent "unresolvable" uncertainty as these events will not have occurred at the time of major choice. Beliefs are individual specific and based on current information, which, as discussed below, can be a mixture of public and private information. In the next sections, we refer to these beliefs as "self" beliefs, e.g., beliefs about what the individual would earn if she graduated with a business degree. Self beliefs are distinct from the "population" beliefs that students hold about the population distribution of some major characteristics, e.g., beliefs about the average earnings in the population of individuals who graduate with a business degree.

Individuals choose the college major that maximizes expected utility at period  $t = 0$ :

$$V_0^* = \max\{V_{0,k}, \dots, V_{0,K}\},$$

---

<sup>3</sup>In our data, we find that a student's self-reported ability rank in each major is highly correlated with self-reported expected future earnings in the field.

<sup>4</sup>Note for simplicity that (1) ignores any real separation of the  $t = -1$  and  $t = 0$  periods. We implicitly assume that the period  $t = -1$  is "just" before the decision making period in  $t = 0$ . Alternatively, we could write:  $V_{-1,k} = \gamma_k + v(a_k) + \beta EV_{0,k}$ . However, this is only a slight change from the present model since the discount rate would not be identified separately from the scale of the  $\eta_k$  shocks (2), and we can capture differences in utility flows from future post-graduation activities with a shift in the utility function (3).

At  $t = -1$ , each individual's expected probability of majoring in each of the  $k$  majors given beliefs is then given by integrating over the distribution of resolvable uncertainty:

$$\pi_k = \int 1\{V_{0,k}^* = V_0^*\} dF(\eta), \quad (4)$$

where  $F(\eta)$  is the joint distribution of  $\eta_1, \dots, \eta_K$ , and  $\sum_{k=1}^K \pi_k = 1$ .

### 3 Identification

In this section, we explore identification of the model using three types of data: i) commonly used revealed choice data in which we observe one choice of college major for each individual along with earnings in this major, ii) a cross-section of baseline (pre-treatment) beliefs, and iii) panel data including both pre- and post- treatment beliefs.

#### 3.1 Identification Using Actual Choice Data

We first consider identification under the typical revealed preference data in which we observe for each individual  $i$  their actual choice of major (i.e., the data are collected after college graduation). In revealed preference data, we typically observe a set of indicators for major choice, some measure(s) of ability, and some realizations of future events, such as future earnings in the chosen major. Let  $d_{1,i}, \dots, d_{i,K}$  be the set of indicators for these choices such that  $d_{k,i} = 1\{V_{0,k,i} = V_{0,i}^*\}$  for all  $k$ . From these revealed choices, we can identify the probability that each major is chosen:

$$\begin{aligned} P_k &\equiv pr(d_{k,i} = 1) \\ &= \int \pi_{k,i} dQ(\gamma_{1,i}, \dots, \gamma_{K,i}, a_{1,i}, \dots, a_{K,i}, G_i(X|t, 1), \dots, G_i(X|t, K)), \end{aligned}$$

where  $\sum_{k=1}^K P_k = 1$ .  $Q(\cdot)$  is the population distribution of tastes, abilities, and beliefs about future post-graduation events. Note that  $P_k$  is distinct from  $\pi_{k,i}$ :  $P_k$  is the probability major  $k$  was chosen, which is revealed in post-graduation data, whereas  $\pi_{k,i}$  is the belief about the future probability that major  $k$  will be chosen.

With this revealed preference data, the researcher faces the task of constructing elements of the utility function from actual observed data. In general, this requires four additional layers of assumptions:

i) an assumed mapping between revealed or actual post-graduation earnings to beliefs about earnings (or any other elements of post-graduation utility) when the major was chosen,

- ii) an assumed model for counterfactual beliefs about earnings (or any other elements of post-graduation utility) in majors not chosen,
- iii) an assumed mapping between measures of ability to beliefs about ability in each major,
- iv) an assumed distribution of tastes for all majors.

The prior literature makes various types of assumptions along these dimensions. Freeman (1971, 1975) for example assumes an adaptive expectations mapping between realized earnings and beliefs about earnings. Siow (1984) and Zarkin (1985) make perfect foresight (rational expectations) assumptions. Implicitly these models also assume that earnings are the same for all individuals. Other work, including Bamberger (1986), Berger (1988), Flyer (1997), Eide and Waehrer (1998), Montmarquette et al. (2002), and Befy et al. (2011) allow for some heterogeneity in earnings, across chosen and counterfactual majors, but assume rational expectations. Arcidiacono (2004) uses realized grade information during college and an assumed learning model in order to map grade measures to beliefs about ability in each major. Finally, previous research assumes some distribution of tastes for majors, usually an extreme value taste distribution, as in Berger (1988) and Arcidiacono (2004), or a normality assumption, as in Befy et al. (2011).

This approach overlooks the fact that subjective expectations may be different from objective measures, assumes that formation of expectations is homogeneous, makes nonverifiable assumptions on expectations, and uses choice data to infer decision rules conditional on maintained assumptions on expectations. This can be problematic since observed choices might be consistent with several combinations of expectations and preferences, and the list of underlying assumptions may not be valid (see Manski, 1993, for this inference problem in the context of how youth infer returns to schooling, and Manski, 2004, for a detailed discussion on this).

### 3.2 Identification Using Baseline Beliefs

We next turn to considering the identification if we have baseline beliefs data only, and do not have the post-treatment information from our information experiment. This is the data available, for example, in Delavande (2008), van der Klaauw and Wolpin (2008), Zafar (2009), Giustinelli (2010), Arcidiacono et al. (2011), Attanasio and Kaufmann (2011), and van der Klaauw (2011). The benefit of collecting belief information for outcomes in all possible choices is that this allows the researcher to relax assumptions about i) the mapping between realizations and beliefs for outcomes in the choice made, and ii) beliefs for outcomes in counterfactual choices not chosen.

In order to make the potential source of bias transparent, let the vector of relevant future events  $X$  be divided into a subset of observed (to the researcher, in the data) events  $X^o$  and unobserved events  $X^u$ :  $X = [X^o X^u]$ . Note in our context "observed" means future events that

the researcher asks respondents' expectations about and "unobserved" means any other events not inquired about. For any given student respondent  $i$ , we observe at the time of our survey (period  $t = -1$ , prior to college major choice):

D1) self-reported expectations of graduation for each of the  $K$  majors:  $\pi_{1,i}, \dots, \pi_{K,i}$ ,

D2) individual beliefs about the distribution of post-graduation future events conditional on major choice  $G_i^o(X^o|1, t), \dots, G_i(X^o|K, t)$  for all  $t = 1, \dots, T$ , and

D3) individual beliefs about ability in each of the majors  $a_{1,i}, \dots, a_{K,i}$ .

$G_i^o(X^o|k, t)$  are the observed beliefs which are self-reported by respondents in the survey. The distribution of the unobserved events, covering those events not collected in the beliefs data, is given by  $G_i^u(X^u|k, t)$ .

Given this data, we next investigate how much of the underlying choice model can be identified. Assume that the resolvable uncertainty preference shocks for each major are distributed i.i.d. extreme value across major choices and across each individual. With this assumption, (4) is

$$\pi_{k,i} = \frac{\exp\{\gamma_{k,i} + v(a_{k,i}) + \sum_{t=1}^T \beta^{t-1} \int u(X) dG_i(X|t, k)\}}{\sum_{j=1}^K \exp\{\gamma_{j,i} + v(a_{j,i}) + \sum_{t=1}^T \beta^{t-1} \int u(X) dG_i(X|t, j)\}}. \quad (5)$$

In the convenient log odds form, we can write the log odds of student  $i$  completing major  $k$  relative to a reference major  $\tilde{k}$  as

$$\begin{aligned} r_{k,i} &\equiv \ln \pi_{k,i} - \ln \pi_{\tilde{k},i} \\ &= \gamma_{k,i} - \gamma_{\tilde{k},i} + v(a_{k,i}) - v(a_{\tilde{k},i}) + EV_{1,k,i} - EV_{1,\tilde{k},i} \end{aligned} \quad (6)$$

Distinguishing between observed and unobserved events, we have

$$r_{k,i} = \gamma_{k,i} - \gamma_{\tilde{k},i} + v(a_{k,i}) - v(a_{\tilde{k},i}) + EV_{1,k,i}^o - EV_{1,\tilde{k},i}^o + \epsilon_{k,i}, \quad (7)$$

where  $\epsilon_{k,i} = EV_{1,k,i}^u - EV_{1,\tilde{k},i}^u$ ,

$$EV_{1,k,i}^o = \sum_{t=1}^T \beta^{t-1} \int u(X) dG_i^o(X^o|k, t),$$

$$EV_{1,k,i}^u = \sum_{t=1}^T \beta^{t-1} \int u(X) dG_i^u(X^u|k, t).$$

$\epsilon_{k,i}$  represents the "error" associated with the missing information on beliefs about post-graduation events not collected in the survey. This is simply the belief data counterpart to omitted variable error in revealed preference data, e.g., "missing" information about earn-

ings in counterfactual majors. Without loss of generality, we normalize  $\gamma_{\tilde{k},i} = 0$  for all  $i$  and  $E[\epsilon_{k,i}] = 0$  for all  $k$ .<sup>5</sup>

Collecting information about beliefs about earnings and ability has the advantage of obviating the need for assumptions mapping realized measures of ability to beliefs about ability in all fields. However, without any further modeling restrictions, we cannot separately identify the relative taste for each major  $\gamma_{ki}$  from the expected post-graduation future utility. The lack of identification holds since we can fully rationalize the data on expected choice probabilities as  $u(X) = 0$  for any vector  $X$  and  $r_{k,i} = \gamma_{k,i}$  for all  $k \neq \tilde{k}$ . Separately identifying  $EV_{1,k,i}$  from tastes could be achieved through a parametric restriction on the joint distribution of taste parameters  $\gamma_{k,i}$  (e.g., assuming a joint extreme value or normal distribution of tastes).<sup>6</sup> In the next section we propose a new strategy for identification using additional data derived from experimentally perturbed beliefs.

### 3.3 Identification using Experimental Variation

This section provides the basis for separately identifying tastes for majors from other utility components using experimental perturbations of beliefs. Our innovation is to note that if we can perturb the beliefs of the individuals so that at least some individuals form new beliefs  $G'_i(X|k) \neq G_i(X|k)$ , we could identify a parameterized utility function  $u(X)$ . We perturb individual beliefs by providing individuals information on general population characteristics regarding earnings and labor supply among those who have graduated with various majors (see Data section). To the extent that the individuals' self beliefs about earnings and other characteristics are i) linked to their beliefs about the population distribution of these characteristics and ii) they are mis-informed about the population characteristics, this new information may cause some individuals to update their own self beliefs. We use our experimental data to test whether individuals are mis-informed and to examine the extent to which individuals update their own self beliefs based on this new information. As we discuss below, we find substantial errors in population beliefs and logical self belief updating in response to our information treatment.

An important distinction between our panel generated using experimental variation and other longitudinal information on beliefs is that we collect beliefs data over a short period

---

<sup>5</sup>To see that there is no loss of generality, note that the original model and the model with  $\gamma_{\tilde{k},i} = 0$  for all  $i$  are equivalent by adding the major  $\tilde{k}$  taste parameter and return to the original model as  $\tilde{u}(X) = \gamma_{\tilde{k},i} + u(X)$ .

<sup>6</sup>For example, in our notation, Arcidiacono et al. (2011) assume that  $\delta_{k,i} = (\eta_{k,i} + \gamma_{k,i})$  is distributed i.i.d. extreme value. We make the same parametric assumption about the resolvable uncertainty  $\eta_{k,i}$ , but relax this assumption for the permanent taste component  $\gamma_{k,i}$ . As described below, our model is then a mixed logit model which uses the experimental perturbation of beliefs to generate panel data to separately identify a taste component.

of time, where the period before and after the information is provided in our experiment is separated by only a few minutes. This is in contrast to other studies (e.g., Lochner 2007; Stinebrickner and Stinebrickner 2010, 2011; Delavande, 2011; Paula, Shapira, and Todd, 2011) where the separation between beliefs observations is much longer, typically months or years. We can then credibly claim that the utility function, most notably the individual and major specific taste parameters, are truly time invariant in our context, and our experimentally derived panel data satisfies the standard fixed effects assumptions.

After providing information on the population distribution, our information treatment experiment augments the baseline information on self beliefs (D1, D2, and D3), with

D1') post-treatment self-reported expectations of graduating with each of the  $K$  majors:  $\pi'_{1,i}, \dots, \pi'_{K,i}$ ,

D2') post-treatment individual beliefs about the distribution of post-graduation future events conditional on major choice  $G_i^{o'}(X^o|1, t), \dots, G_i^{o'}(X^o|K, t)$ , and

D3') individual beliefs about ability in each of the majors  $a'_{1,i}, \dots, a'_{K,i}$ .

Assuming the distribution of resolvable uncertainty given by the taste shocks  $\eta$  is independent of the information treatment which perturbs beliefs, we can write the individual post-minus pre-treatment difference in the log odds of majoring in each major (relative to a reference major  $\tilde{k}$ ) as

$$r'_{k,i} - r_{k,i} = [\ln \pi'_{k,i} - \ln \pi'_{\tilde{k},i}] - [\ln \pi_{k,i} - \ln \pi_{\tilde{k},i}]$$

$$= v(a'_{k,i}) - v(a'_{\tilde{k},i}) - [v(a_{k,i}) - v(a_{\tilde{k},i})] + EV_{1,k,i}^{o'} - EV_{1,\tilde{k},i}^{o'} - [EV_{1,k,i}^o - EV_{1,\tilde{k},i}^o] + \epsilon'_{k,i} - \epsilon_{k,i}, \quad (8)$$

where

$$EV_{1,k,i}^{o'} = \sum_{t=1}^T \beta^{t-1} \int u(X) dG_i^{o'}(X^o|k, t).$$

Given this structure and a parameterized utility and ability functions  $u(X, \theta)$  and  $v(a_{k,i}, \alpha)$ , with finite dimensional unknown parameter vectors  $\theta$  and  $\alpha$ , we assume the following moment condition, which is the basis of our estimation strategy:

$$E[\Delta \epsilon_{k,i} | h(Z_i, \theta, \alpha)] = 0, \quad (9)$$

where  $\Delta \epsilon_{k,i} = \epsilon'_{k,i} - \epsilon_{k,i}$ ,  $Z_i = [G_i^o(X|1, t), \dots, G_i^o(X|K, t), G_i^{o'}(X|1, t), \dots, G_i^{o'}(X|K, t)]$ , and

$$h(Z_i, \theta, \alpha) = v(a'_{k,i}, \alpha) - v(a'_{\tilde{k},i}, \alpha) - [v(a_{k,i}, \alpha) - v(a_{\tilde{k},i}, \alpha)]$$

$$\begin{aligned}
& + \sum_{t=1}^T \beta^{t-1} \int u(X, \theta) dG_i^{o'}(X^o|k, t) - \sum_{t=1}^T \beta^{t-1} \int u(X, \theta) dG_i^{o'}(X^o|\tilde{k}, t) \\
& - \left[ \sum_{t=1}^T \beta^{t-1} \int u(X, \theta) dG_i^o(X^o|k, t) - \sum_{t=1}^T \beta^{t-1} \int u(X, \theta) dG_i^o(X^o|\tilde{k}, t) \right].
\end{aligned}$$

Note that with our data collection, the collection of beliefs for each individual, given by the vector  $Z_i$ , are *data* since we elicit these beliefs in our survey design.

