How Do Northwestern Students Infer Returns to MMSS? 
Major Choice Behavior of Northwestern Students 

Haksoo Lee 
Advised by Professor Charles Manski 

Abstract 

This study investigates how college students form their expectations on each major and how they choose their majors in relation to their expectations. In particular, we investigate: (i) the relationship between students perceived ability in each field of study and the likelihood of majoring in that field, and (ii) the relationship between students expected major-specific earnings at age 30 and the likelihood of majoring in each respective field. We find that major-specific earnings expectation plays a significant role in major choice, especially for underclassmen. We find that female students are less influenced by expected academic performance in their major choice, contrary to the findings of Figlio, et al. (2017).

Moreover, we provide upper-bounds to the returns to Economics/MMSS major combination relative to alternative majors; in particular, we estimate the upper-bound for mean log-wage increase for majoring in Economics/MMSS relative to Humanities to be 0.532.

Keywords: college majors; information; subjective expectations; returns to schooling
Acknowledgement

I thank my advisor, Professor Charles Manski for his guidance, patience, and kindness. I thank Professor Joseph Ferrie for his leadership in our thesis seminar. I thank Professor Jeffrey Ely for his leadership in the MMSS program—through which I matured as a scholar and as a person. I thank Nicole Schneider for her attention and support. I thank my parents—who have encouraged me always—for their close presence away. I thank my friends who have made my four years at Northwestern memorable—and I thank JiHo Kim for our time together in MMSS. All too precious, I will remember and cherish.

Introduction

The primary object of this study is to understand how much do college major choices depend on future earnings, perceived ability in each field, and non-pecuniary tastes for each field respectively. In studying determinants of college major choices, this study responds to a broader question: How do youth form their expectation? A better understanding of college major choices is of importance as college major choice represents occupation-specific human capital investment decisions—and as human capital in turn plays a pivotal role in our economy. Denison (1962) and Schultz (1961) argue that human capital affects the growth of an individual's wage, firms' productivity, and national economy. Lepak and Snell (1999) underscore human capital in firms' competitive advantage. Romer (1986) identifies human capital to be a driver of national economic growth. Fang (2006) estimates that college education increases attendees productivity by 40%.

The modern literature on college major choices indeed has found choice of major plays a critical role in determining career paths and future earnings. Weinberger (1999) finds that college graduates with mathematical majors earn more than those with non-mathematical majors. Kim (2015) finds that lifetime earnings depend crucially on choice of major; furthermore, he argues that the choice of major is more pertinent to future earnings than the choice of getting a Bachelor's degree itself. Indeed, evolving distribution of college major choices plays a key role in understanding various socioeconomic phenomena of present, such as the gender wage gap and widening inequality. Eide (1994) attributes the narrowing of gender wage gap to the convergence of male and female college major choices in the 1980s. Weinberger (1998) studies racial and gender wage differentials among college graduates by considering college major choices.

Literature review

Decision-making under uncertainty depends critically on our subjective expectations of the future. At each stage of our lives, we form expectations about the future given who we are. The question was initially motivated by an effort to develop a model of participation for the MMSS program.
are and what we know at that moment. Based on subjective expectations, we make critical choices of our lives: We choose whom to marry and which career to pursue contingent on our imperfect expectations on our uncertain future. Yet, there has not been substantial effort to understand how youth form their expectations. Traditionally, economists have been skeptical of subjective data. Thus, the efforts to study schooling decisions concentrated on forming reasonable assumptions on expectations formation rather than collecting data on expectations.

Reasonable assumptions include the assumption that expectation formation is homogeneous and the assumption that youth condition their belief on the same variables and process information in the same way. A great many studies have assumed that youth have rational expectation. For instance, Willis and Rosen (1979) assume that youth know the actual process generating life-cycle incomes conditional on personal covariates, such as sex, armed force status, and ability.

In response to the culture of making reasonable assumptions, Manski (1993) critiques that rudimentary treatment of expectations has placed the economics of education at an impasse. Manski (1993) discusses two-fold identification problems. One of the problems is that without knowing how youth perceive the returns to school it is not possible to infer decision process behind his decision process. Indeed, if schooling choice is a function of returns to schooling, knowing schooling choice only is insufficient to identify the function (that is, the decision rule) whereas knowing both the perceived returns to schooling and the schooling choice can identify the decision rule. Manski (1993) observes that the youth forming expectation faces the same kind of inferential problem as the econometrician who wants to infer returns to schooling. Since the economists themselves make different inferential assumptions, he argues, there is little reason to believe that expectation formation is homogeneous.

Responding to the need for collecting and employing subjective expectations data, this study contributes to the expanding literature that incorporates subjective expectations to decision-making under uncertainty. Delavande (2008) investigates the choice of contraceptive devices by eliciting subjective expectations for their effectiveness. Zafar (2011) collects a panel data of major-specific expectations from Northwestern undergraduates to study the choice of college major. In his study, he elicits the students expectations on various major related post-graduation outcomes, such as future earnings and future academic performances. For instance, he asks what is the percent chance of graduating with GPA greater than 3.5. Zafar finds that the revision of the expected academic performance was rational in the sense that negative realization of GPA adjusts the expected GPA negatively. In addition, the study finds learning on academic performance influenced the decision to switch major. The methods pioneered in Zafar (2011) and Wiswall and Zafar (2014) are benchmarked for this study.
4 Model

4.1 Major choice model of Wiswall and Zafar (2014)

4.1.1 Specification

In this section, we specify a random utility model of major choice, adopting from Wiswall and Zafar (2014). Let $i$ denote a Northwestern undergraduate choosing a major. She chooses major $j$ from her major choice set $J$. We restrict the space of $J$ to be composed of six elements—$J$ contains Economics/MMSS (Econ), Engineering/Computer Science (Eng), Natural Sciences/Mathematics (NatSci), Humanities (Hum), Arts (Arts), No Graduation (Dropout).

We consider two time periods $t = 0$ and $t = 1$. At $t = 0$, $i$ is yet to be sure of her major, and at $t = 1$ she finalizes her decision. The timing consideration reflects that the major choice decision is made under uncertainty. That is, some of the uncertainties regarding her major choice is resolved between $t = 0$ and $t = 1$; we incorporate such resolvable uncertainty component as preference shocks in the random utility model. As such, we posit the random utility model for $i$ at $t = 0$

$$U_{0,i,j} = EU_{1,i,j} + \alpha \cdot \log(a_j) + \gamma_j + \eta_j$$

where each term is defined as the following:

$EU_{i,j}$ represents post-graduation utility component corresponding to pecuniary aspect of major $j$ following graduation. With a rich set of data, it is possible to construct a sophisticated $EU_{i,j}$ by incorporating expectations on marital status, labor force participation, time-discount, and expected growth of expected earnings over career. However, the survey data for this study is insufficient to support such sophisticated construction of $EU_{i,j}$. Thus, we make a simplifying assumption that reduces $EU_{i,j}$ to be the expected earnings at age 30 if one is assumed to work full-time.

$\log(a_j)$ reflects on perceived ability of $i$ if she were to major in $j$. The model posits that the higher the perceived ability in $j$, the lower would be effort cost of completing major $j$ and the higher would be the chance that she performs well; thus, we expect $\alpha$ to be positive.

$\gamma_j$ reflects on time-invariant taste for major $j$. $\gamma_j$ incorporates such characteristics as delight of major-specific coursework (one can without difficulty imagine Tolstoy deriving not so much delight from a coursework in mathematics). $\gamma_i$ can also be thought of as representing non-pecuniary preference for major specific graduation outcome (that is, Tolstoy may not be so inclined to the life style particular to mathematicians).

$\eta_j$ reflects possible preference shocks between $t = 0$ and $t = 1$. It reflects on the uncertainty that is to be resolved when a major choice is made by $i$. The resolution of uncertainty however does not regard learning on major-specific ability or post-graduation utility. The modeling assumption thus restricts the role of uncertainty in major-choice context to the perception of resolvable uncertainty at $t = 0$. Furthermore, it may be argued that uncertainty component to $EU_{i,j}$ remain unresolved at $t = 1$. 


4.1.2 Identification

The difficulty in identification of the model lies in the unobservable taste component, \( \gamma_j \). Observe that cross-sectional data would not allow separate identification of parameters as we cannot account for the time-invariant taste component. Faced with the identification problem, we adopt fixed-effect identification strategy by leveraging experimentally derived panel of beliefs. Given the utility specification above, subjective choice probability of \( i \) for \( j \) is given by

\[
q_{i,j} = Q_i[U_{0,i,j} > U_{0,i,k} : k \neq j]
\]

where \( Q_i \) is a continuous subjective distribution that \( i \) places on \( \eta \). We assume \( \eta_j \) are i.i.d. across students and major choices and has Type 1 extreme value distribution. Then, the choice probabilities can be written in multinomial logit form

\[
q_{i,j} = \frac{e^{x_{ij}\beta_i}}{\sum_{k=1}^J e^{x_{ik}\beta_i}}
\]

for \( j \in J \) where \( x_{i,j} \) are independent variables discussed above. Applying log-odds transformation, we have the linear mixed logit model

\[
r_{i,j} = \log \left( \frac{q_{i,j}}{q_{i,Hum}} \right) = (EU_{i,j} - EU_{i,Hum}) + \alpha \cdot (\log(a_{i,j}) - \log(a_{i,Hum})) + \psi_{i,j}
\]

where the combined error \( \psi_{i,j} = \gamma_{i,j} - \gamma_{i,Hum} + \epsilon_{i,j} \) consists of relative taste for \( j \) with respect to Humanities and an additional error \( \epsilon \).

The information treatment allows perturbation of beliefs on perceived ability (grade-treatment) and beliefs on future earnings (salary-treatment). Thus, we have the following FE model:

\[
r'_{i,j} - r_{i,j} = [(EU_{i,j} - EU_{i,Hum})' - (EU_{i,j} - EU_{i,Hum})] + \alpha \cdot [(\log(a_{i,j}) - \log(a_{i,Hum}))' - (\log(a_{i,j}) - \log(a_{i,Hum}))]
\]

where \( x' \) is the post-treatment variable of \( x \). Given the nature of information treatments, it can reasonably posited that the unobserved changes in beliefs, \( \epsilon'_{i,j} - \epsilon_{i,j} \), is mean-independent from the changes in earnings-expectations and perceived-ability in major \( j \) (see Wiswall and Zafar (2014) for further discussion on the FE identification strategy). Thus, by forming a fixed-effect model, the model parameters can be identified.

4.1.3 Rounding at extremity

In the previous section, we considered the linear mixed logit model of the form

\[
\log \left( \frac{q_{ij}}{q_{i1}} \right) = (x_{ij} - x_{i1}) \cdot \beta_i = (x_{ij} - x_{i1}) \cdot b + u_{ij}
\]
for $j = 2, \ldots, J$ and where $\beta_i = b + \eta_i$ and $u_{ij} = (x_{ij} - x_{i1}) \cdot \eta_i^2$.

In this section, we resolve the problem of rounding at extremities for the above model. Observe that rounding of subjective probability at extremities (0 and 1) is problematic since it results in the log-odds of $\pm \infty$. In the absence of rounding at extremities, we may invoke the symmetry of preference with respect to $b$ to estimate the parameters of the model. Symmetry implies $u_{ij}$ has median zero conditional on $x_i$. Thus, we estimate the following linear median regression model:

$$M[\log \left( \frac{q_{ij}}{q_{i1}} \right) | x] = b \cdot (x_{ij} - x_{i1})$$

where $b$ can be estimated by LAD. Observe that the median is invariant under any transformation that preserve its rank in the sample. Hence, we take the following approach:

(i) We replace the extreme values (0 and 1) with the values such as 0.001 and 0.999 respectively.
(ii) We estimate the model by LAD.