Our identification assumption states that any changes in beliefs about unobserved events, contained in the  $\Delta\epsilon_{k,i}$  term, is mean-independent of the function of observed changes in beliefs given by  $h(Z_i, \theta)$ . This assumption is satisfied if the relative tastes for each major, given by the  $\gamma_{k,i}$  terms, the post-graduation utility function itself  $u(X, \theta)$ , and the current effort cost ability function  $v(a_{k,i}, \alpha)$  are invariant between the pre and post-treatment period. Given that the pre- and post- information treatment periods are separated by a few minutes in our survey design, we find this to be reasonable assumption.<sup>7</sup>

### 3.4 Example

We next consider a simple example to provide some intuition for our information experiment based identification strategy. Suppose there is a just a single post-graduation period  $T = 1$ , two majors  $k$  and  $\tilde{k}$ ,  $X$  includes one element (earnings)  $X = [w]$ , and we ignore the role of major specific ability. Students in period  $t = -1$  self-report their expected distribution of earnings given their beliefs. Suppose the utility function takes the simple linear form  $u(X) = \theta w$ ,  $\theta \geq 0$ .  $\theta$  is the marginal utility of earnings: high  $\theta$  indicates that college major choices are sensitive to earnings (relative to tastes), and low  $\theta$  indicates that college major choices are insensitive to earnings. In our empirical estimation, we consider richer life-cycle specifications of the utility function and collect an array of data about future events associated with majors.

In this simple example, pre-treatment expected post-graduation utility for student  $i$  is then:

$$\begin{aligned}
EV_{1,k,i}^o &= \sum_{t=1}^T \beta^{t-1} \int \theta w dG_i^o(w|k, t) \\
&= \theta \bar{w}_{k,i},
\end{aligned}$$

where  $\bar{w}_{k,i}$  is individual  $i$ 's beliefs about the average earnings she would receive if she were to graduate with major  $k$ . The information treatment provides new information to the student

---

<sup>7</sup>Note as in the typical panel model with homogeneous elements, we do not require that ALL individuals update their beliefs, only that some individuals update their beliefs. This is because we restrict the post-graduation utility function to be homogeneous, but allow heterogeneity in fixed taste parameters. In general if we have many belief changes, we could identify rich patterns of heterogeneity in the utility function as well.

on the population distribution of earnings, and following the information treatment, student  $i$  revises her beliefs about her future earnings in each major  $k$  and her future probability of graduating with each degree. Expected post-graduation utility for student  $i$  post-treatment is then:

$$\begin{aligned} EV_{1,k,i}^{o'} &= \sum_{t=1}^T \beta^{t-1} \int \theta w dG_i^{o'}(w|k, t) \\ &= \theta \bar{w}'_{k,i} \end{aligned}$$

The post- minus pre- treatment difference in log probabilities (relative to a reference major  $\tilde{k}$ ) is given by

$$r'_{k,i} - r_{k,i} = \theta(\bar{W}'_{k,i} - \bar{W}_{k,i}), \quad (10)$$

where  $\bar{W}'_{k,i} = \bar{w}'_{k,i} - \bar{w}'_{\tilde{k},i}$  and  $\bar{W}_{k,i} = \bar{w}_{k,i} - \bar{w}_{\tilde{k},i}$ . In this example, the data consist of post- and pre-treatment self probabilities of major in major  $k$  and a reference major  $k'$  ( $r'_{k,i}, r'_{\tilde{k},i}, r_{k,i}, r_{\tilde{k},i}$ ), and post- and pre- treatment beliefs about expected earnings in both majors: ( $\bar{w}'_{k,i}, \bar{w}'_{\tilde{k},i}, \bar{w}_{k,i}, \bar{w}_{\tilde{k},i}$ ). Re-arranging (10), we then have the following moment condition that identifies the marginal utility of earnings  $\theta$ :

$$\theta = E\left[\frac{r'_{k,i} - r_{k,i}}{\bar{W}'_{k,i} - \bar{W}_{k,i}}\right]. \quad (11)$$

The intuition for our identification strategy is clearly seen in (11). The numerator of this expression measures the extent of the relative probability revision from the pre-treatment to post-treatment period. The denominator of (11) measures the extent of the relative revision in self beliefs about earnings. The ratio of the revision of the self-reported revision in probabilities versus the revision in earnings identifies the marginal utility of earnings in major choice. If there is a large revision in probabilities relative to a small revision in earnings, then we conclude that  $\theta$  is large and earnings are an important factor in major choice. If however, there is little revision in probabilities relative to a large revision in earnings, then we conclude that  $\theta$  is low, and other factors such as tastes or abilities, not earnings, are the predominant consideration in major choice.<sup>8</sup>

---

<sup>8</sup>There is scope of course to consider heterogeneity in the utility function (e.g.  $\theta_i$  varies with  $i$ ), in addition to heterogeneity in permanent taste parameter  $\gamma_{k,i}$ . We discuss this possibility below in the estimation.



## 4 Data

### 4.1 Administration

Our data is from an original survey instrument administered to New York University (NYU) undergraduate students over a 3-week period, during May-June 2010. NYU is a large, selective, private university located in New York City. The students were recruited from the email list used by the Center for Experimental Social Sciences (CESS) at NYU. The study was limited to full time NYU students who were in their freshman, sophomore, or junior years, were at least 18 years of age, and US citizens. Upon agreeing to participate in the survey, students were sent an online link to the survey (constructed using the SurveyMonkey software). The students could use any internet connected computer to complete the survey. The students were given 2-3 days to start the survey before the link became inactive, and were told to complete the survey in one sitting. The survey took approximately 90 minutes to complete, and consisted of several parts. Students were not allowed to revise answers to any prior questions after new information treatments were received. Many of the questions had built-in logical checks (e.g. percent chances of an exhaustive set of events such as majors had to sum to 100). Students were compensated \$ 30 for successfully completing the survey.

### 4.2 Survey Instrument

The survey instrument consists of three distinct stages:

STAGE 1) Initial Stage: Respondents were asked about their population and self beliefs

STAGE 2) Intermediate Stage: Respondents were randomly selected to receive 1 of 4 possible information treatments. The information was reported on the screen and the respondents were asked to read this information before they continued. Respondents were then re-asked about population beliefs (on areas they were not provided information about) and self beliefs.

STAGE 3) Final Stage: Respondents were given all of the information contained in each of the 4 possible information treatments. Respondents were then re-asked about their self beliefs.

For the purposes of estimating the choice models in this paper, we used only the initial Stage 1 self beliefs (pre-treatment) and the final Stage 3 beliefs. Because of time constraints not all beliefs questions were asked in the intermediate second stage.

The information treatment consisted of statistics about the earnings and labor supply of the US population. Some of the information was general (e.g., mean earnings for all US workers), while other information was specific to individuals who had graduated in a specific major (e.g., mean earnings for all male college graduates with a degree in business or economics). Appendix Table A1 lists all of the information treatments. The information treatments were calculated by the authors using the Current Population Survey (for earnings and employment for the

general and college educated population) and the National Survey of College Graduates (for earnings and employment by college major). Details on the calculation of the statistics used in the information treatment are in the Appendix; this information was also provided to the survey respondents.

Our goal was to collect information on consequential life activities that would plausibly be key determinants of the utility gained from a college major. Because of time constraints, we were forced to make difficult choices in the aggregation of college majors and the breadth of belief questions. We aggregate college majors to 5 groups: 1) Business and Economics, 2) Engineering and Computer Science, 3) Humanities and Other Social Sciences, 4) Natural Sciences and Math, and 5) Never Graduate/Drop Out. We provided the respondents a link where they could see a detailed listing of college majors (taken from various NYU sources) and how each of these college majors mapped into the aggregate major categories. Given that we include a never graduate/drop out category, this list of college majors is exhaustive. Thus, we forced the self reported percent chance of majoring in these categories to sum to 100. Before the official survey began, survey respondents were first required to answer a few simple practice questions in order to familiarize themselves with the format of the questions.

Because we wanted to approximate life cycle utility from each major, we collected beliefs about both initial earnings- just after college graduation, and for later periods, when earnings might be believed to be much higher. We collected post-graduation beliefs for three periods: i) first year after college graduation (when most respondents would be aged 22-24), ii) when the respondent would be aged 30, and iii) when the respondent would be aged 45. At each of those periods, we ask respondents for their beliefs about their own earnings (including measures of dispersion), work status (not working, part time, full time), probability of marriage, and spouse's earnings. An example question on expected earnings at age 30: "*If you received a Bachelor's degree in each of the following major categories and you were working FULL TIME when you are 30 years old what do you believe is the average amount that you would earn per year?*"<sup>9</sup> The instructions emphasized to the respondents that their answers should reflect their own beliefs, and not use any outside information.<sup>10</sup>

Our questions on earnings were intended to elicit beliefs about the distribution of future earnings. We asked three questions on earnings: beliefs about expected (average) earnings,

---

<sup>9</sup>We also provided definitions of working full time ("working at least 35 hours per week and 45 weeks per year"). Individuals were instructed to consider in their response the possibility they might receive an advanced/graduate degree by age 30. Therefore, the beliefs about earnings we collected incorporated beliefs about the possibility of other degrees earned in the future and how these degrees would affect earnings. We also instructed respondents to ignore the effects of price inflation.

<sup>10</sup>We included these instructions: "*This survey asks YOUR BELIEFS about the earnings among different groups. Although you may not know the answer to a question with certainty, please answer each question as best you can. Please do not consult any outside references (internet or otherwise) or discuss these questions with any other people. This study is about YOUR BELIEFS, not the accuracy of information on the internet.*"

beliefs about the percent chance earnings would exceed \$35,000, and percent change earnings would exceed \$85,000. As detailed below, we use this information to estimate individual specific distribution of earnings beliefs. Beliefs about spouse's earnings conditional on *own* major were also elicited in a similar way.

The probability of marriage was elicited as follows: "*What do you believe is the percent chance that you will be married by age 30 if you received a Bachelor's degree in each of the following?*"

Beliefs about labor supply were elicited conditional on marriage. For example, labor supply conditional on being not married at age 30 was asked as follows: "*What do you believe is the percent chance of the following: (1) You are working full time; (2) You are working part time; (3) You are not working at all, when you are 30 years old if you are NOT married and you received a Bachelor's degree in each of the following?*"

Respondents were also asked about their spouse's labor supply and field of study, conditional on own field of study. Beliefs about average hours of work for each major were also asked. The full survey questionnaire is available from the authors upon request.

### **4.3 Sample Selection and Descriptive Statistics**

Our sample is constructed using the following steps. First, we drop 6 students who report that they are in the 4th year of school or higher, violating the recruitment criteria. Second, we censor reported beliefs about full time annual earnings (population or self earnings) so that earnings below \$10,000 are recorded as \$10,000 and earnings reported above \$500,000 are recorded as \$500,000. Third, we drop nearly 25 percent of the sample who report too radical changes in age 30 earnings (change of positive \$50,000 or negative \$50,000 between initial and final information treatments) for any of the majors. Fourth, we exclude individuals who report a change in graduation probabilities of greater than 0.5 in magnitude in any of the 5 major categories. The latter sample selection requirements eliminates a minority of respondents who either made errors in filling out the survey or simply did not take the survey seriously. In addition, we recode all reported extreme probabilities of 0 to 0.001 and 1 to 0.999. This follows Blass et al. (2010) who argue that dropping individuals with extreme probabilities would induce a sample selection bias in the resulting estimates.

The final sample consists of 359 individual observations and  $359 \times 5 \times 2 = 3,590$  total (person  $\times$  major  $\times$  pre and post treatment) responses. 36 percent of the sample (129 respondents) is male, 40 percent is white and 45 percent is Asian. The mean age of the respondents is about 20, with 40 percent of respondents freshmen, 37 percent sophomores, and the remaining 24 percent juniors. The average grade point average of our sample is 3.5 (on a 4.0 scale), and the students have an average Scholastic Aptitude Test (SAT) math score of 709, and a verbal

score of 691 (with a maximum score of 800). These correspond to the 93rd percentile and 95th percentile of the math and verbal score population distributions, respectively. Therefore, our sample represents a high ability group of college students.

## 5 Earnings Beliefs and Belief Updating

We begin our data analysis by focusing on expected earnings at age 30. Beliefs about other future events are discussed below and incorporated into the life-cycle model of post-graduation utility. Here we examine heterogeneity in beliefs about population average earnings, self beliefs about what each individual expects to earn in different majors, self beliefs about spouse's earnings conditional on own major, and revisions in self beliefs following the information treatment.

### 5.1 Population Beliefs About Earnings

We asked the following question for a randomly selected subset of respondents: "*Among all male college graduates currently aged 30 who work full time and received a Bachelor's degree in each of the following major categories, what is the average amount that you believe these workers currently earn per year?*" For another randomly selected group of respondents, we asked the corresponding question for women. A subset of respondents were asked the population earnings for both males and females.

Table 1 reports the mean and standard deviation of male and female respondents' beliefs about US population earnings of men and women by the 5 major fields, including college drop-out, the no degree "major". Examining first the beliefs among male students, we see that mean male belief of age 30 female full time earnings varies from \$30,100 for college drop-outs to \$65,900 for graduates with degrees in economics or business. Students believe humanities and arts has the lowest average earnings among the graduating majors (\$48,400). Engineering and computer science graduates are believed to have earnings close to economics and business, followed by natural science majors. There is considerable heterogeneity in beliefs as indicated by the large standard deviation in population beliefs. For example, for the economics and business field, the 5th percentile of the belief distribution in our sample is \$10,000, the 50th percentile is \$70,000, and the 95th percentile is \$100,000.

Based on responses of students who reported population earnings for both males and females, we can construct the perceived gender gap in earnings. This is reported in column (5) of the table. Males expect a wage gap in their favor in each of the five major fields, with the gap varying from -3.23% for humanities/arts to -7.41% in college drop-out.

The lower panel of Table 1 shows that female students have beliefs similar to those of male students about relative earnings in the majors, and expect the highest average earnings in

economics or business, followed by engineering and computer science, and the lowest earnings in humanities and arts among the graduating majors. However, relative to male students, female students believe average earnings to be higher in all fields for both females and males. Female students, like their male counterparts, perceive a wage gap in favor of men in all the fields, but report a larger gender gap in earnings for all graduating majors than men.

### 5.1.1 Errors in Population Beliefs

Columns (2) and (4) of Table 1 report the percent "error" in these beliefs relative to the information treatment "truth" we provided (see Table A1 for true population earnings that were revealed in the information treatments). We calculate errors as truth minus belief, so that a positive (negative) error indicates that the student under-estimates(over-estimates) the truth. As students revise their self earnings in response to the information treatment, the sign of the error should match the sign of the self earnings revision: positive errors should cause an upward self earnings revision and negative errors should cause a downward self earnings revision. We find support for this kind of logical updating below.

Table 1 reports that the mean percent error is positive for the majority of the fields and sub-samples, indicating that on average students are under-estimating the earnings in most fields (exceptions are mean errors for humanities/arts for female respondents, and economics/business for both male and female respondents, which are over-estimated). The errors in many categories are substantial, with students under-estimating full time earnings for engineering and computer science graduates by 7.3 and 23.4 percent, depending on sub-group and sample. Reflecting the dispersion in baseline beliefs, there is considerable heterogeneity in errors, with non-trivial numbers of students making both positive and negative errors in all categories. The top panel of Figure A1 shows the male student distribution of errors regarding full time men's earnings with a economics or business degree. While the mean of this error distribution is 6.74 percent, the 5th percentile is -34.2 percent and the 95th percentile is 86.6 percent.

The last two columns of Table 1 show that, while both male and female students correctly perceive the wage gap to be negative, i.e., in favor of males in all fields, they substantially underestimate the wage gender gap, with male students underestimating the gender gap more than female students. This underestimation is particularly striking for the "not graduate" category where the actual gender gap is -27.6 percent (i.e., earnings are 27.6 percent higher for male college drop-outs relative to corresponding females), but female students expect it to be close to zero and male students expect it to be about -7 percent. Engineering/computer science and humanities/arts are the only fields where the discrepancy between the actual and perceived wage gender gap is less than 10 percentage points.

## 5.2 Self Beliefs About Earnings

Next, we turn to self beliefs about *own* earnings at age 30 if the respondent were to graduate in each major. For all respondents, we asked "*If you received a Bachelor's degree in each of the following major categories and you were working full time when you are 30 years old what do you believe is the average amount that you would earn per year?*" The first column of Table 2 provides the average and standard deviation of the distribution of reported self earnings in our sample before the information treatment was provided. The second column of Table 2 provides the percent revision in self earnings after the information treatment. In general, students believe their self earnings will exceed the population earnings for the US, with the average self earnings across all of the major fields higher than the corresponding average population belief about earnings reported in Table 1. Looking across majors in column (1), we see that self earnings beliefs follow the same pattern as the population beliefs, with students believing their earnings will be highest if they complete a major in the economics/business and engineering/computer science categories, and lowest if they do not graduate or graduate in a humanities and arts field. There is a clear pattern of a perceived gender gap in self earnings as the average beliefs about self earnings for men exceeds those for women. Like the population beliefs, there is substantial heterogeneity in self beliefs, as seen in the large standard deviations (relative to the means). The middle panel in Figure A1 shows the distribution of male beliefs for earnings if they were to complete a major in economics or business. The 5th percentile of the distribution is \$50,000, the 50th percentile is \$90,000, and the 95th percentile is \$150,000.

Table A2 provides the baseline, pre-treatment, correlation in earnings across fields. We see that for both male and female students, there is a generally high correlation in self earnings across fields: Individuals who believe they will have high earnings in one field also believe they will have high earnings in other fields. This cross-major correlation is higher for men than women, indicating that women believe their earnings advantage is more specialized. Comparing the correlations across fields, we see a higher correlations in earnings belief across technical or mathematical intensive fields like economics/business and engineering/computer science compared to humanities/arts and economics/business.

### 5.2.1 Revisions of Self Beliefs

The second column of Table 2 reports the mean and standard deviation of the distribution of percent post minus pre-treatment changes in self beliefs about earnings. There is considerable heterogeneity in the revisions of self beliefs. Looking across categories, the average of the percent revisions distribution varies from about -47 percent (downward revision) to +37 percent (upward revision). For both male and female students, average revisions in the two highest earning categories—economics/business and engineering/computer science—are negative, while average

revisions in the lowest earning field—the not graduate category—are substantially positive. As indicated by the standard deviations, within categories there is considerable heterogeneity. The bottom panel of Figure A1 shows the dispersion in male students’ revisions for earnings in economics/business: the 5th percentile of the percentage earnings revision is -38 percent, the 50th percentile is zero percent, and the 95th percentile is +33 percent. For female students, the 5th, 50th and 95th percentiles are -40 percent, -12.5 percent, and +30 percent, respectively.

### 5.2.2 Uncertainty of Self Beliefs

While a very large literature has studied the average returns to schooling choices, there is little empirical work on the role risk plays in educational choices (Saks and Shore, 2005; Nielsen and Vissing-Jorgensen, 2006). Attanasio and Kaufmann (2011) is the only other study that collects data on risk perceptions of schooling choices. We asked respondents about the percent chance that their own earnings at age 30 would exceed \$35,000 and \$85,000.<sup>11</sup> We fit each student’s response to these questions as well as the reported average earnings for each field to a log-normal distribution, and obtain individual field specific parameters of the earnings distribution. The third column of Table 2 shows the average and standard deviation of the individual standard deviations of the earnings distributions for each field before the information treatment was provided. Male students believe the variance to be the largest for economics/business and engineering amongst graduating majors. For females, highest uncertainty is perceived for economics/business and natural sciences. The highest level of uncertainty is reported for the not graduate category by both male and female students. This is not surprising because the not graduate category is the least likely to be chosen by our respondents.

Column (4) of Table 2 reports the uncertainty of respondents excluding those who report the field to be their most likely major. This is to test if the perceived earnings uncertainty in a field is different conditional on whether the respondent intends to choose it or not. Of the eight possible pairwise comparisons (of whether the uncertainty of in-major students is equivalent to that of out-major students), only one is rejected at the 5% level. This suggests that students intending to major in a field do not have any less uncertainty about earnings than those who do not intend to major in the field. Column (5) of the table reports the uncertainty in earnings post-treatment. Earnings uncertainty decreases across all majors for both males and females (with the exception being the not graduate field for female students).

---

<sup>11</sup>The question was asked as follows: "What do you believe is the percent chance that you would earn: (1) At least \$85,000 per year, (2) At least \$35,000 per year, when you are 30 years old if you worked full time and you received a Bachelor’s degree in each of the following major categories?"

### 5.3 Beliefs about Spouse Earnings

One potentially important consideration of major choice may be the types of potential spouses one might marry. Recent empirical papers suggest that investment in education generates returns in the marriage market. Ge (2010) estimates a structural dynamic (partial equilibrium) model of college attendance using the NLSY 1979, and shows that marriage plays a significant role in a female's decision to attend college. Lafortune (2010) shows that a worsening of marriage market conditions spurs higher pre-marital investments—in particular for males—in her sample of second-generation Americans born around the turn of the twentieth century, and argues that part of this occurs through the anticipated shift in after-marriage bargaining power. Attanasio and Kaufmann (2011) find that marriage market considerations are important in females' schooling choices in Mexico. Evidence of the effect of marriage market considerations on educational choices is inferred indirectly in these studies. We investigate this in a direct way, and asked respondents about the earnings of their potential spouse *if* they were to be married at age 30 and their spouse worked full-time: "*What do you believe is the average amount that your spouse would earn per year if you received a Bachelor's degree in each of the following major categories?*" Importantly, we emphasized to respondents that they were to report beliefs about their spouse's earnings conditional on their *own* major, not the potential spouse's major. Column (6) of Table 2 reports the mean and standard deviation of beliefs about spouse's earnings. Compared to beliefs about own earnings in column (1), male students believe their spouse's earnings will be below their own earnings in every major category, while female students believe their spouse's earnings will exceed their own earnings. There are substantial differences in spousal earnings across own major choices, with both male and female students expecting their spouse's earnings to be the highest if they themselves majored in economics/business, and lowest if they graduated in humanities/arts (among graduating majors). The relative spousal earnings for own major are similar to the relative self earnings for own major. These patterns indicate that students perceive sorting of spouses by own major choice, and is suggestive of assortative mating by field of study.<sup>12</sup>

Column (7) of Table 2 indicates that the information treatment induced considerable revisions in beliefs about spousal earnings, with the mean of the distribution of spousal beliefs shifting upward in almost all cases.

### 5.4 Self Beliefs and Population Beliefs

We next examine whether population beliefs regarding earnings and associated errors relate to self beliefs and self beliefs revisions. Table 3 estimates a series of reduced form regressions. In

---

<sup>12</sup>The fact that there is assortative mating by education (more precisely, years of schooling) in the US is well documented (Mare, 1991; Pencavel, 1998).



the first 3 columns, we use only the baseline, pre-treatment data, and the dependent variable is the individual's (log) expected self earnings in each field. We pool all of the majors together, and in some specifications include separate intercepts or major-specific fixed effects (dummy variables). We regress self earnings in each field on the individual's (log) belief about the population average earnings in that field. The estimates indicate that population beliefs are strongly and statistically significantly related to beliefs about self earnings. The log-log form of the regressions gives the coefficient estimates an "elasticity" interpretation: the coefficient of 0.96 in column (1) indicates that a 1 percent increase in population beliefs about average earnings increases beliefs about own earnings by 0.96 percent. The estimated relationship is reduced only slightly as we add major-specific fixed effects and covariates for individual characteristics.

Columns (4) and (5) of Table 3 examine whether the *revisions* in self-earnings are related to *errors* in population beliefs. These regressions indicate the extent to which the information treatments we provide influence individual beliefs about earnings. We regress log earnings revision in self earnings (post minus pre-treatment) on the log relative error about population earnings ( $\log(\text{truth}/\text{belief})$ ). The coefficient estimates are positive and statistically significant at the 5 percent level. The coefficient estimate of 0.196 indicates that a 1 percent error (underestimate of population earnings) is associated with a 0.196 percent upward revision of self earnings. The relatively "inelastic" response of revisions in self beliefs to population errors suggests that self beliefs about earnings are not entirely linked to the type of public population information we provide. Heterogeneous private information on the abilities and future earnings prospects of individuals may cause individuals to have an inelastic response to population information.

## 6 Major Choice and Post-Graduation Utility

We next examine how beliefs about elements of future, post-graduation utility, including own earnings, relate to self-reported beliefs about majoring in the different fields. In this section we report estimates from a number of reduced form type regressions, and in the following section we report estimates from a structural life-cycle utility model.

### 6.1 College Major Beliefs

Self beliefs about the probability of graduating with a major in each of the categories were elicited as follows: "*What do you believe is the percent chance (or chances out of 100) that you would either graduate from NYU with a major in the following major categories or that you*

would never graduate/drop-out (*i.e.*, you will never receive a Bachelor's degree from NYU or any other university)?" Percent chance was converted to (0, 1) probabilities.

Table 4 provides descriptive statistics of the expected major field probabilities for male and female students. For male students, the most likely major is economics/business at 38 percent, followed by humanities/arts at 32 percent. For women, the most likely major is humanities at 50 percent followed by economics/business at 27 percent. The probability of not graduating at all is about 3 percent for men and 2 percent for women. Figure A2, which presents the distribution of (log) expected major field probabilities for male and female students, shows there is considerable dispersion in beliefs about future degrees. The distributions are bi-modal for most majors, with a considerable mass of individuals reporting a small or no chance of majoring in each field and another mass of individuals reporting a large or near perfect certainty of graduating in the field.

Figure 1 provides the post minus pre-treatment change in log beliefs for male and female students about majoring in each field (relative to humanities):  $r_{k,i} - r'_{k,i}$  from equation (8). The mean of the distribution of log odds changes is positive for all fields and for both male and female students (see last column of Table 4), indicating that after the information treatment, students on average revised their expected probability of majoring in non-humanities/arts fields upward relative to humanities/arts. However, as indicated by Figure 1, there were a substantial number of male and female respondents who revised their expected relative major choice downward, and believed they were more likely to major in humanities/arts relative to the other majors. About 1/3 of the sample reported no change in the probability of majoring in any of the fields following the information treatment. The largest upward changes occurred for the high earning fields (economics/business and engineering/computer science), especially for women. For example, the average log odds for male students of majoring in economics/business increased by 28 percentage points, from pre-treatment odds of 61 percent more likely to major in economics/business relative to humanities to 89 percent post-treatment. For women, the log odds of majoring in business/economics relative to humanities increased 53 percentage points from -132 percent to -79 percent (negative odds indicate more likely to major in humanities/arts than business/economics). After the information treatment, women are still more likely to major in humanities/arts than business/economics, but the difference in expected probabilities declined substantially.

## 6.2 College Major Beliefs and Self Beliefs about Own Earnings

We next examine the relationship between beliefs about college major choices and future earnings. The first three columns of Table 5 estimate a series of reduced form regressions using log expected probability of majoring in each field (relative to humanities/arts) as the dependent variable and log self beliefs about earnings at age 30 (relative to humanities/arts) as the

independent variable. The regressions take the form:

$$(\ln \pi_{k,i} - \ln \pi_{\tilde{k},i}) = \beta_0 + \beta_1(\ln \bar{w}_{k,i} - \ln \bar{w}_{\tilde{k},i}) + C'_i \delta + \nu_k + \omega_{k,i} \quad (12)$$

where from (6), the residual error term is

$$\omega_{k,i} = \gamma_{k,i} - \gamma_{\tilde{k},i} + v(a_{k,i}) - v(a_{\tilde{k},i}) + \epsilon_{k,i}.$$

The residual error in this cross-sectional regression consists of unobserved relative tastes  $\gamma_{k,i} - \gamma_{\tilde{k},i}$ , unobserved relative abilities  $v(a_{k,i}) - v(a_{\tilde{k},i})$ , and a component  $\epsilon_{k,i}$ , which reflects all other residual components. The reference major  $\tilde{k}$  in these regressions is humanities/arts,  $\bar{w}_{k,i}$  is beliefs about age 30 earnings in major  $k$ ,  $C_i$  is a vector of individual specific characteristics, and  $\nu_k$  is a major  $k$  fixed effect.

The log-log format of these regressions gives the estimates of  $\beta_1$  a "choice elasticity" interpretation. We estimate that a 1 percent increase in beliefs about self earnings in a major (relative to self earnings in humanities/arts) increases the log odds of majoring in that field (relative to humanities/arts) by about 2 percent. This estimate is robust to the inclusion of a wide array of individual characteristics and major fixed effects. The estimates indicate that beliefs about future relative self earnings are strongly associated with beliefs about future relative major choices: individuals appear to select into majors that they believe will provide them with the highest earnings. Importantly, because we have beliefs about earnings for all fields, this type of regression avoids the selection issue inherent in using actual major choice and the actual earnings in that one major, and omitting counterfactual earnings in majors not chosen.

The regressions in columns (1)-(3) of Table 5 are cross-sectional based regressions using only the baseline pre-treatment beliefs. As described in the identification section, the major drawback to using only baseline beliefs is that one cannot separately identify taste or ability components from earnings components. In the reduced form of (12), the residual  $(\gamma_{k,i} + \epsilon_{k,i})$  contains individual components reflecting individual variation in tastes and abilities in each of the majors. A concern is therefore the cross-sectional estimates of the relationship between choices and earnings could be biased if beliefs about earnings are correlated with beliefs about ability or tastes for the majors. To resolve this problem, column (4) of Table 5 estimates the reduced form model (12) in individual (within) differences to net out the individual taste and ability components  $(\gamma_{k,i} - \gamma_{\tilde{k},i} + v(a_{k,i}) - v(a_{\tilde{k},i}))$ :

$$\begin{aligned} & [(\ln \pi'_{k,i} - \ln \pi'_{\tilde{k},i}) - (\ln \pi_{k,i} - \ln \pi_{\tilde{k},i})] \\ & = \beta_0 + \beta_1[(\ln \bar{w}'_{k,i} - \ln \bar{w}'_{\tilde{k},i}) - (\ln \bar{w}_{k,i} - \ln \bar{w}_{\tilde{k},i})] + \nu_k + \epsilon'_{k,i} - \epsilon_{k,i}, \end{aligned} \quad (13)$$

where  $\pi'_{k,i}$  and  $\bar{w}'_{k,i}$  are post-treatment observations of choice probabilities and expected earn-

ings. The estimates of this model are equivalent to adding individual fixed effects (FE) as individual dummy variable indicators to (12).

Using the post- and pre-treatment panel data with individual FE, we estimate  $\beta_1$ , the choice elasticity with respect to beliefs about earnings, at 0.28. The FE estimate is substantially smaller than the estimate of around 2 using the cross-sectional OLS estimator. The FE estimate is statistically significant at the 7 percent level (p-value of 0.067), and significantly different from the cross-sectional/OLS estimates in Columns (1)-(3) at the 5 percent level. The difference between the FE/panel and OLS/cross-sectional estimates suggests that the individual tastes and ability components are positively correlated with beliefs about earnings, and this positive correlation is severely upwardly biasing the estimates in the cross-section.

### **6.3 College Major Beliefs and Self Beliefs about Spouse's Earnings**

We next turn to adding other potential elements of post-graduation utility to the reduced form log odds regression framework. As mentioned above, a potential consideration of major choice may be the types of potential spouses one might marry. Therefore, beliefs about spouse's earnings may be related to college major choice. Columns (5)-(8) of Table 5 examine the responsiveness of beliefs about major choices to spousal earnings. Columns (5) and (6) use the cross-sectional design, including major fixed effects and individual covariates. Column (5) shows that beliefs about spousal earnings are statistically significantly related to the beliefs about major choice for both male and female students, with the estimate being larger for male students. In column (6) we include both own earnings and spousal earnings. A one percent increase in spousal earnings in a major (relative to humanities/arts) increases the odds of graduating with that major by about 1.01 percent for males (p-value 0.003), and 0.395 percent for females (p-value 0.116). Including spousal earnings reduces slightly the coefficient on own earnings to 1.94 (from around 2.15 in column (3)). Own earnings continue to be a statistically significant factor for major choice, with spousal earnings having a considerably lower effect on major choice. Turning to the fixed effects estimates using the post and pre-treatment differences in columns (7) and (8), we see that both own and spousal earnings revisions are positively related to revisions in beliefs about major choice. The own earnings elasticity in column (8) including spousal earnings is slightly smaller than that including own earnings alone, and the coefficient is significantly different from zero at the 11 percent level (p-value 0.109). The spousal earnings coefficient is smaller than the own earnings coefficient for both males and females (p-value of 0.107 for males, and 0.132 for females), indicating that own earnings are more important to major choice than spouse's earnings. Self and spousal earnings are jointly significant in the regression reported in column (8).