4.2 Classical model of stated-choice

In this section, we consider a random utility model of stated-choices by applying the methodology discussed in Blass et al. (2008). In contrast to the previous model, $i$ responds to a choice scenario where she is asked to choose one major out of the set of alternatives $J$. We derive the data for stated-choices from our choice probability data by setting the alternative with the highest reported probability to be 1 and setting other majors to be 0. Note, in this model, we deliberately do not take advantage of reported subjective choice probabilities and deliberately consider less-informative data.

We specify the random utility model for student $i$ at $t = 0$ as:

$$U_{0,i,j} = x_{ji} \cdot \beta_i + \eta_{ji}$$

where $x_j$ is the vector containing observable characterics, in particular $EU_{1,j}$ and $\log(a_j)$.

A key distinction from the above model lies in that $\eta_j$ is known to the decision maker, but not to the researcher. We assumed that $\eta_i \equiv (\eta_{ji} : j = 1, \ldots, J)$ is i.i.d. conditional on

---

2With the assumption that cross-sectional distribution of $\beta$ is statistically independent from $x$, we have $E(\beta) = b$ and $E(u_{ij}|x) = 0$. Note further that $u_{ij} = (x_{ij} - x_{i1}) \cdot \eta_i$ implies $u$ is heteroskedastic, but it does not affect consistency of estimation.

3Suppose student $i$ reports that she is 60% likely she would major in Economics and 40% likely she would major in Humanities, and student $j$ reports that she is 99% likely she would major in Economics and 1% likely she would major in Humanities. Observe that, if they were to report one choice out of the two alternatives, they would both report Economics. And in the data the varying degree of certainty in their choices would not be reflected.

4In particular, we compare the results from the stated-choice model to the pooled cross-sectional OLS estimates of (1) above.

5We derive $EU_{1,j}$ and $\log(a_j)$ as previously.
$x_i \equiv (x_{ji} : j = 1, \cdots, J)$ and has Type I extreme value distribution. Then, it follows the probability of choosing major $j$ has the multinomial logit form

$$P(y = j|x_i, \beta_i) = \frac{e^{x_{ij}\beta_i}}{\sum_{k=1}^{J} e^{x_{ik}\beta_i}}.$$  

And, with a further assumption that $\beta$ is statistically independent of $x$ with the density $f(\beta|\theta)$ in the population (which is known up to parameter $\theta$), by McFadden and Train (2000), we have

$$P(y = j|x_i, \theta) = \int \frac{e^{x_{ij}\beta_i}}{\sum_{k=1}^{J} e^{x_{ik}\beta_i}} f(\beta|\theta)d\beta.$$  

Then, the model can be estimated by maximum likelihood.

4.3 Maximum score estimation

In this section, we consider the Maximum Score Estimation of Manski (1975, 1985) to estimate the main regression model (1).  

Recall that in the standard multinomial logit form we supposed that $\eta_{hj}$ is i.i.d. with Type I extreme value distribution. Instead, by adopting maximum score estimation, we may suppose weaker conditions that:

(i) Each individual $i$ places median zero on $\eta_{hj} - \eta_{hk}$.
(ii) The cross-sectional distribution of $\beta$ is symmetric.

4.3.1 Binary choice setting

For a given agent, denote as previously the utility from choosing major $j \in J$ to be

$$u_j = x_j'\beta + \eta_j.$$  

Consider an econometrician using choice data on two choices $i$ and $j$. Let $a = 1, \cdots, n$ be observations, in which each agent selects one from $i$ and $j$.

Assuming the rank ordering property above, by Fox (2007), we may estimate $\beta$ consistently using the pairwise maximum-score estimator. The pairwise maximum-score estimator maximizes

$$Q_{ij}^n(\beta) = \frac{1}{n} \sum_{n=1}^{n} (1[ai] \cdot 1[x_{ai}'\beta > x_{aj}'\beta] + 1[aj] \cdot 1[x_{aj}'\beta > x_{ai}'\beta]),$$  

where $1[\cdot]$ is the indicator function and $1[ai]$ outputs 1 if the agent selects $i$ in the observation $a$.

---

6 Revelt and Train (1998), for example, uses stated-choice data to estimate mixed logit models of this kind.

7 The maximum score estimator approach in this section adopts from Blass, et al. (2008) and Fox (2007).

8 (i) and (ii) imply that, for a given agent, and for $i,j \in J$, $x_i'\beta > x_j'\beta$ if and only if $P(i|x, J, \beta) > P(j|x, J, \beta)$.  

---
4.3.2 More than two choices

Fox (2007) generalizes the above objective function to include more choices. The pairwise maximum-score objective function using choice data on $K$ choices is

$$Q^K_n(\beta) = \sum_{i=1}^{K-1} \sum_{j=i+1}^K \frac{1}{n} \sum_{n=1}^n (1[ai] \cdot 1[x_{ai}\beta > x_{aj}\beta] + 1[aj] \cdot 1[x_{aj}\beta > x_{ai}\beta]).$$

Then, as the indicator function is not differentiable, we may solve the above optimization problem using a grid-search.

5 Data

This section discusses the survey administration, the survey instrument, and descriptive sample statistics.