## 6.4 College Major Beliefs and Self Beliefs about Ability

Ability in each major could be a factor in expectations about future earnings, and may affect the likelihood of a student completing required coursework necessary to graduate in each major. We asked the following question: "*Consider the situation where either you graduate with a Bachelor's degree in each of the following major categories or you never graduate/drop out. Think about the other individuals (at NYU and other universities) who will graduate in each of these categories or never graduate/drop out. On a ranking scale of 1-100, where do you think you would rank in terms of ability when compared to all individuals in that category?*" To provide easier interpretation, we re-scaled the ability beliefs such that 100 represents highest ability and 1 represents lowest ability. Table 6 provides descriptive statistics for the ability rank beliefs. In general, male students believe they have higher relative ability than female students- this is consistent with evidence that women tend to be less confident than men (Weinberger, 2004; Niederle and Vesterlund, 2007). For both male and female students, lowest believed ability is in engineering and computer science (53 for male students and 46 for female students). The highest average beliefs about ability for women are in humanities, whereas for male students it is in the not graduate category.

The second column of Table 6 reports the ability revisions after the information treatment. For almost all categories, the average ability revision is upward: After receiving the earnings and labor supply information, the students believe they are more able than they were before. This likely reflects that fact that most students under-estimated the average earnings in the population. The only exception to the positive ability updating was humanities/arts for female students where the average ability rank fell somewhat following the information treatment.

We next turn to examining whether self beliefs about ability relate to beliefs about future major choices. Columns (9)-(12) of Table 5 examine the responsiveness of beliefs about major choices to ability. Ability rank in a major (relative to ability rank in humanities/arts) is positively and significantly related to reported log odds of graduating in that major (relative to humanities/arts). Column (9) indicates that a 1 percent increase in ability rank in a major is associated with a about a 2/3 percent increase in odds of completing that major. In column (10), we add self beliefs about own earnings and spouse's earnings at age 30. Reflecting the positive correlation between ability beliefs and self earnings, the ability rank coefficient and self earnings coefficient are both smaller than when either are included separately in the regression. Self earnings, spouse's earnings, and ability are all jointly statistically significant in these regressions with the expected positive sign on each.

Turning to the individual fixed effect analysis using revisions in log odds as the dependent variable and revisions in ability as the regressor in column (11), we see a smaller coefficient of 0.11 on log rank ability than in the cross-sectional analysis. Mirroring the results with

own earnings, it appears that the unobserved individual specific taste component is positively correlated with beliefs about ability and this positive correlation biases upward the ability coefficient in the cross-sectional analysis in columns (9) and (10). Adding own earnings and spouse's earnings in column (12) has little effect on the ability coefficient, and it continues to be precise at the 5% level (p-value 0.015). Compared to the previous specification omitting ability (column (8)), the own earnings coefficient is slightly smaller and less precise, while the spouse's earnings coefficients are little changed.

## 7 Structural Estimates

### 7.1 Empirical Model of Post-Graduation Utility

In this section, we develop the specification of post-graduation utility (periods  $t = 1, \dots, T$ ). Each individual from college graduation to retirement makes a series of decisions regarding labor supply and marriage. At college graduation, we assume each individual is single and has obtained a degree in particular field  $k = 1, \dots, K$ .

In defining the utility function, we distinguish between two states: married and single. The flow utility in period  $t$  if the agent is single is given by  $U_{S,t} = u_S(c_{S,1,t})$ , where  $c_{S,1,t}$  is the individual's period  $t$  consumption. The own utility for an individual if married is given by  $U_{M,t} = u_M(c_{M,1,t}, c_{M,2,t})$ , where  $c_{M,1,t}$  is consumption of the individual and  $c_{M,2,t}$  is the consumption of the individual's spouse.  $U_{M,t}$  defines the *own* utility flow in period  $t$  from being married, not the household total utility for both spouses. Our specification of the utility function allows for the possibility that the individual agent may derive utility from the consumption of his or her spouse. Flow utility over the two states is then given by  $U_t = m_t U_{M,t} + (1 - m_t) U_{S,t}$ , where  $m_t = 1$  indicates marriage, and  $m_t = 0$  indicates single status at period  $t$ . The future events in  $u(X)$  from (3) are then the sequence of own and spousal consumption across both the married and single states:  $X^o = [\{c_{S,1,t}, c_{M,1,t}, c_{M,2,t}, m_t\}_{t=1}^T]$ .

We specify the utility functions with CRRA forms. When single, the utility function is given by  $u_S(c_{S,1,t}) = \phi_1 \frac{c_{S,1,t}^{1-\rho_1}}{1-\rho_1}$ , with  $\phi_1 \in (0, \infty)$  and  $\rho_1 \in (0, \infty)$ .  $1/\rho_1$  is the intertemporal elasticity of substitution (IES) for own consumption (in this specification,  $\rho_1$  is the coefficient of relative risk aversion). When married, we specify a commonly used specification where utility is a sum of own and spouse's utility:  $u_M(c_{M,1,t}, c_{M,2,t}) = u_{M,1}(c_{M,1,t}) + u_{M,2}(c_{M,2,t})$ .

Own utility while married uses the same preference structure while single (although the consumption level may be different under marriage, as we describe below):  $u_{M,1}(c_{M,1,t}) = \phi_1 \frac{c_{M,1,t}^{1-\rho_1}}{1-\rho_1}$ . Spousal preferences over consumption are allowed to be different from preferences over own consumption:  $u_{M,2}(c_{M,2,t}) = \phi_2 \frac{c_{M,2,t}^{1-\rho_2}}{1-\rho_2}$ , with  $\phi_2 \in (0, \infty)$  and  $\rho_2 \in (0, \infty)$ .  $1/\rho_2$  provides the

IES for spouse’s consumption.

We use the individual’s self beliefs about own earnings and labor supply and use the individual’s self beliefs about potential spousal earnings and labor supply to define consumption levels under the single and married states. We do not model borrowing and savings and assume consumption in each period is equal to current period earnings.<sup>13</sup> Because we ask individuals about full time equivalent earnings, we combine the beliefs about labor supply and full time earnings to define earnings in any given period. Own and spousal earnings are modeled as  $y_{1,t} = w_{FT,1,t}FT_{1,t} + w_{FT,1,t}(h_{PT,1,t}/h_{FT,1,t})PT_{1,t}$  and  $y_{2,t} = w_{FT,2,t}FT_{2,t} + w_{FT,2,t}(h_{PT,2,t}/h_{FT,1,t})PT_{2,t}$ , where  $w_{FT,q,t}$  are full time earnings ( $q = 1$  own,  $q = 2$  spouse),  $FT_{q,t} \in \{0, 1\}$  is an indicator if working full-time,  $PT_{q,t} \in \{0, 1\}$  is an indicator for working part-time,  $h_{FT,q,t}$  is full time hours, and  $h_{PT,q,t}$  is part-time hours. For each potential major, we ask respondents for their beliefs about the probability of working full or part-time, if single or married, the probability their potential spouse works full or part-time if married, and beliefs about average hours of work for each major. We allow an individual’s beliefs about the future distribution of full-time and part-time probabilities to depend on marriage, and therefore earnings and consumption also depend on marriage.

Consumption conditional on marriage is then given by  $c_{S,1,t} = y_{1,t}$  (own consumption when single),  $c_{M,1,t} = \kappa_1(y_{1,t} + y_{2,t})$  (own consumption when married), and  $c_{M,2,t} = (1 - \kappa_1)(y_{1,t} + y_{2,t})$  (spousal consumption when married).  $\kappa_1 \in (0, 1)$  is the share parameter which indicates how much of total household earnings is consumed by each spouse.<sup>14</sup>

## 7.2 Estimation

We estimate the parameters of the utility function using the pre- and post-information beliefs. Because of time limitations, we were forced to ask a limited set of questions: we cannot ask respondents to report full time earnings for all post-graduation periods and we cannot ask an infinite number of questions in order to provide a non-parametric estimate of the distribution of beliefs. Section C in the Appendix describes our approximations of the full life-cycle beliefs from the given data. It is important to emphasize that these approximations of beliefs are entirely individual specific: we make no assumption regarding the distribution of beliefs in the

---

<sup>13</sup>One has two alternatives in adding borrowing and savings behavior to a model such as this. First, following the earnings and labor supply questions, one could directly ask respondents about future consumption, borrowing, savings, or asset levels. However, framing these types of questions in a meaningful way for respondents may be quite difficult. Second, one could use traditional observational data to estimate a model of borrowing and saving and combine this model with the current model allowing consumption to be endogenous given earnings and labor supply.

<sup>14</sup>We have also experimented with functions that allow public goods, such that consumption of each spouse when married can exceed total resources. In some preliminary estimation, we found that these more general models were at best only weakly identified.

population.

The estimator is based on the moment condition (9). Using the within post-pre treatment difference, the non-linear least squares (NLS) estimator for  $\theta$  and  $\alpha$  is given by

$$(\hat{\theta}, \hat{\alpha}) = \arg \min \sum_{i=1}^N \sum_{k=1}^K [r'_{k,i} - r_{k,i} - \{h(Z_i, \theta, \alpha)\}]^2 \quad (14)$$

where  $h(Z_i, \theta, \alpha)$  (defined earlier) is a non-linear function of parameters. The utility function parameters to be estimated include  $[\rho_1, \psi_1, \rho_2, \psi_2]$ . We set  $\kappa_1 = 1/2$  as we found it difficult to separately identify the consumption share parameter from parameters governing the marginal utility of consumption. The ability function is parameterized as  $v(a) = \ln a$ .  $\beta$  is assumed to be 0.95 and  $T = 55$ . The combined parameters then consists of the taste for each major  $\gamma_1, \dots, \gamma_K$  and the post-graduation utility function parameters  $\theta$ . We allow for different utility function parameters for male and female students.<sup>15</sup>

### 7.3 Model Estimates

Table 7 provides the parameter estimates for two versions of the structural model. Model 1 is our main model. The marginal utility of own consumption (when single) is given by  $\phi_1 c_{S,1,t}^{-\rho_1}$ . We estimate  $\phi_1$  to be 0.23 for male students and 0.20 for female students, and the curvature parameter (relative risk aversion)  $\rho_1$  to be 4.43 for males and 5.20 for females. Both estimates are on the high end of previous estimates, but similar to the estimate in Nielsen and Vissing-Jorgensen (2006). The larger estimate of relative risk aversion for females is consistent with several studies that conclude that women are more risk averse than men in their choices (Eckel and Grossman, 2008; Croson and Gneezy, 2009). The high  $\rho$  estimates could be driven by the fact that our sample reports very high probabilities of completing a degree in humanities (Table 4), and humanities is one of the fields with the lowest uncertainty in earnings (columns (3) and (5) of Table 2). Own value of spouse's consumption has lower values of  $\phi_2$  and  $\rho_2$ . The coefficient on log ability rank is similar to the estimate in the reduced form of around 0.11 for both male and female students.

With the estimated parameters of the utility and ability functions, we can use the choice pre- and post- treatment choices to estimate each individual's taste for each major (relative to humanities/arts), given by  $\gamma_{k,i}$ . Table A3 provides statistics for the distribution of the estimated  $\gamma_{k,i}$  taste parameters (relative to humanities/arts which is normalized to 0). We see a distinct gender difference in tastes: On average, male students have a strong taste for

---

<sup>15</sup>In the estimation we also include a vector of revision fixed effects/intercepts that capture any mean differences in revisions by major. These revision fixed effects can be consistently estimated by estimating the mean revision for each major (relative to the reference major). The estimator (14) is then computed by de-meaning the  $h(\cdot)$  by these estimated revision fixed effects.



economics/business majors over humanities/arts (positive  $\gamma_{k,i}$ ), but average tastes for female students are negative for all majors, indicating a strong preference for humanities/arts over all other fields. Interesting, the median male taste for economics/business majors is negative and close to zero, indicating a skewed taste distribution. Figure 2 provides a direct look at the distribution of tastes for majors for men and women, respectively. Both distributions show some bimodality, but the most frequent mode for male students' tastes is near 0, whereas for the female students' tastes the mode is negative.

## 7.4 Using Cross-Sectional Data Only

We also estimated a second model using only the cross-sectional data and assuming a parametric distribution for college major tastes. The estimates of this model are intended to illustrate the "value added" of our panel data information experiment which allows us to flexibly estimate the distribution of unobserved tastes. For this restricted model, we assumed that the college major taste terms  $\gamma_k$  are distributed Type 1 extreme value with gender and major specific means. We estimated this model using only the pre-treatment data, thereby forming a cross-sectional dataset. This is essentially the same type of parametric taste restriction and data structure as Arcidiacono et al. (2011), although we use our life-cycle consumption utility specification and our data on own earnings and hours, marriage, and spousal earnings and hours. The estimates for this model are reported in the last column of Table 7. We obtain estimates that generally have larger degrees of relative risk aversion, and several times larger estimates for the ability component. The alternate model estimates also imply a lower marginal utility of own and spousal consumption.

## 7.5 Sample Fit

Next, we assess the fit of the estimated models, compared to the reported major choice probabilities in the data. Table 8 computes the predicted probabilities of major choice using the estimated parameters from each model. The unrestricted model fits the choice probabilities quite well, for both males and females, with only slight deviations between predicted model probabilities and those from the actual data.

The second model, using a parametric restriction on the taste distribution fits the choice probabilities substantially worse. There are large differences in predictions from this model relative to the actual data. For example, this model predicts only 23 percent of males choose the economics/business field, compared to the actual proportion of 38 percent. Given this low sample fit, we can soundly reject the parametric restriction on the taste distribution.

In Table 8 we also report estimates of a model in which we impose risk neutrality ( $\rho_1 = \rho_2 =$

0) on the unrestricted, panel data estimates. The risk neutral restricted model has considerably worse sample fit than the unrestricted risk aversion model. For example, the risk neutral model predicts 55 percent of males will choose the economics/business field, compared to the actual data probability of 38 percent or the predicted unrestricted model probability of 39 percent.

## 7.6 Choice Elasticities

We next interpret the estimated model (unrestricted Model 1) in terms of implied responsiveness of major choices to changes in self earnings. For each major, we increase beliefs regarding own earnings by 1 percent in every period. How much more likely would individuals be to major in each major due to this increase in earnings? We compute choice elasticities given by

$$\xi_{k,i} = \frac{\partial \pi_{k,i}}{\partial w_{FT,1,t}} \frac{w_{FT,1,t}}{\pi_{k,i}} \times 100.$$

Note that these choice elasticities depend on the estimated utility function parameters, and given the non-separability of tastes, abilities, and  $u(X, \theta)$ , also depend on the distribution of tastes and abilities.

Figure 3 graphs the distribution of the  $\xi_{k,i}$  choice elasticities in our samples of male and female students. Table 9 reports the mean of this distribution. A value of  $\xi_{k,i} = 0.1$  indicates that individual  $i$  would increase her probability of majoring in major  $k$  by 0.1 percent for a 1 percent increase in own earnings each period. From the figures it is clear that there is substantial heterogeneity in the responsiveness of individuals to changes in earnings. While some individuals would have a near zero response to the change in earnings, other individuals would have a substantial, albeit inelastic, response. The average response for male students is higher in most majors. The mean elasticity is considerably higher in the no grad field than in the other fields. This may be due to the relatively low beliefs about earnings in this major combined with the estimated concavity of the utility function with respect to consumption.

## 7.7 Decomposition of the Determinants of College Major Choices

Table 10 uses the estimated unrestricted model to decompose the college major choices into the constituent components. Our decomposition procedure starts by creating a baseline where every major choice is equally likely. We accomplish this by setting each respondent's beliefs about earnings, ability, hours of work, marriage, spousal characteristics (spousal earnings and hours), and tastes equal to the corresponding level for the humanities/arts major. Therefore, at the baseline, the odds of majoring in each of the remaining majors (relative to humanities/arts) is  $\pi_{k,i}/\pi_{\tilde{k},i} = 1$ . After establishing this baseline, we then progressively re-introduce each individual's major specific beliefs and tastes into the estimated choice model in order to

capture the marginal contribution of each component. The magnitude by which the relative odds of majoring in each field changes as we add a component measures the importance of this component. Table 10 reports the choice probability at each stage of the decomposition averaged over all of the sample respondents.

### 7.7.1 Male Students

In the first panel, we decompose major choices for male students only. Focusing on the first row, we see that re-introducing each individual's beliefs about his own earnings in each major increases the average odds of majoring in economics/business (relative to humanities/arts) from the baseline of 1 to 1.11, or a +0.11 marginal increase in odds. The increase in the average odds of majoring in economics/business reflects the earnings advantage most individuals perceive from graduating with an economics/business degree, evaluated at the estimated utility function parameters. In contrast, adding self beliefs about own earnings reduces the odds of not graduating from a baseline of 1 to 0.8124. Individuals are now less likely to believe they will not graduate given lower expected earnings from not graduating.

The remaining columns progressively add other model components, and the entries in Table 10 reflect the marginal gain of each component, given the other preceding components are included. Thus, adding beliefs about own ability in Column (2) only slightly reduce the odds of majoring in economics/business from 1.11 (including beliefs about own earnings) to about 1.10 (including both beliefs about own earnings and own ability). It is likely that the high positive correlation of beliefs about earnings and ability implies that marginal contribution of each is rather small. The marginal contribution of ability has the largest negative effect on majoring in engineering/computer science. The negative sign on the own ability components indicates that individuals perceive higher "study effort" due to either lower ability or greater difficulty in economics/business relative to humanities/arts, and thus this factor reduces the odds of majoring in economics/business.

Column (3) of Table 10 re-introduces beliefs about own work hours for each major. Because higher work hours increase total earnings (and there is no disutility from work), this tends to increase the odds of majoring in economics/business the most, and tends to reduce the odds of not graduating, given beliefs of higher unemployment spells with this major.

Column (4) adds spousal characteristics, including probability of marriage, spousal earnings, and spousal hours. The column indicates the marginal contribution of beliefs about gains in the marriage market from choosing different majors. These gains are positive and highest for economics/business but negative for not graduating.

Finally, Column (5) adds the remaining determinant of major choice, the vector of estimated major specific tastes. Tastes have a modest effect on choice to major in economics/business,

increasing the log odds by 0.0931. Tastes in this case then complement the other positive contributions to choosing the economics/business major, with the exception of ability. However, tastes have a large and negative effect on choosing the other majors. The negative sign on this component indicates that on average male students have high dis-taste for these majors (relative to humanities/arts). But the high negative taste is offset somewhat, with the exception of the not graduate category, by the positive contribution from own earnings and spousal characteristics.

### 7.7.2 Female Students

The second panel of Table 10 calculates the decomposition for female students. In comparing the male and female decompositions, it is clear that own earnings differences are a substantially smaller factor in college major choice for women than men. For ability, the reverse is true as ability differences across majors are a more important difference for women than men. For women, the negative component from ability, reflecting lower perceived ability in these majors relative to humanities/arts, (more than) offsets the positive earnings advantage. This was not true for men as the ability component, with the exception of engineering/computer science, is quite minor relative to the earnings component.

For the other components, own hours and spousal characteristics play relatively small marginal roles, with the exception of the not graduate category, where beliefs about poor spousal characteristics reduces the probability of not graduating. As with male choices, the taste component is large. This suggests that while the other determinants of college major choices, including earnings and ability, are meaningful, the taste component at the time of college major decision-making is dominant.

Column (4) shows that including spousal characteristics doesn't change the log-odds for graduating majors, but decreases the log-odds for the not graduate category. This suggests that returns in the marriage market are generated by simply going to college, and the college major itself does not matter much in this aspect.

### 7.7.3 Gender Ratio

The last panel of Table 10 directly assesses the contribution of the model components to the ratio of female to male major choices. Women are considerably more likely to major in humanities/arts than other majors: In our sample (before information treatment) the average female probability of majoring in humanities is 0.5, compared to 0.32 for men. The last panel of Table 10 calculates the relative odds for women versus men for each major (relative to humanities/arts):

$$\frac{\pi_{k,i}(\text{women})/\pi_{\tilde{k},i}^{\text{women}}}{\pi_{k,i}(\text{men})/\pi_{\tilde{k},i}^{\text{men}}}$$

In the pre-treatment sample, this ratio for economics/business is 0.46, reflecting that women are less likely to major in economics/business relative to humanities/arts than men.

As with the previous decomposition, we start with a baseline in which men and women are equally likely to choose all majors, and hence the female-male odds ratio is 1. In column (1) we see that adding beliefs about own earnings begins to create a gap between men’s and women’s college major choices. Adding earnings beliefs, reduces the economics/business female-male ratio from 1 to 0.95 (-0.05 marginal reduction). Similar negative reductions are evident for engineering/computer science and natural sciences. This increase in the gap between men and women occurs because men have generally higher earnings beliefs in these fields relative to humanities/arts than women (column (1) of Table 2). The exception is the not graduate category in which the female-male ratio actually increases to a female advantage from 1 at the baseline to 1.047 (+0.047 marginal gain).

In Column (2), we see that ability differences between men and women cause a further increase in the gender gap in major choice. Differences in beliefs about ability exacerbate the tendency for men to major in non-humanities subjects more than women. This is because men have higher ability beliefs in these beliefs relative to humanities/arts than women (column (1) of Table 6). On the other hand, gender differences in beliefs about own hours and spousal characteristics have only a minor effect on the gender gap. Finally, in Column (5), adding gender differences in major specific tastes substantially increases the gender gap. This finding suggests that pre-college determinants of tastes, as distinct in our framework from beliefs about earnings, ability, hours, and spousal characteristics, causes the majority of the gender difference in college major choices.

## 8 Conclusion

This paper seeks to shed light on the determinants of college major choice. While there is a recent and growing literature that uses subjective expectations data to understand schooling choices, our approach is unique in several ways. First, our survey has an innovative experimental feature embedded in it, which generates a panel of beliefs. This allows us to explore the extent to which students tend to be misinformed about the population distribution of earnings. We show that this experimental variation in beliefs can be used to identify the distribution of tastes non-parametrically. Second, we collect data on earnings uncertainty, which are usually not available in observational data. Third, instead of using indirect proxies, we provide the

first direct evidence of the role of marriage market returns on schooling choice. The fit of the model that excludes each of these additional dimensions (panel beliefs and non-parametric taste distribution, earnings uncertainty, marriage market returns) is substantially worse than that of our richer model, indicating that incorporating each of these dimensions is important.

We find that, in the context of major choice, earnings differences across majors is a more important factor for men than women, and ability differences matter more for women than men. However, tastes for majors are a dominant factor for both males and females. Even accounting for other characteristics such as earnings, labor supply, and ability, we find that females have a strong taste for humanities/arts while male students have a strong relative taste for economics/business. We also estimate substantial heterogeneity in tastes within gender, with the distribution of relative tastes estimated to be bimodal.

In our framework, "tastes" are defined at the point when students are in college. These could be tastes for major-specific outcomes realized in college, such as the enjoyability of coursework, or major-specific post-graduation outcomes, such as non-pecuniary aspects of jobs. It is important to note that tastes in our framework are distinct from ability and future earnings, though they may be correlated with them (which we do find to be the case). Differences in tastes may arise exogenously because of innate differences (Kimura, 1999; Baron-Cohen, 2003), or they may be endogenously determined by earlier interactions with peers and parents (Altonji and Blank, 1999). Understanding the originations of differences in tastes is not investigated in the current study, and is an important area of future research.

Despite our sample consisting of very high ability students enrolled at an elite university, we find that our survey respondents have biased beliefs about the distribution of earnings in the population. As shown in Table 2, the mean errors in population full-time earnings are substantial, varying from -17% (overestimation by 17%) to 28% (underestimation of population earnings by 28%), depending on the sub-group and the major. In Sections 5 and 6, we show that students sensibly revise their self beliefs as well as their probabilistic choices in response to this information. This suggests that information campaigns focused on providing accurate information on returns to schooling could have a large impact on beliefs and choices of students. While such campaigns have been conducted in developing countries (Jensen, 2010; Nguyen, 2010), our results make a case for such interventions in developed countries as well.<sup>16</sup>

A possible alternative to our quasi-experimental approach is the methodology used in Blass et al. (2010), who estimate preferences for electricity reliability by asking survey participants to value various bundles of electricity generation bills and outage probabilities. The shortcoming of their counterfactual scenarios approach is that it may be difficult to operationalize meaningful

---

<sup>16</sup>One study that we are aware of in a developed setting is that of Bettinger et al. (2011) who find that providing information on financial aid and assistance in filling out federal financial aid forms improves college access.

counterfactual scenarios for some applications of interest- it is not clear how one would pose simple counterfactual situations in complicated occupational choice contexts, such as college major choices.

How students revise their beliefs and choices in an experimental framework like ours where the information is presented to the respondent may be very different from the change in their behavior where they acquire the information themselves (Hertwig et al., 2004). While it is challenging to identify changes in information sets in actual panels (Zafar, 2011), an important question for future research is to explore how students' beliefs and choices evolve over longer time horizons, and how persistent the impact of revealed information is on students' behavior.

## References

- [1] Altonji, Joseph, and Rebecca Blank. 1999. "Race and Gender in the Labor Market." In *Handbook of Labor Economics*, Volume 3c, ed. Orley Ashenfelter and David Card, 3144-3259. Elsevier Science.
- [2] Arcidiacono, Peter. 2004. "Ability Sorting and the Returns to College Major." *Journal of Econometrics*, 121(1-2): 343-375.
- [3] Arcidiacono, Peter, Joseph Hotz, and Songman Kang. 2011. "Modeling College Major Choices using Elicited Measures of Expectations and Counterfactuals." *Journal of Econometrics*, forthcoming.
- [4] Attanasio, Orazio, and Katja Kaufmann. 2011. "Education Choices and Returns on the Labour and Marriage Markets: Subjective Expectations, Gender and Decision Making." Working Paper.
- [5] Bamberger, Gustavo. 1986. "Occupation Choice: The Role of Undergraduate Education." Ph.D. Dissertation, University of Chicago.
- [6] Baron-Cohen, Simon. 2003. *The Essential Difference: Men, Women, and the Extreme Male Brain*. London: Allen Lane.
- [7] Beffy, Magali, Denis Fougere, and Arnaud Maurel. 2011. "Choosing the Field of Study in Post-Secondary Education: Do Expected Earnings Matter?" *The Review of Economics and Statistics*, Forthcoming.
- [8] Berger, Mark. 1988. "Predicted Future Earnings and Choice of College Major." *Industrial and Labor Relations Review*, 41(3): 418-29.
- [9] Bettinger, Eric, Bridget Long, Philip Oreopoulos, and Lisa Sanbonmatsu. 2011. "The Role of Simplification and Information in College Decisions: Results from the H&R Block FAFSA Experiment" Working Paper.
- [10] Betts, Julian. 1996. "What do Students Know about Wages? Evidence from a Survey of Undergraduates." *Journal of Human Resources*, 31(1): 27-56.

- [11] Blass, Asher, Saul Lach, and Charles Manski. 2010. "Using Elicited Choice Probabilities to Estimate Random Utility Models: Preferences for Electricity Reliability." *International Economic Review*, 51(2): 421-440.
- [12] Brown, Charles, and Mary Corcoran. 1997. "Sex Based Differences in School Content and the Male-Female Wage Gap." *Quarterly Journal of Economics*, 99: 31-44.
- [13] Chiappori, Pierre-Andre, Murat Iyigun, and Yoram Weiss. 2009. "Investment in Schooling and the Marriage Market." *American Economic Review*, 99(5): 1689-1713.
- [14] Croson, Rachel, and Uri Gneezy. 2009. "Gender Differences in Preferences." *Journal of Economic Literature*, 47(2): 448-74.
- [15] Dey, Judy, and Catherine Hill. 2007. "Behind the Pay Gap." American Association of University Women Educational Foundation Research Report.
- [16] Delavande, Adeline. 2008. "Pill, Patch or Shot? Subjective Expectations and Birth Control Choice." *International Economic Review*, 49(3): 999-1042.
- [17] Delavande, Adeline. 2011. "HIV/AIDS-related Expectations and Risky Sexual Behavior in Malawi." Working Paper.
- [18] Eckel, Catherine, and Philip Grossman. 2008. "Differences in the Economic Decisions of Men and Women: Experimental Evidence." In *Handbook of Experimental Economics Results*, ed. Charles Plott and Vernon Smith, 509-519. New York: Elsevier.
- [19] Eide, Eric., and Geetha Waehrer. 1998. "The Role of the Option Value of College Attendance in College Major Choice." *Economics of Education Review*, 17(1): 73-82.
- [20] Flyer, Fredrick. 1997. "The Influence of Higher Moments of Earnings Distributions on Career Decisions." *Journal of Labor Economics*, 15(4): 689-713.
- [21] Freeman, Richard. 1971. *The Market for College-Trained Manpower: Study in the Economics of Career Choice*. Cambridge: Harvard University Press.
- [22] Freeman, Richard. 1975. "Supply and Salary Adjustments to the Changing Science Manpower Market: Physics, 1948-1975." *American Economic Review*, 65(1): 27-39.
- [23] Ge, Suqin. 2010. "Women's College Decision: How Much Does Marriage Matter?" *Journal of Labor Economics*, forthcoming.
- [24] Gemici, Ahu, and Matthew Wiswall. 2011. "Evolution of Gender Differences in Post-Secondary Human Capital Investments: College Majors at the Intensive Margin." Working Paper.
- [25] Grogger, Jeff, and Eric Eide. 1994. "Changes in College Skills and the Rise in the College Wage Premium." *Journal of Human Resources*, 30(2): 280-310.
- [26] Giustinelli, Pamela. 2010. "Uncertain Outcomes and Child-Parent Decision Making in Curriculum Choice: What Data Do We Need to Tell Them Apart?" Working Paper.



- [27] Hertwig, Ralph, Greg Barron, Elke Weber, and Ido Erev. 2004. "Decisions from Experience and the Weighting of Rare Events." *Psychological Science*, 15 (8): 534-539.
- [28] Iyigun, Murat, and Randall Walsh. 2007. "Building the Family Nest: Premarital Investments, Marriage Markets, and Spousal Allocations." *Review of Economic Studies*, 74(2): 507-535.
- [29] Jensen, Robert. 2010. "The (Perceived) Returns to Education and the Demand for Schooling." *Quarterly Journal of Economics*, 125(2): 515-548.
- [30] Katz, Lawrence, and Kevin Murphy. 1992. "Changes in Relative Wages, 1963-87: Supply and Demand Factors." *Quarterly Journal of Economics*, 107(1): 35-78.
- [31] Kaufmann, Katja. 2009. "Understanding the Income Gradient in College Attendance in Mexico: The Role of Heterogeneity in Expected Returns to College." Working Paper.
- [32] Kimura, Doreen. 1999. *Sex and Cognition*. Cambridge: MIT Press.
- [33] Lafortune, Jeanne. 2010. "Making Yourself Attractive: Pre-Marital Investments and the Returns to Education in the Marriage Market." Working Paper.
- [34] Lochner Lance. 2007. "Individual Perceptions of the Criminal Justice System." *American Economic Review*, 97(1): 444-460.
- [35] Manski, Charles. 1993. "Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?" in *Studies of Supply and Demand in Higher Education*, ed. Charles Clotfelter and Michael Rothschild, 43-57. Chicago: University of Chicago Press.
- [36] Manski, Charles. 2002. "Identification of Decision Rules in Experiments on Simple Games of Proposal and Response." *European Economic Review*, 46(4-5): 880-891.
- [37] Manski, Charles. 2004. "Measuring Expectations." *Econometrica*, 72(5): 1329-1376.
- [38] Mare, Robert. 1991. "Five Decades of Educational Assortative Mating." *American Sociological Review*, 56(1): 15-32.
- [39] Montmarquette, Claude, Kathy Cannings, and Sophie Mahseredjian. 2002 "How Do Young People Choose College Majors?" *Economics of Education Review*, 21(6): 543-556.
- [40] Nielsen, Helena, and Annette Vissing-Jorgensen. 2006. "The Impact of Labor Income Risk on Educational Choices: Estimates and Implied Risk Aversion." Working Paper.
- [41] Nguyen, Trang. 2010. "Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar," MIT Working Paper.
- [42] Niederle, Muriel, and Lise Vesterlund. 2007. "Do Women Shy away from Competition? Do Men Compete too Much?" *Quarterly Journal of Economics*, 122(3): 1067-1101.
- [43] Paglin, Morton, and Anthony Rufolo. 1990. "Heterogeneous Human Capital, Occupational Choice, and Male-Female Earnings Differences." *Journal of Labor Economics*, 8: 123-144.