5.1 Administration

The data for this study originates from a survey instrument administered on Northwestern undergraduate students. The survey had been conducted from March to May 2017. The students were asked to participate by email, by public Facebook postings, by individual encounters at Norris University Center and the Main Library of Northwestern University. The email lists were generously provided by the MMSS major coordinator and by professors of certain undergraduate classes. The recruitment material and the actual survey instrument are included in Appendix. The survey was limited to full-time undergraduate students. The survey is administered through Qualtrics, an online survey platform. The participants accessed the survey through an anonymized link. In the beginning of the survey, the participants were told to complete the survey in one seating (participants were informed that responses filled in after thirty minutes will be invalidated). Revise of previous responses was not allowed. The survey included several logical checks such as that checks whether the total sum of subjective probabilities of exhaustive events is 1. Three waves of the survey were administered. After each iteration, minor clarifications were made, and a few typos were fixed. The initial iteration respondents were not provided with financial incentive; in the last two iterations, participants were promised $20 conditional on winning a lottery that will draw ten students at random.

---

9 I thank Nicole Scheneider for her help in the times of difficulty
10 Emails were sent out to the students of Introduction to Macroeconomics, Public Finance, Money and Banking, Constitutional Law, Machine Learning. I especially thank Professor Walker for his support for my study. Time permitting, students of other disciplines such as Arts and Journalism would have been reached out to; the sample collected as such would have been less susceptible to selection bias.
11 The participants were asked to provide a contact email for financial compensation. It was made clear in the beginning of the survey the personal email would bear no relation to the actual survey responses so that anonymity would be insured. The participants were informed that the lottery winners will be announced publicly on a set date on Facebook for transparency.
5.2 Survey Instrument

The survey instrument forms a single-survey panel data of students, eliciting their subjective expectations about future events, such as graduating majors, graduating GPAs, and earnings at age 30. A panel of beliefs allows identification of the choice model developed above.

The survey proceeds in three stages. In the first stage, the survey elicits (i) each participant’s subjective probabilities of majoring in each of six possible major fields (Economics/MMSS, Engineering/Computer Science, Natural Sciences/Math, Humanities, Arts, and Not Graduate), (ii) subjective probabilities of earning at least $35,000 and $85,000 at age 30 if she were to major in each field, and (iii) her perceived ability in each field expressed in percentile rank. In the second stage, the participants were provided with one of two possible information treatments. One provides post-graduation starting salary information from Class of 2015 as disclosed by Northwestern Career Advancement, and the other poses a hypothetical situation in which the participant learns a new information about her ability in each field. In the third stage, the survey re-elicits the subjective expectations through the same set of questions as in the first stage. The survey concludes with demographic questions that elicit such information as SAT scores, education level of parents, socioeconomic status, and current GPA. The full survey is included in Appendix.

5.3 Sample selection and descriptive statistics

A total of 130 students participated in the survey, 106 of whom responded in full and 24 of whom responded to at least the first half of the survey. For analysis, a ceiling of $500,000 was imposed on the earnings expectations at age 30. We record extreme probabilities, 0 and 1, with 0.001 and 0.999 respectively. Such modification of data responds to an argument made by Blass et al. (2010) that dropping individuals who report extreme probabilities may lead to selection bias. The utility model of this study incorporates the perceived ability component and the expected post-graduation utility component as represented by earnings expectation at age 30. Such beliefs were elicited by asking:

(i) Consider a situation in which you will graduate with a Bachelors degree in each of the following major categories. Think about other individuals at Northwestern who will graduate with a major in each category. In PERCENTILE, in terms of your graduating GPA, what would be your rank in each major category compared to other students at Northwestern who will graduate with a major in each category?

Note: 99 percentile would mean your GPA would be in the TOP 1%; correspondingly, 1 percentile would mean that your GPA would be in the BOTTOM 1%(or equivalently 99% of the people would have higher GPA than you).

The participants were directed to only one of the two information treatment, and in the grade-treatment the participants were only hypothesized about a grade outcome of a class that they have not taken (these were insured by Qualtrics coding of the survey flow).

The survey was approved by the IRB with the study number STU00204728.
(ii) When you are 30 YEARS OLD, if you are working FULL-TIME and you received a Bachelors in each of the following major categories, what do you believe is the percent chance that you would earn:

At least $35,000 per year
At least $85,000 per year

Observe that by asking for subjective probabilities for earnings realizations at age 30 the survey elicits subjective probability distribution of future earnings, which is more informative than reports of mean. A summary of sample characteristics is shown in Table 1.

6 Reduced-form analysis

In this section, we examine patterns of future earnings expectation and of perceived ability. We also examine the effects of information treatments on beliefs. We confirm that students revise their beliefs in a logical manner. Lastly, we examine the relationship between the expected earnings belief and the major choice probabilities.

6.1 Expected future earnings beliefs

Table 2 documents earnings at age 30 beliefs of the students for each field of study. Column (1) show the mean and standard deviation (SD) of the earnings distribution. From the pre-treatment responses, we observe that the expected self-earnings are the highest for Engineering (at $357,800), followed by Economics and Kellogg Certificate combination at ($357,400), and Economics and MMSS combination at ($344,700). The expected self-earnings belief was the lowest for the Not Graduate category (at $96,500). We observe from high SD a significant heterogeneity in future self-earnings beliefs. Similar results hold for the post-treatment observations. Column (2) and (3) show the mean and SD of percentage-revision and absolute-revision in future self-earnings beliefs respectively. The belief revisions also show significant heterogeneity. We observe that the absolute revision is substantially larger than the percentage revision.

Table 3 considers only those students whose revisions were less than $50,000 who received the salary-treatment. The general findings of Table 2 persist. Note the earnings expectation is the highest for the Economics and MMSS combination (at $387,900).