- [44] Paula, Aureo de, Gil Shapira and Petra Todd. 2011. "Behavioral Determinants of HIV/AIDS: A Dynamic Model of Marriage and Cheating in Malawi." PIER Working Paper 11-005.
- [45] Pencavel, John. 1998. "Assortative Mating by Schooling and the Work Behavior of Wives and Husbands." *American Economic Review*, 88(2): 326-329.
- [46] Saks, Raven, and Steven Shore. 2005. "Risk and Career Choice." *Advances in Economic Analysis & Policy*, 5(1).
- [47] Siow, Aloysius. 1984. "Occupational Choice under Uncertainty." *Quarterly Journal of Economics*, 52(3), 631-45.
- [48] Stinebrickner, Todd, and Ralph Stinebrickner. 2010. "Learning about Academic Ability and the College Drop-out Decision." University of Western Ontario Working Paper 20086.
- [49] Stinebrickner, Todd, and Ralph Stinebrickner. 2011. "Math or Science? Using Longitudinal Expectations Data to Examine the Process of Choosing a College Major." Working Paper.
- [50] Turner, Sarah, and William Bowen. 1999. "Choice of Major: The Changing (Unchanging) Gender Gap." *Industrial and Labor Relations Review*, 52(2): 289-313.
- [51] van der Klaauw, Wilbert, and Kenneth Wolpin. 2008. "Social Security and Savings and Retirement Behavior of Low Income Households." *Journal of Econometrics*, 145(1-2): 21-42.
- [52] van der Klaauw, Wilbert. 2011. "On the Use of Expectations Data in Estimating Structural Dynamic Choice Models." Working Paper.
- [53] Weinberger, Catherine. 1998. "Race and Gender Wage Gaps in the Market for Recent College Graduates." *Industrial Relations*, 37(1): 67-84.
- [54] Weinberger, Catherine. 2004. "Just Ask! Why Surveyed Women Did Not Pursue Information Technology Courses or Careers?" *IEEE Technology and Society*, 23(2): 28-35.
- [55] Wiswall, Matthew. 2006. "Skill Composition and the Returns to Post-Secondary Education in the United States." Working Paper.
- [56] Zarkin, Gary. 1985. "Occupational Choice: An Application to the Market for Public School Teachers." *Quarterly Journal of Economics*, 100(2), 409-446.
- [57] Zafar, Basit. 2009. "College Major Choice and the Gender Gap." Staff Report No. 364, Federal Reserve Bank of New York.
- [58] Zafar, Basit. 2011. "How do College Students Form Expectations." *Journal of Labor Economics*, 29(2): 301-348.

Figure 1: Distribution of Changes in Log Expected Graduation Probabilities (Relative to Humanities/Arts)

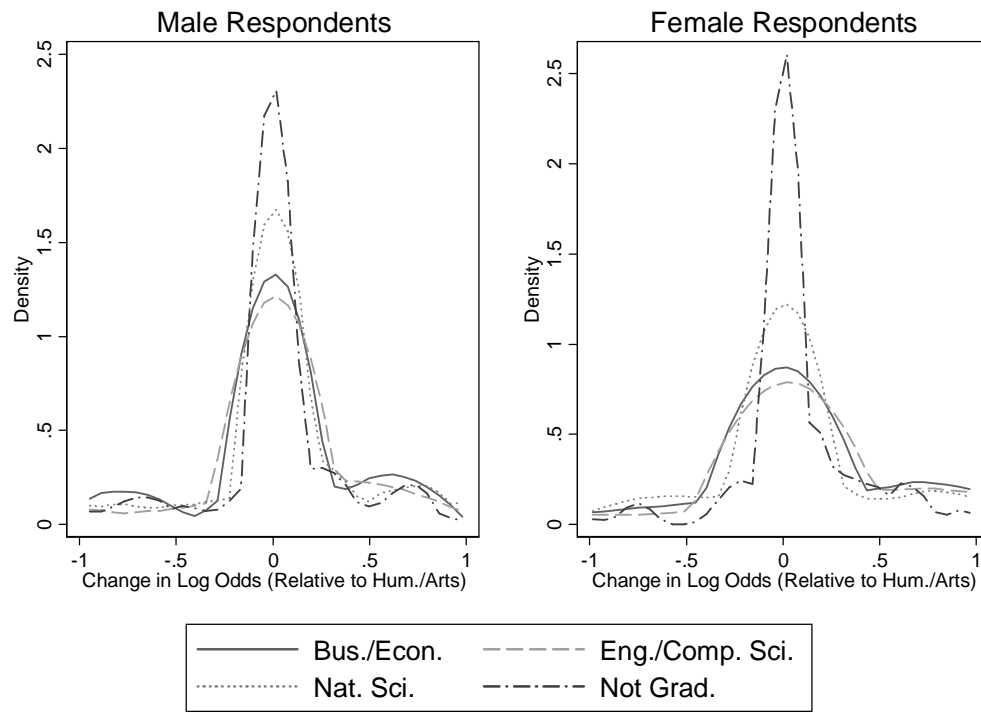


Figure 2: Distribution of Individual Fixed Taste (Rel to Hum./Arts) Component  $\gamma_{ik}$

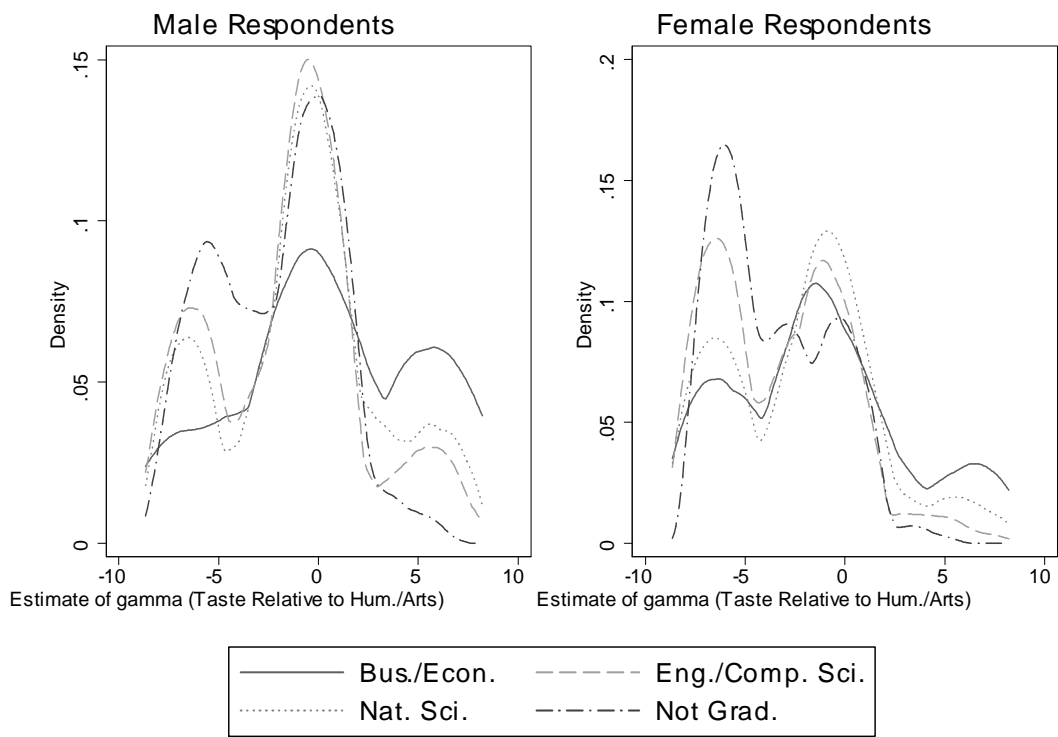


Figure 3: Distribution of Choice Elasticities

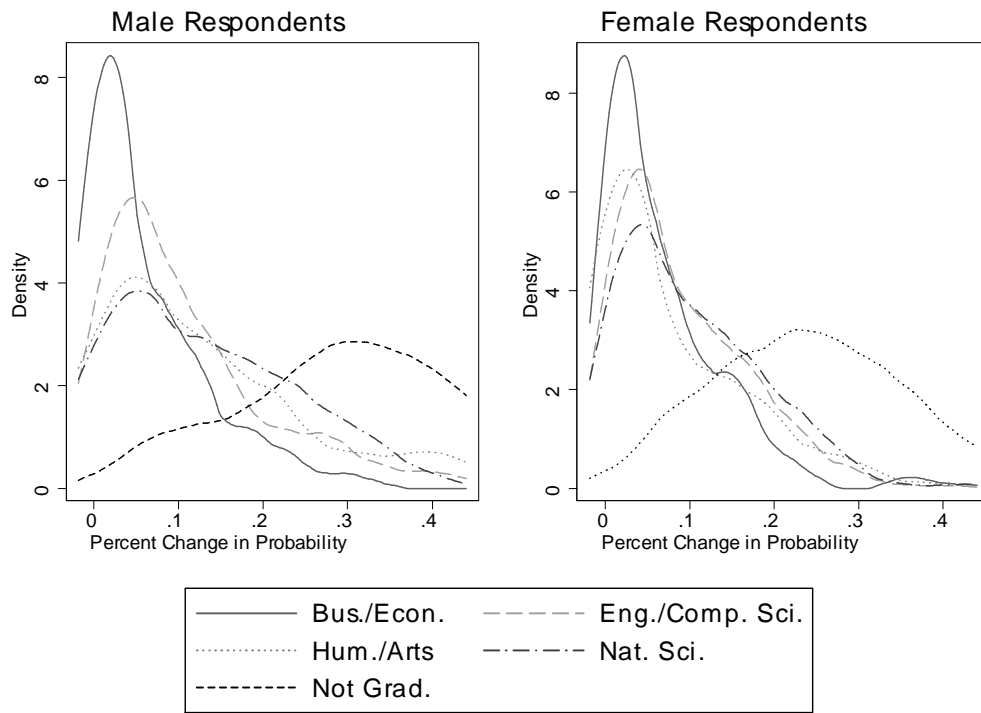


Table 1: Summary Statistics: Mean and Standard Deviation of Elicited Beliefs

	(1)	(2)	(3)	(4)	(5)	(6)
	Belief about Women	Percent Error $(\frac{\text{Truth} - \text{Belief} * 100}{\text{Truth}})$	Belief about Men	Percent Error $(\frac{\text{Truth} - \text{Belief} * 100}{\text{Truth}})$	Belief Abt Wage Gender Gap <sup>a</sup>	True Wage Gender Gap
<b>Sample: Male Students</b>						
Economics/Business	mean (std.)	-8.59 (40.73)	6.95 (2.94)	6.74 (39.41)	-4.87 (6.06)	-18.53
Engineering/Comp. Sci.	mean (std.)	14.33 (62.37)	6.31 (2.58)	23.41 (31.32)	-4.26 (6.54)	-8.85
Humanities/Arts	mean (std.)	1.57 (37.2)	4.86 (1.67)	8.11 (31.62)	-3.23 (9.06)	-7.15
Natural Sciences	mean (std.)	5.74 (48.81)	5.99 (3.1)	17.44 (42.66)	-3.68 (6.51)	-17.31
Not Graduate	mean (std.)	13.13 (34.52)	3.38 (1.3)	29.27 (27.13)	-7.41 (11.66)	-27.61
<b>Sample: Female Students</b>						
Economics/Business	mean (std.)	-17.14 (45.21)	8.07 (4.38)	-8.31 (58.77)	-6.07 (11.33)	-18.53
Engineering/Comp. Sci.	mean (std.)	13.00 (33.57)	7.64 (5.14)	7.31 (62.43)	-6.30 (12.53)	-8.85
Humanities/Arts	mean (std.)	-3.03 (35.42)	5.63 (2.85)	-6.40 (53.85)	-4.84 (10.16)	-7.15
Natural Sciences	mean (std.)	-1.48 (65.21)	6.74 (3.57)	7.15 (49.16)	-5.96 (9.43)	-17.31
Not Graduate	mean (std.)	8.16 (60.03)	3.46 (1.39)	27.52 (28.99)	-0.04 (82.15)	-27.61

Notes: Beliefs (columns 1, & 5) are in \$10,000's. Other columns are percentages. Sample for columns 1 & 2 (3 & 4) includes respondents who received the female (male), individual, or college treatments in the intermediate stage; Sample for columns 5 & 6 includes respondents who received either individual or college treatments.

<sup>a</sup> Wage gap is defined as 100\* (female population earnings-male population earnings)/male population earnings.

Table 2: Earnings and Earnings Revisions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Self Earnings Pre	Self % Revision $(\frac{\text{Post-Pre} * 100}{\text{Pre}})$	Self S.d. Pre <sup>a</sup>	Self S.d. (not in major <sup>b</sup> )	Self S.d. Post	Spouse's Earnings Pre <sup>c</sup>	Spouse % Revision $(\frac{\text{Post-Pre} * 100}{\text{Pre}})$
<b>Sample: Male Students</b>							
Economics/Business	mean (std.)	-47.35 (29.00)	8.76 (18.65)	5.95** (12.69)	7.15 (15.67)	7.66 (4.573)	31.36 (57.76)
Engineering/Computer Sci	mean (std.)	-0.341 (25.04)	7.85 (18.15)	8.2 (18.69)	5.86 (13.24)	7.14 (4.51)	4.22 (36.92)
Humanities/Arts	mean (std.)	7.79 (3.36)	5.83 (14.94)	6.73 (18.45)	5.32 (12.71)	5.99 (4.41)	3.05 (30.61)
Natural Sciences	mean (std.)	7.31 (35.60)	7.18 (17.42)	7.32 (16.76)	6.29 (15)	7.13 (5.87)	10.57 (90.32)
Not Graduate	mean (std.)	30.65 (61.00)	14.43 (31.2)	14.43 (31.2)	11.43 (26.24)	4.31 (4.66)	14.94 (69.65)
<b>Sample: Female Students</b>							
Economics/Business	mean (std.)	-7.03 (39.63)	7.34 (12.91)	6.87 (11.76)	6.57 (12.87)	9.17 (4.34)	-0.965 (58.79)
Engineering/Computer Sci	mean (std.)	-0.116 (28.26)	7.78 (14.85)	7.77 (14.98)	7.01 (14.76)	8.04 (2.90)	9.15 (64.05)
Humanities/Arts	mean (std.)	4.77 (31.11)	6.29 (12.73)	5.97 (9.63)	5.93 (14.68)	6.48 (2.39)	10.27 (79.87)
Natural Sciences	mean (std.)	5.21 (44.94)	7.7 (15.79)	8.18 (16.78)	7.4 (16.66)	7.71 (4.07)	11.38 (92.28)
Not Graduate	mean (std.)	37.19 (62.79)	11.69 (25.69)	11.69 (25.69)	12.13 (26.82)	4.94 (5.62)	34.19 (130.32)

Notes: Earnings and S. d. (standard deviation) of earnings are in \$10,000's.

<sup>a</sup> Self S.d. pre is the standard deviation obtained by fitting the respondent's responses about mean earnings, percent chance of earnings exceeding \$35,000 and \$85,000 to a log-normal distribution.

<sup>b</sup> Out of major are respondents for whom the field is NOT their stated most likely field of graduation. \*\* Differences in s.d. of in-major and out-major respondents are statistically significant at the 5% level (pairwise ttest).

<sup>c</sup> Spouse's earnings pre are beliefs about expected earnings of the student's spouse, conditional on the student's own major (not the spouse's major).

Table 3: Population and Self Beliefs

	(1)	(2)	(3)	(4)	(5)
Dependent Variable:	Log Self Earnings	Log Self Earnings	Log Self Earnings	Log Earnings Revision (Post-Pre)	Log Earnings Revision (Post-Pre)
Log Population Earnings Beliefs	0.961 (0.0437)	0.876 (0.0641)	0.904 (0.0622)		
Log Population Earnings Errors log(Truth/Belief)				0.196 (0.0455)	0.161 (0.0423)
Individual Covariates?	NO	NO	YES	–	–
Major Dummies?	NO	YES	YES	NO	YES

Notes: Individual covariates include an indicator for gender; indicators for Asian, Hispanic, black, or other race (white race is omitted category), overall grade point average (GPA); scores on the verbal and mathematics SAT; indicators for whether the student’s mother and father attended college; parents’ income; and indicators for non-reported (missing) SAT scores, GPA, parental education or parental income. Major dummies include indicators for the remaining majors: economics/business, engineering/computer sci, natural science, and no graduation.



Table 4: Expected Probability of Completing a Degree in Specific Majors

		Before <sup>a</sup>	Before (Rel. Hum./Arts) <sup>b</sup>	Revisions Post-Pre Treat.	Log Odds Rev. (Rel. Hum./Arts) <sup>c</sup>
<b>Male Students</b>					
Econ./Bus.	mean	0.378	0.0547	-.00251	0.276
	(std.)	(0.381)	(0.679)	(0.115)	(1.69)
Eng./Comp.Sci.	mean	0.0940	-.230	0.0167	0.496
	(std.)	(0.171)	(0.452)	(0.0873)	(2.00)
Hum./Arts	mean	0.324	–	-.0289	–
	(std.)	(0.373)	–	(0.109)	–
Nat. Sci.	mean	0.179	-.145	0.0182	0.347
	(std.)	(0.279)	(0.535)	(0.110)	(1.89)
Not Graduate	mean	0.0275	-.296	-.00374	0.127
	(std.)	(0.0650)	(0.377)	(0.0627)	(1.93)
<b>Female Students</b>					
Econ./Bus.	mean	0.268	-.235	0.0164	0.528
	(std.)	(0.348)	(0.676)	(0.0960)	(1.75)
Eng./Comp.Sci.	mean	0.0529	-.450	0.0212	0.659
	(std.)	(0.127)	(0.442)	(0.0683)	(1.95)
Hum./Arts	mean	0.503	–	-.0421	–
	(std.)	(0.389)	–	(0.126)	–
Nat. Sci.	mean	0.159	-.344	0.00303	0.340
	(std.)	(0.261)	(0.553)	(0.0965)	(1.74)
Not Graduate	mean	0.0184	-.485	0.00133	0.0944
	(std.)	(0.0537)	(0.396)	(0.0317)	(1.77)

Notes: This table reports the mean self belief about completing each of the majors. Probabilities are reported on a 0 - 100 scale, and then normalized to 0 - 1. The standard deviation is in parentheses.

<sup>a</sup> Reported before receiving info treatments.

<sup>b</sup> Probability in major - Probability in Humanities.

<sup>c</sup>  $\text{Log}(\text{Post Probability in major} / \text{Post Probability in Humanities}) - \text{Log}(\text{Pre Probability in major} / \text{Pre Probability in Humanities})$ .

Table 5: Graduation Expectations and Expected Earnings

Dep. Variable:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
		Log Odds of Major Rel. to Hum.	Log Odds of Major Rel. to Hum.	Log Odds Revision (Post-Pre)	Log Odds of Major Rel. to Hum.	Log Odds of Major Rel. to Hum.	Log Odds Revision (Post-Pre)	Log Odds Revision (Post-Pre)	Log Odds of Major Rel. to Hum.	Log Odds of Major Rel. to Hum.	Log Odds Revision (Post-Pre)	Log Odds Revision (Post-Pre)
Log Self Earnings (Rel to HumArts)	1.98 (0.161)	2.07 (0.242)	2.15 (0.227)	0.279 (0.152)	1.62 (0.354)	1.01 (0.339)	0.270 (0.167)	0.243 (0.152)	0.664 (0.0594)	1.14 (0.331)	0.115 (0.0473)	0.206 (0.154)
Log Self Earnings Rev (Post - Pre)												
Log SPOUSE Earn (Rel to Hum) × Male					1.15 (0.247)	0.395 (0.251)	0.201 (0.133)	0.223 (0.254)				
Log SPOUSE Earn Revision (Post-Pre) × Male							0.270 (0.167)	0.228 (0.166)				0.249 (0.166)
Log SPOUSE Earn Revision (Post-Pre) × Female							0.201 (0.133)	0.187 (0.132)				0.185 (0.134)
Log Ability Rank (Rel. to Hum.)									0.664 (0.0594)	0.594 (0.0600)		
Log Ability Rank Rev. (Post - Pre)											0.115 (0.0473)	0.111 (0.0476)
Indiv Covariates?	NO	NO	YES	–	YES	YES	–	–	YES	YES	–	–
Major Dummies?	NO	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Total Observ.:	1436	1436	1436	1436	1436	1436	1436	1436	1436	1436	1436	1436
Individuals:	359	359	359	359	359	359	359	359	359	359	359	359

Notes: Self Expected Earnings is expected earnings at age 30 if the student were to graduate in one of five potential majors. Heteroskedastic cluster robust standard error in parentheses. Standard errors are adjusted for clustering at the individual level for the models which include individual covariates. All other explanatory variables vary at the individual and major level. Individual covariates are the same as in Table 3.

Table 6: Ability Rank

		Self Ability Before	Ability Revision (Post - Pre)
<b>Sample: Male Students</b>			
Economics/Business	mean	64.96	7.19
	(std.)	(28.59)	(20.28)
Engineering/Computer Science	mean	53.82	12.50
	(std.)	(29.77)	(25.21)
Humanities/Arts	mean	67.98	6.64
	(std.)	(29.62)	(23.67)
Natural Sciences	mean	64.81	3.65
	(std.)	(27.65)	(21.58)
Not Graduate	mean	69.99	1.15
	(std.)	(38.31)	(35.30)
<b>Sample: Female Students</b>			
Economics/Business	mean	59.18	6.28
	(std.)	(26.98)	(23.65)
Engineering/Computer Science	mean	45.63	11.10
	(std.)	(28.95)	(28.81)
Humanities/Arts	mean	73.81	-.191
	(std.)	(24.33)	(22.39)
Natural Sciences	mean	56.81	5.82
	(std.)	(28.64)	(24.51)
Not Graduate	mean	55.01	13.70
	(std.)	(43.36)	(43.33)

Notes: Ability ranking is measured on a 100 point scale, with 100 being top rank and 1 lowest rank.

Table 7: Structural Model Post-Graduation Parameter Estimates

	<i>Model 1</i>		<i>Model 2</i>	
	(Panel Data)		(Cross-Sectional Data Only)	
	Males	Females	Males	Females
Own Utility				
$\phi_1$	0.2333 (0.0613)	0.1985 (0.0172)	0.0828 (0.0003)	0.1170 (0.0008)
$\rho_1$	4.4379 (1.0694)	5.1919 (1.0246)	6.4819 (0.0033)	5.4128 (0.0096)
Spouse Utility				
$\phi_2$	0.3435 (0.2774)	0.3146 (0.0395)	0.0813 (0.0008)	0.1868 (0.0091)
$\rho_2$	3.8003 (0.8480)	4.0965 (1.0593)	1.2190 (0.0381)	1.5482 (0.1542)
Ability $\alpha$	0.1113 (0.0699)	0.1053 (0.0421)	0.6221 (0.0667)	0.4301 (0.0360)

Notes: Bootstrapped standard errors in parentheses calculated from 50 bootstrap repetitions.

Table 8: Sample Fit

	Data	Model 1	Model 2	Model 3
		(Unrestricted)	(Cross-Sectional Data Only)	(Risk Neutral: $\rho_1 = \rho_2 = 0$ )
<b>Male Students Prob. of Majoring in...</b>				
Economics/Business	0.3782	0.3861	0.2305	0.5540
Engineering/Comp. Sci.	0.0940	0.0953	0.0332	0.2166
Humanities/Arts	0.3235	0.3130	0.6123	0.0984
Natural Sciences	0.1787	0.1877	0.1090	0.1216
Not Graduate	0.0275	0.0179	0.0149	0.0094
<b>Female Students Prob. of Majoring in...</b>				
Economics/Business	0.2684	0.2771	0.5363	0.5717
Engineering/Comp. Sci.	0.0529	0.0583	0.0661	0.1498
Humanities/Arts	0.5031	0.4908	0.2550	0.1215
Natural Sciences	0.1591	0.1588	0.1293	0.1430
Not Graduate	0.0184	0.0150	0.0133	0.0141

Table 9: Own Earnings Choice Elasticities: Average Percent Change in Probability of Graduating in Each Major with a 1% Increase in Own Earnings in that Major

	Male Students	Female Students
% $\Delta$ Prob Bus/Econ	0.0728	0.0459
% $\Delta$ Prob Eng/Comp. Sci.	0.1121	0.0702
% $\Delta$ Prob Hum./Arts	0.1531	0.0733
% $\Delta$ Prob Nat. Sci.	0.1589	0.0811
% $\Delta$ Prob No Grad.	0.3475	0.2272

Table 10: Decomposition of the Determinants of College Major Choices

	(1)	(2)	(3)	(4)	(5)	
	Change in Odds Relative to Humanities/Arts					
	Baseline Equal Odds	Add Own Earnings	Add Own Ability	Add Own Hours	Add Spousal Charact.	Add Own Tastes
<b>Male Students</b>						
Econ./Bus.	1.0000	0.1079	-0.0046	0.0197	0.0174	0.0931
Eng./Comp. Sci.	1.0000	0.0982	-0.0415	0.0079	0.0127	-0.7729
Nat. Sci.	1.0000	0.0505	-0.0039	0.0024	0.0044	-0.4537
Not Grad.	1.0000	-0.1795	-0.0260	-0.0251	-0.0848	-0.6274
<b>Female Students</b>						
Econ./Bus.	1.0000	0.0536	-0.0366	0.0082	0.0128	-0.4732
Eng./Comp. Sci.	1.0000	0.0423	-0.0816	0.0080	0.0137	-0.8637
Nat. Sci.	1.0000	0.0253	-0.0443	0.0046	0.0076	-0.6696
Not Grad.	1.0000	-0.1412	-0.0958	-0.0142	-0.0652	-0.6532
<b>Female/Male Ratio</b>						
Econ./Bus.	1.0000	-0.0490	-0.0292	-0.0089	-0.0027	-0.4524
Eng./Comp. Sci.	1.0000	-0.0509	-0.0399	0.0008	0.0020	-0.5218
Nat. Sci.	1.0000	-0.0239	-0.0387	0.0022	0.0033	-0.4033
Not Grad.	1.0000	0.0467	-0.0864	0.0129	0.0254	-0.4651