6.1.1 Notes on the derivation of expected earnings distribution

We obtained the expected earnings distribution parametrically by fitting the log-normal distribution. That is, for the parameters \((m, q)\) of the log-normal distribution \(F\), we compute for each student \(j\)

\[
\text{Argmin}_{m, q} \sum_{i=1}^{2} [F_{ji} - F(y_{ji}; m, q)]^2
\]

where \(F_{j1} = P(\text{earnings at age 30} \leq 35)\), \(F_{j2} = P(\text{earnings at age 30} \leq 85)\), \(y_{j1} = 35\), \(y_{j2} = 85\). We impose bounds on candidate \((m, q)\) by setting \(m \in [0, 4.5]\) and \(q \in [0, 4]\). Such restrictions are reasonable given the distribution of parameters obtained without restriction. Then, we calculate the mean and standard deviation from thus obtained parameters. For the means above 50 we truncate it to 50.
6.2 Major choice probabilities

Subjective major choice probabilities were elicited by asking:

What do you believe is the percent chance that you would either graduate from Northwestern with a major in the following major categories (considering your PRIMARY MAJOR) or that you would never graduate/drop-out?
Table 4 shows descriptive statistics of graduating major expectation. From Column (1), we observe that Economics/MMSS has the highest mean pre-treatment probability of majoring. Such result may be due to selection bias as Economics major is over-represented in the sample (cf. Table 1). Perhaps not surprisingly, Drop-out category has the lowest mean probability. Standard deviation is high across all majors, which is indicative of heterogeneity of beliefs. Column (2) and (3) document revision of beliefs following information treatment. Both columns show large standard deviations. Notably, Column (3) shows that the log-odds
of majoring in Economics/MMSS increased by 14 percentage points relative to Humanities. Column (4) shows that many students reported extreme probabilities (0 or 1) of majoring in each major category. Notably, for Arts and Drop-out categories, the percentage of extreme probabilities is over 70%. A closer look at the data shows that many reported 0 for the two categories (rather than 1). Thus, it may be inferred that Arts and Drop-out are not in the choice set of many students in the sample.

6.3 College major choice probabilities and beliefs about own earnings and ability

In this section, we examine the relationship between the major choice probabilities and the beliefs about future earnings and ability. We regress the major choice probability on the (log)-expected future earnings, the perceived ability and various control variables. That is, we perform regression of the form:

\[
\log q_{k,i} = \beta_0 + \beta_1 \log w_{k,i} + \beta_2 \log a_{k,i} + \delta C_i + \nu_k + \psi_{k,1}
\]

where \(q_{k,i}\) is the reported probability of graduating with major \(k\); \(w_{k,i}\) is the expected earnings at age 30 if \(i\) were to major in \(k\); \(a_{k,i}\) is the perceived ability in major \(k\) in percentile-rank; \(C_i\) is a vector of individual-specific characteristics such as gender and race; \(\nu_k\) major dummies incorporating major \(k\) fixed effect; and \(\psi_{k,1} = \gamma_k - \gamma_{\tilde{k}} + \epsilon_k\) is the composite error term consisting of relative taste for major \(k\) (with respect to major \(\tilde{k}\)) and the additional residual error. We set reference major \(\tilde{k}\) to be Humanities.

The log-log format of the regression allows elasticity interpretation of the coefficient \(\beta_1\). Column (1) of Table 5 documents the results from the pooled cross-sectional OLS regression. In this regression, the pre-treatment observations were pooled with the post-treatment observations to estimate the model. We estimate that 1% increase in expected earnings (relative to Humanities) would result 0.3% increase in the odds of majoring (relative to Humanities). We find the result to be significant at 1% significance level. This estimate is far more conservative than the estimate of Wiswall and Zafar (2014), which reports an
estimate of 1.6. The result also shows that the perceived ability has a strong positive
correlation with the major choice probability. Moreover, from Panel A, which considers
the full-sample, we observe that the choice elasticity with respect to perceived ability is positive
at 1% significance level. Note, as students report future earnings expectations and perceived
ability for all major categories, the regression resolves the selection problem, which would
manifest if one were to use the chosen major and its expected earnings. Column (2) of
Table 5 reports the estimates from the median regression. As discussed previously, the LAD
estimation is robust under rounding in reported probabilities. We observe that the LAD
estimates tend to be more conservative than those of OLS.

A key weakness for such pooled-OLS is that one cannot separately identify the taste
component $\gamma$ from the earnings component; if the taste component is correlated with the
earnings component, then the cross-sectional estimate of the choice elasticity with respect
to expected earnings would be biased. We resolve this problem by taking individual within-
difference to clear the time-invariant taste component:

$$
\left( \log q_{k,i} - \log q_{\tilde{k},i} \right) - \left( \log q_{k,i} - \log q_{\tilde{k},i} \right) = \beta_0 + \beta_1 \left[ \left( \log w_{k,i} - \log w_{\tilde{k},i} \right) \right] \\
+ \beta_2 \left[ \left( \log a_{k,i} - \log a_{\tilde{k},i} \right) \right] \\
+ \nu_k + \epsilon_{k,1} - \epsilon_{\tilde{k},1}
$$

Column (3) of Table 5 reports the estimates of $\beta_1$ and $\beta_2$. The FE estimates are con-
trasted with the OLS estimates. We no longer observe a positive relationship between the
expected earnings and the major choice probability (the FE estimate of $\beta_1$ is not significant
at 10% significance level). The difference suggests that the unobserved taste component is
positively correlated with the earnings expectation, and thus results in an upward bias of
the estimate in OLS. Considering Panel B, which restricts the sample to the underclassmen
(freshmen and sophomores), we observe higher and more significant estimates of the choice
elasticities; the choice elasticity with respect to expected earnings (the Panel B FE estimate)
is significant at 5% significance level.

From the series of regressions, we conclude that, at least for underclassmen, the earnings
expectation plays a significant role in the choice of major. However, for upperclassmen
(juniors and seniors), the earnings expectation plays only a marginal role in their choice of
major. It may even be argued that the earnings expectation and perceived ability play
no role in their decision of major choice as the restricted sample of juniors and seniors show
that both estimates of $\beta_1$ and $\beta_2$ are not significant at 15% significance level.

Such speculation resonates with the introspection that I had no intention of switching
my major after the second year at Northwestern. Such conviction in my major choice fol-
lowed from several reasons. First, as speculated in Wiswall and Zafar (2014), switching
cost was prohibitively high; as an international student paying full-tuition, if such decision

---

14 The p-value for $\beta_1$ in the Panel A FE estimate is 0.931
15 The p-values for $\beta_1$ and $\beta_2$ are 0.360 and 0.279 respectively.
would extend undergraduate study by one year, it would have costed me extra $70,000. And it is noteworthy that the cost of switching is contingent on the proximity of two majors considered—that is, it was more expensive to switch from Industrial Engineering into Arts or Journalism, having had no exposure to them, and less expensive to switch into economics/MMSS whose prerequisites could be fulfilled by some of the engineering coursework. Therefore, the initial choice of study has an effect of reducing the major choice set significantly. Second, if I were to switch into another career path, say a career in law, I would prefer to adjust my coursework within the chosen major and defer the conclusive transition to the graduate school (say, an extra year of philosophy at Oxford). I speculate one of the reasons for the popularity of economics major to be its generality and low entry cost. With a degree in economics, a student may plausibly prepare for an entry into various industries and even law school. Also, the bare requirement for obtaining economics degree does not enforce high level of mathematical maturity. Mathematics, I have observed, due to its cumulative nature, can be a significant barrier to entry into many disciplines.

Interpretation of the estimates of $\beta_2$ is more difficult as the pooled cross-sectional OLS estimates show a strong positive correlation, yet the FE estimates do not support such findings. It can be speculated that the students surveyed had trouble considering a hypothetical scenario of receiving a (generally) negative grade feedback. If the students revised downward their perceived ability in the relevant major field in their post-treatment responses but not revised their major choice probabilities, negative estimates would result. Indeed, if we restrict the sample to those who received the salary-treatment, any discernable relationship between the perceived ability and major choice probability disappears. However, such speculation is not well founded as the sample size for such regression is too limited to be conclusive.

7 Stated-choice analysis

In this section, we report the results for the classical stated-choice model discussed in Section 4.3. Recall we estimate

$$P(y = j|x_i, \theta) = \frac{e^{x_i \beta_i}}{\sum_{k=1}^{J} e^{x_i \beta_i}} f(\beta|\theta) d\beta$$

(where $x$ includes the expected earnings and perceived ability) by the Maximum Likelihood.

---

16 Before deciding on a major, I have been a student of various disciplines. I entered the college as an industrial engineering major, flirted with computer science, and finally decided on mathematics, economics/MMSS. I can personally testify that there is no associated switching cost within McCormick School of Engineering, if the decision is made during the freshman year, as all engineering majors take (mostly) the same set of classes.

17 The students were asked hypothesize about a situation in which they received B- in a course they have not taken. Most students in our sample have GPA above 2.7

18 In such regression on the restricted sample, $\beta_1$ and $\beta_2$ had p-values of 0.846 and 0.396 for Panel A and 0.253 and 0.449 for Panel B respectively. The result of the regression is included in Appendix.
<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Log-odds of major relative to Hum</th>
<th>Log-odds revision (post-pre)</th>
<th>Fixed-effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) OLS</td>
<td>(2) LAD</td>
<td>(3) OLS</td>
</tr>
<tr>
<td><strong>Panel A: Full sample</strong></td>
<td><strong>Pooled Cross-section</strong></td>
<td><strong>Individual covariates and major dummies included</strong></td>
<td></td>
</tr>
<tr>
<td>Log self-earnings</td>
<td>0.299*** (0.089)</td>
<td>0.328*** (0.127)</td>
<td>0.00806 (0.0932)</td>
</tr>
<tr>
<td>Log self-earnings revision</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log perceived-ability</td>
<td>0.864*** (0.156)</td>
<td>0.765*** (0.185)</td>
<td>-0.119** (0.0457)</td>
</tr>
<tr>
<td>Log perceived-ability revision</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>R²</strong></td>
<td>0.474</td>
<td>0.195</td>
<td></td>
</tr>
<tr>
<td>Total observations</td>
<td>406</td>
<td>406</td>
<td>190</td>
</tr>
<tr>
<td>Number of respondents</td>
<td>56</td>
<td>56</td>
<td>51</td>
</tr>
<tr>
<td><strong>Panel B: Underclassmen (Freshmen and sophomore)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log self-earnings</td>
<td>0.453*** (0.134)</td>
<td>0.369*** (0.257)</td>
<td>0.200** (0.0852)</td>
</tr>
<tr>
<td>Log self-earnings revision</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log perceived-ability</td>
<td>0.803*** (0.209)</td>
<td>0.809*** (0.183)</td>
<td>-0.105** (0.0478)</td>
</tr>
<tr>
<td>Log perceived-ability revision</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>R²</strong></td>
<td>0.447</td>
<td>0.195</td>
<td></td>
</tr>
<tr>
<td>Total observations</td>
<td>214</td>
<td>214</td>
<td>100</td>
</tr>
<tr>
<td>Number of respondents</td>
<td>29</td>
<td>29</td>
<td>27</td>
</tr>
</tbody>
</table>

**Notes:** Heteroskedastic cluster robust standard error in parentheses; SEs were adjusted for clustering at the individual level; individual covariates include an indicator for gender; indicators for race; GPA indicators (0-3.0, 3.0-3.5, 3.5-3.8, 3.8-4.0); SAT math/verbal scores; indicators for parents’ education level (BA or more); an indicator for non-reported SAT; indicators for socioeconomic status; major dummies. Standard error in
Table 6: Major choice likelihood and expected earnings/perceived ability
(classical stated-choice model)

<table>
<thead>
<tr>
<th></th>
<th>Mixed-logit model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Maximum Likelihood</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
</tr>
</tbody>
</table>

**Panel A: Full-sample**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Log self-earnings belief</td>
<td>1.094***</td>
</tr>
<tr>
<td></td>
<td>(0.346)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.002</td>
</tr>
<tr>
<td>Log perceived-ability</td>
<td>10.722***</td>
</tr>
<tr>
<td></td>
<td>(2.226)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.000</td>
</tr>
<tr>
<td>Total observations</td>
<td>963</td>
</tr>
</tbody>
</table>