# A Appendix

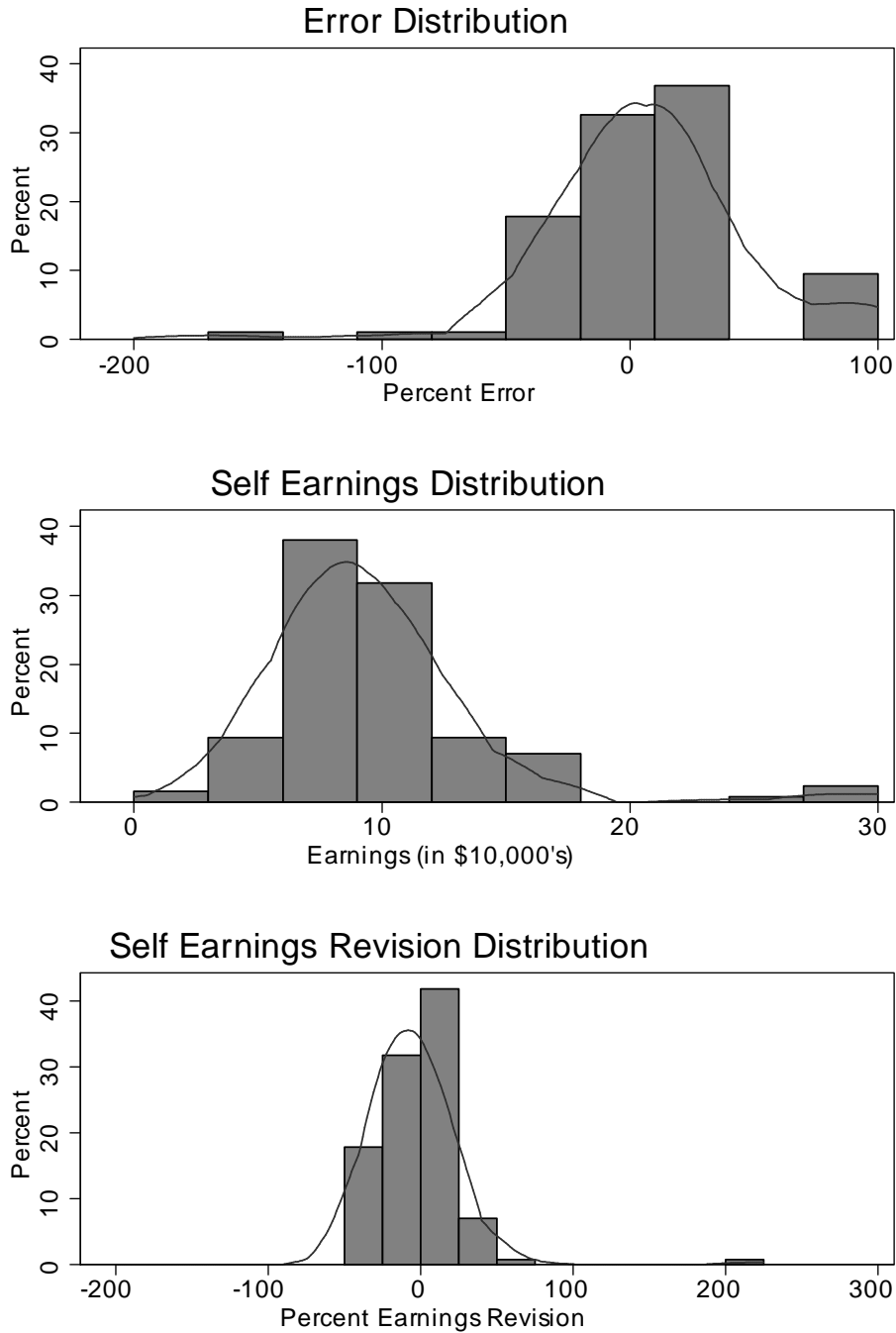


Figure A1: Male Expectations of Male Econ/Business Earnings

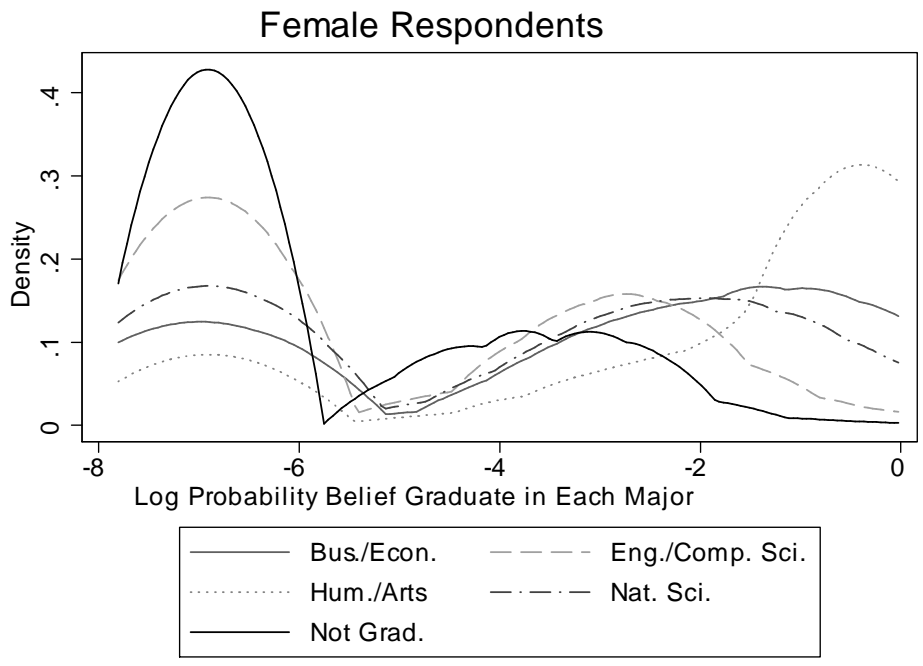
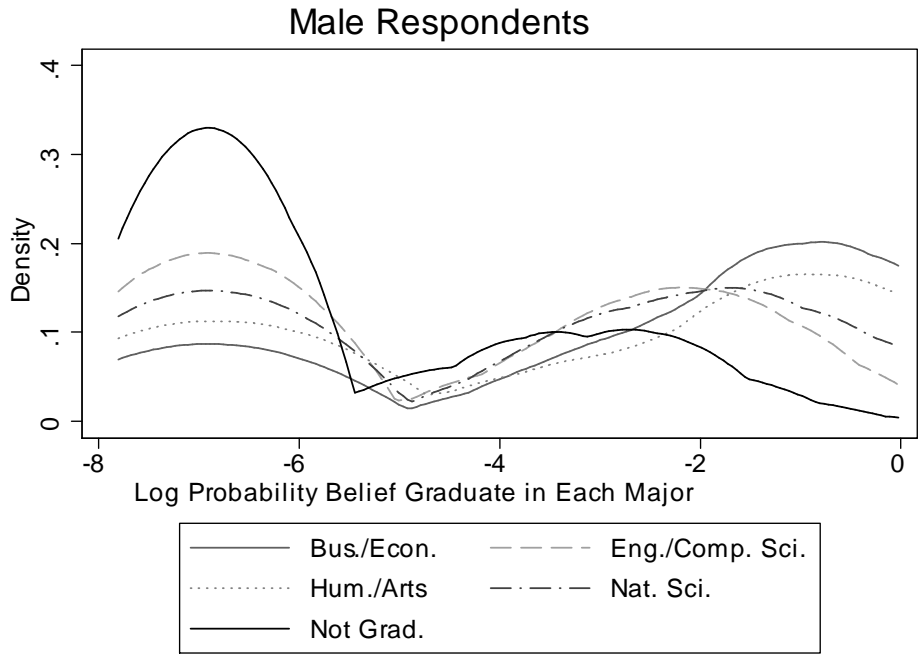


Figure A2: Distribution of Expected (Log) Graduation Probabilities

Table A1: Information revealed in the Treatments

**Individuals Treatment**

The following information is from the US Census Bureau.	
Among all individuals (including college and non-college graduates) aged 30:	
The percentage that are working full time is	59.80%
The percentage of those that are working full time who are women is	42.70%
The average annual earnings of those that are working full time is	\$45,726
The percentage of those that are working full time that earn more than \$35,000 per year is	59.00%
The percentage of those that are working full time that earn more than \$85,000 per year is	7.30%

**College Treatment**

The following information is from the US Census Bureau.	
Among all college graduates currently aged 30:	
The percentage that are working full time is	69.80%
The percentage of those that are working full time who are women is	52.80%
The average annual earnings of those that are working full time is	\$60,376
The percentage of those that are working full time that earn more than \$35,000 per year is	80.70%
The percentage of those that are working full time that earn more than \$85,000 per year is	14.80%

**Female Treatment**

The following information is from the US Census Bureau.					
Among all female college graduates aged 30 who received a Bachelor's degree in major (M):					
The percentage that are working full time is	Econ	Eng	Hum	Nat	No Grad
The average annual earnings of those that are working full time is	60.6%	72.8%	52.3%	55.3%	51.6%
The percentage of those that are working full time that earn more than \$35,000 per year is	\$60,730	\$75,086	\$49,154	\$60,021	\$34,603
The percentage of those that are working full time that earn more than \$85,000 per year is	85.5%	99.0%	72.2%	84.0%	44.9%
	27.5%	26.9%	8.0%	8.5%	1.6%

**Male Treatment**

The following information is from the US Census Bureau.					
Among all male college graduates aged 30 who received a Bachelor's degree in major (M):					
The percentage that are working full time is	Econ	Eng	Hum	Nat	No Grad
The average annual earnings of those that are working full time is	93.5%	91.6%	77.6%	81.9%	72.1%
The percentage of those that are working full time that earn more than \$35,000 per year is	\$74,542	\$82,377	\$52,937	\$72,583	\$47,803
The percentage of those that are working full time that earn more than \$85,000 per year is	92.4%	95.2%	78.8%	90.6%	65.2%
	31.5%	33.6%	8.7%	24.2%	5.7%

**Also Revealed to All Respondents in Final Stage**

The percentage of those who are women is	34.70%	18.20%	55.20%	48.00%	42.30%
--	--------	--------	--------	--------	--------



Table A2: Correlation in Self Earnings Across College Majors

Panel A: Male Students					
	Econ/Bus	Eng/Comp.	Hum./Arts	Nat Sci.	No Grad.
Econ/Bus	1.00				
Eng/Comp.	0.794	1.00			
Hum./Arts	0.374	0.366	1.00		
Nat Sci.	0.540	0.591	0.778	1.00	
Not Grad.	0.662	0.797	0.719	0.799	1.00
Panel B: Female Students					
	Econ/Bus	Eng/Comp.	Hum./Arts	Nat Sci.	No Grad.
Econ/Bus	1.00				
Eng/Comp.	0.602	1.00			
Hum./Arts	0.446	0.431	1.00		
Nat Sci.	0.483	0.546	0.431	1.00	
Not Grad.	0.186	0.206	0.360	0.0445	1.00

Table A3: Distribution of Estimated Taste Parameters (Relative to Humanities/Arts)

	Econ./Bus.	Eng./Comp.Sci	Nat. Sci.	No Grad.
<b>Male Students</b>				
Mean	0.507	-1.38	-0.764	-2.07
(Std.)	(4.47)	(3.71)	(3.90)	(3.01)
Median	-0.0381	-0.464	-0.198	-1.59
<b>Female Students</b>				
Mean	-1.36	-3.13	-2.06	-3.53
(Std.)	(4.21)	(3.28)	(3.67)	(2.80)
Median	-1.55	-2.83	-1.61	-3.96

## **B Information on Survey Design and Information Treatments**

Description of data sources provide to survey respondents:

Sources:

1) CPS: The Current Population Survey (CPS) is a monthly survey of about 50,000 households conducted by the Bureau of the Census for the Bureau of Labor Statistics. The survey has been conducted for more than 50 years. The CPS is the primary source of information on the labor force characteristics of the U.S. population. The sample is scientifically selected to represent the civilian non-institutional population.

2) NSCG: The 2003 National Survey of College Graduates (NSCG) is a longitudinal survey, designed to provide data on the number and characteristics of individuals. The Bureau of the Census conducted the NSCG for the NSF (National Science Foundation). The target population of the 2003 survey consisted of all individuals who received a bachelor's degree or higher prior to April 1, 2000.

Methodology:

1) CPS: Our CPS sample is taken from the March 2009 survey. Full time status is defined as "usually" working at least 35 hours in the previous year, working at least 45 weeks in the previous year, and earning at least \$10,000 in the previous year. Average employment rates, average earnings, and percent with greater than \$35,000 or \$85,000 earnings is calculated using a sample of 2,739 30 year old respondents.

2) NSCG: We calculate inflation adjusted earnings using the Consumer Price Index. The salary figures we report are therefore equivalent to CPS figures in 2009 March real dollars. Full time status is defined as in the CPS sample. Given the need to make precise calculations for each field of study group, we use the combined sample of 30-35 year old respondents and age adjust the reported statistics for 30 year olds. This sample consists of 14,116 individuals. To calculate average earnings, we use an earnings regression allowing for separate age intercepts, one each for 6 ages 30-35. The predicted value of earnings from the regression is used as the estimate of average earnings for 30 year olds. For the percent full time employed, and percent with earnings greater than \$35,000 and \$85,000, we use a logit model to predict these percentages for 30 year olds and include a separate coefficient for each of the 6 ages 30-35.

## C Estimation Details

This Appendix describes the approximation of beliefs we use to construct expected lifetime utility from each major. To make clear the relationship between the beliefs questions, which are conditioned on future ages of the respondents, we index age  $q = 22, \dots, 55$ , rather than use time. At period  $t = 1$  (first post-graduation period) in the lifecycle model we assume individuals are aged 22.

### C.1 Beliefs about Own Earnings

For each individual, for each major, and for both the pre- and post- treatment periods, we have 7 data points: i) expected earnings immediately after graduation, ii) expected earnings at age 30, iii) belief that own earnings would exceed \$35,000 at age 30, iv) belief that own earnings would exceed \$85,000 at age 30, v) expected earnings at age 45, vi) belief that own earnings would exceed \$35,000 at age 45, vii) belief that own earnings would exceed \$85,000 at age 45. With 5 major categories, this provides  $5 \times 7 \times 2 = 70$  data points on beliefs about own earnings for each individual respondent.

From this data, we estimate a Normal distribution approximation to individual beliefs about the distribution of earnings for all periods. For each individual  $i$ , we assume beliefs about earnings in major  $k$  follow

$$\ln w_{FT,1,q,i,k} \sim N(\mu_{1,q,i,k}, \sigma_{1,q,i,k}^2),$$

where

$$\mu_{1,q,i,k} = \mu_{1,i,k}^0 + \mu_{1,i,k}^1 q + \mu_{1,i,k}^2 q^2,$$

$$\sigma_{1,q,i,k} = \sigma_{1,i,k}^0 + \sigma_{1,i,k}^1 q.$$

This parameterization allows beliefs in earnings to grow with age  $q$ , following the standard concave pattern. We also allow the variance in beliefs about own earnings to vary over time by allowing the variance parameter to depend on age. The individual specific beliefs parameters consist of  $\omega_{i,k} = [\mu_{i,k}^0, \mu_{i,k}^1, \mu_{i,k}^2, \sigma_{i,k}^0, \sigma_{i,k}^1]$ . We estimate these parameters using a method of simulated moments (MSM) estimator. For any given parameter vector  $\omega_{i,k}$ , we form a sequence of simulated earnings beliefs draws. From this sequence of earnings draws, we construct the simulated counterpart to the 7 statistics detailed above. Our estimator then chooses the  $\omega_{i,k}$

parameters that minimize the quadratic distance between the simulated and actual data beliefs. Note that we estimate  $\omega_{i,k}$  for all individual, majors, and for the pre- and post-treatment states separately.<sup>17</sup>

## C.2 Beliefs about Spouse's Earnings

For self beliefs about future spouse's earnings, we use a similar approximation method. For beliefs about spouse's earnings we economized on data question given the length of survey collection and only asked about the equivalent i)-v) beliefs for spouses. We follow the same model and estimation procedure for spouse's earnings beliefs as with own earning beliefs and estimate a potentially different vector  $\omega_{i,k}$  of parameters for spouses.

$$\ln w_{FT,2,q,i,k} \sim N(\mu_{2,q,i,k}, \sigma_{2,q,i,k}^2),$$

where

$$\mu_{2,q,i,k} = \mu_{2,i,k}^0 + \mu_{2,i,k}^1 q + \mu_{2,i,k}^2 q^2,$$

$$\sigma_{2,q,i,k} = \sigma_{2,i,k}^0 + \sigma_{2,i,k}^1 q.$$

## C.3 Beliefs about Own Labor Supply

For labor supply, we asked respondents to report their beliefs about the probability they would work either full-time, part-time, or not all, conditional on marriage. We asked this information for two time periods: age 30 and age 45. We also asked population beliefs by major about the average hours each individual believes a full time individual works in each major. To conserve on time, this question was only asked in the final post-treatment part of the survey, but the full/part/no work probability question was asked both in the pre- and post- treatment periods. We construct an approximation to the hours beliefs for all periods by assuming full time hours by marriage.

We construct the hours distribution (conditional on marriage  $m_{q,i,k} \in \{0, 1\}$ ) as

---

<sup>17</sup>In order to remove outliers that can happen by chance in the simulated wages, we enforce an earnings ceiling and floor as in the original data. We replace all simulated full-time earnings exceeding \$500,000 with \$500,000 and all simulated earnings less than \$10,000 with \$10,000.

$$h_{1,q,i,k} = \begin{cases} \bar{h}_{1,i,k} & \text{w/ prob. } pr(FT_{1,q,i,k} = 1 | m_{q,i,k}) \\ 20 & \text{w/ prob. } pr(PT_{1,q,i,k} = 1 | m_{q,i,k}) \\ 0 & \text{w/ prob. } 1 - (pr(FT_{1,q,i,k} = 1 | m_{q,i,k}) + pr(PT_{1,q,i,k} = 1 | m_{q,i,k})). \end{cases},$$

where  $\bar{h}_{i,k} = \bar{h}_{30,i,k}1\{q \leq 35\} + \bar{h}_{45,i,k}1\{q > 35\}$  is individual  $i$ 's belief about average full time hours in major  $k$ , which depends on age. Beliefs about part-time hours are assumed to be 20 hours for all individuals and majors.

#### C.4 Beliefs about Spouse's Labor Supply

The distribution of spouse's hours is modeled symmetrically with own labor supply. We therefore set full time hours for spouse's labor supply to 40.

$$h_{2,q,i,k} = \begin{cases} \bar{h}_{2,i,k} & \text{w/ prob. } pr(FT_{2,q,i,k} = 1) \\ 20 & \text{w/ prob. } pr(PT_{2,q,i,k} = 1) \\ 0 & \text{w/ prob. } 1 - (pr(FT_{2,q,i,k} = 1) + pr(PT_{2,q,i,k} = 1)). \end{cases},$$

where  $\bar{h}_{i,k} = \bar{h}_{30,i,k}1\{q \leq 35\} + \bar{h}_{45,i,k}1\{q > 35\}$  is individual  $i$ 's belief about opposite gender's average full time hours in major  $k$ , which depends on age.  $pr(FT_{2,q,i,k} = 1)$  and  $pr(PT_{2,q,i,k} = 1)$  are the beliefs of individual  $i$  about her spouse's probability of working full or part-time at age  $t$  if individual  $i$  graduates with major  $k$ .

#### C.5 Beliefs about Marriage

For marriage, we elicited beliefs about the probability the individual is married for 3 time periods: i) first year upon graduation ( $q = 22$ ), ii) age 30, and iv) and age 40. We use a linear function to interpolation beliefs for all years as follows:

$$pr(m_{q,i,k} = 1) = \begin{cases} pr(m_{22,i,k} = 1) & \text{for } q = 22 \\ pr(m_{22,i,k} = 1) + \frac{pr(m_{30,i,k}=1) - pr(m_{22,i,k}=1)}{30-22}(q - 22) & \text{for } 30 < q < 30 \\ pr(m_{30,i,k} = 1) & \text{for } q = 30 \\ pr(m_{30,i,k} = 1) + \frac{pr(m_{45,i,k}=1) - pr(m_{30,i,k}=1)}{45-30}(q - 30) & \text{for } 30 \leq q < 45 \\ pr(m_{45,i,k} = 1) & \text{for } q \geq 45. \end{cases}.$$