**Panel B: Underclassmen (Freshmen and sophomores)**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Log self-earnings belief</td>
<td>0.999**</td>
</tr>
<tr>
<td></td>
<td>(0.446)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.025</td>
</tr>
<tr>
<td>Log perceived-ability</td>
<td>12.265***</td>
</tr>
<tr>
<td></td>
<td>(3.03)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.000</td>
</tr>
<tr>
<td>Total observations</td>
<td>589</td>
</tr>
</tbody>
</table>

*Notes:* The independent variables whose coefficients are random are (log)-self-earnings belief and (log)-perceived-ability expressed in the percentile rank; the number of Halton draws used for simulation is 50,000; the fact that the data is panel is taken into account by specifying identifier variable; standard error in parenthesis; ***, **, * indicate significance at 1%, 5%, 10% respectively.
Immediately visible from Table 6 is that the classical stated-choice model results in estimates of $\beta_1$ and $\beta_2$ that are far greater than those of our baseline model that takes full advantage of elicited subjective probabilities. Table 5 Column (1)—the pooled cross-sectional OLS estimates—reports the estimate of coefficient associated with the expected earnings to be 0.299 whereas the stated-choice model reports 1.094. Similarly, the pooled cross-sectional OLS estimate of the coefficient associated with the perceived ability is 0.864 whereas the stated-choice model estimates 10.722. Albeit different in magnitude, the estimates from the two models agree in the ordinal sense (that is, in both cases, the perceived ability plays more significant role).

Previously, in Section 6.3, we concluded that both earnings expectations and perceived ability in each major field bear positive relationship with the likelihood of majoring in each respective field. Thus, we may reason that in the stated-choice model the role of expected earnings and perceived ability in major choice decision would be biased upward. Suppose student $i$ believes that she would major in Economics with 51% chance and would major in Humanities with 49% chance, and believes that she would rank 51-percentile in Economics (compared to her classmates) and would rank 49-percentile in Humanities. Then, in the stated-choice model, such uncertainty in major choice probability would be veiled as she reports only that she would major in Economics. Consequently, the econometrician observing Economics as her stated-choice would overemphasize the role of marginal advantage in perceived ability in her major choice decision whereas in the case of the baseline model such uncertainty would be taken into account. As expected, Table 6 confirms the advantage of our baseline model employing subjective probability data over the classical stated-choice model.

8 Maximum score estimation

In this section, we report the results from the maximum score estimation discussed in Section 4.4. We argued previously that the vector of parameters that maximize the objective function

$$Q^K_n(\beta) = \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \frac{1}{n} \sum_{n=1}^{n} (1[a_i] \cdot 1[x_{ai}^\prime \beta > x_{aj}^\prime \beta] + 1[a_j] \cdot 1[x_{aj}^\prime \beta > x_{ai}^\prime \beta]).$$

is consistent 19. We solve the above optimization problem by grid-search as the above objective function is not differentiable. We first consider a large, coarse grid to observe a general trend of the scoring function, and then zero-in on the region of interest with a finer grid 20. We pool the pre-treatment responses with the post-treatment responses for estimation 21. The results from the maximum score estimation is unfortunately inconclusive.

19 Here, the vector of observable characteristics $x$ includes the (log)-expected earnings and the (log)-perceived ability.
20 First $[-10, 10]$ by $[-10, 10]$ region with 0.1 interval was considered, and then $[-2, 2]$ by $[-2, 2]$ region with 0.01 interval was considered.
21 We consider 106 students with two responses each; since in each response students respond to each of the five major categories, we have 1,060 observations that can be considered pairwise.
Figure 1: Maximum score estimation by grid-search

Figure 2: Densities of $\beta_1$ and $\beta_2$

Notes: Histograms of grid-searched parameters above; kernel-densities below; the grid [-2,2] by [-2,2] was searched in 0.01 interval; the parameter values counted are those with the maximum score or the maximum score minus one.
The plot of the score function indicates that the limit point of the estimators would occur in the positive region \((0, \infty) \times (0, \infty)\). In particular, the maximum score occurs at \(\hat{\beta}_1 \in [0.01, 0.04]\). However, with the \([-2, 2] \times [-2, 2]\) grid, the numerical optimization merely shows that \(\hat{\beta}_2 \in [0.50, 2.00]\) (see Figure 1) \(^{22}\) Hence, we conclude the following bounds on the parameters
\[
(\beta_1, \beta_2) \in [0.01, 0.04] \times [0.50, \infty].
\]

Although the convergence of parameters cannot be directly observed in the numerical optimization, by counting parameter values at which the maximum score (or that which is close to) occurs, we may fathom the distributions of the parameters. For instance, from the histogram and kernel-density (see the left panel of Figure 2), we may reasonably infer that the distribution of \(\beta_1\) is centered around 0.03. On the other hand, from Figure 2, we observe that the obtained kernel-distribution for \(\beta_2\) is less stable with high variance \(^{23}\) Albeit inconclusive, the maximum score estimation shows that, even under weaker assumptions, the positive relationships that the likelihood of majoring in each field bears with the expected earnings and the perceived ability in that field are maintained.

9 Heterogeneity

9.0.1 Gender

We consider two models of choice to investigate heterogeneity in major choice: the classical stated-choice model and the subjective probability model of Wiswall and Zafar (2014). The stated-choice model can be considered for robustness \(^{24}\). And, conversely, when the two models depict different relationships between groups, we would be alerted to be more cautious in our interpretation of coefficients, in particular, those of the stated-choice model \(^{25}\).

Table 7 illustrates gender heterogeneity in major choice. Figlio, et al. (2017) finds from a revealed choice data that female students are more receptive to negative grade and thus less persistent in pursuing their initially declared major. Yet, contrary to their finding, we find from both models that female students are less influenced by perceived ability (or expected academic performance) than male students. Moreover, the subjective probability model suggests that female students tend to be more receptive to earnings expectations in their major choice.

9.0.2 Race

Table 8 illustrates racial heterogeneity in major choice. Notably, the stated-choice model posits that Asians tend to be more receptive to perceived ability, yet the subjective probability model suggests contrarily that the difference is marginal.

\(^{22}\)Even with a larger grid, the estimate of \(\beta_2\) cannot be bounded above.

\(^{23}\)The kernel-density shifts right if we increase the upper-bound of the search region; thus, the peak of kernel-density for \(\beta_2\) is inconclusive.

\(^{24}\)If both models suggest a certain pattern of heterogeneity, then we would consider the result to be more robust. That is, even if we were to administer a stated-choice survey, we would obtain the same result.

\(^{25}\)Thus, such instance would suggest weakness of the stated-choice model.
<table>
<thead>
<tr>
<th></th>
<th>Mixed-logit model</th>
<th>Cross-sectional pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Maximum Likelihood</td>
<td>OLS</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td><strong>Log self-earnings belief</strong></td>
<td>1.594***</td>
<td>0.252*</td>
</tr>
<tr>
<td></td>
<td>(0.393)</td>
<td>(0.131)</td>
</tr>
<tr>
<td><strong>Log perceived-ability</strong></td>
<td>12.329***</td>
<td>1.747***</td>
</tr>
<tr>
<td></td>
<td>(3.599)</td>
<td>(0.226)</td>
</tr>
<tr>
<td><strong>Total observations</strong></td>
<td>553</td>
<td>428</td>
</tr>
<tr>
<td><strong>Panel B: Female</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Log self-earnings belief</strong></td>
<td>0.442</td>
<td>0.345**</td>
</tr>
<tr>
<td></td>
<td>(0.506)</td>
<td>(0.160)</td>
</tr>
<tr>
<td><strong>Log perceived-ability</strong></td>
<td>10.372***</td>
<td>1.245***</td>
</tr>
<tr>
<td></td>
<td>(2.965)</td>
<td>(0.226)</td>
</tr>
<tr>
<td><strong>Total observations</strong></td>
<td>410</td>
<td>328</td>
</tr>
</tbody>
</table>

**Notes:**
Column (1): The independent variables whose coefficients are random are (log)-self-earnings belief and (log)-perceived-ability in the percentile-rank; the number of Halton draws used for simulation is 50,000; identifier variable was included as we consider a panel data.
Column (2): Heteroskedastic cluster robust standard error in parentheses; standard errors were adjusted for clustering at the individual level; no other control variable was included.
***, ***, * indicate significance at 1%, 5%, 10% respectively.
9.0.3 Socioeconomic status

Table 9 shows heterogeneity in major choice with respect to socioeconomic status. We observe that the stated-choice model suggests (perhaps surprisingly) that the upper/upper-middle class students tend to be more receptive to expected earnings. The subjective probability model, however, suggests such difference is only marginal. We observe also that the two models draw different conclusions regarding the role of perceived ability—the stated-choice model posits the lower/lower-middle class students base their decision on perceived ability more heavily while the subjective probability model posits contrarily. We also note that in all three heterogeneity analysis the stated-choice model produce estimates that are far greater than those of the subjective probability model.

10 Bounding the returns to MMSS

In this section, we bound the returns to the MMSS program relative to other major categories by taking the approach developed in Manski and Pepper (2000).

We view each student \( j \) has a log-wage function \( y_j(t) \) mapping major \( t \in T \) to log-wage \( y \in Y \). \( y_j(t) \) can thus be interpreted as student \( j \)'s production function for human capital.

<table>
<thead>
<tr>
<th></th>
<th>Mixed-logit model</th>
<th>Cross-sectional pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Maximum Likelihood</td>
<td>OLS</td>
</tr>
<tr>
<td>Panel A: Asian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log self-earnings belief</td>
<td>0.801**</td>
<td>0.308**</td>
</tr>
<tr>
<td></td>
<td>0.340</td>
<td>0.131</td>
</tr>
<tr>
<td>Log perceived-ability</td>
<td>16.684***</td>
<td>1.424***</td>
</tr>
<tr>
<td></td>
<td>4.239</td>
<td>0.270</td>
</tr>
<tr>
<td>Total observations</td>
<td>491</td>
<td>377</td>
</tr>
<tr>
<td>Panel B: Non-Asian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log self-earnings belief</td>
<td>1.219**</td>
<td>0.360**</td>
</tr>
<tr>
<td></td>
<td>0.519</td>
<td>0.148</td>
</tr>
<tr>
<td>Log perceived-ability</td>
<td>6.559***</td>
<td>1.475***</td>
</tr>
<tr>
<td></td>
<td>2.010</td>
<td>0.265</td>
</tr>
<tr>
<td>Total observations</td>
<td>472</td>
<td>379</td>
</tr>
</tbody>
</table>

Notes:
Column (1): The independent variables whose coefficients are random are (log)-self-earnings belief and (log)-perceived-ability in the percentile-rank; the number of Halton draws used for simulation is 50,000; identifier variable was included as we consider a panel data.
Column (2): Heteroskedastic cluster robust standard error in parentheses; standard errors were adjusted for clustering at the individual level; no other control variable was included.
***, **, * indicate significance at 1%, 5%, 10% respectively.
We impose a weak ordering on the major choice set $T$ such that

Economics and Kellogg Certificate $\succeq$ Engineering $\succeq$ Economics and MMSS $\succeq$ Economics $\succeq$ Natural Sciences $\succeq$ Humanities $\succeq$ Arts $\succeq$ Drop out

We shall focus on the estimates of the mean returns to major $t$ relative to major $s$ where $s \leq t$. That is, we consider the average treatment effect

$$\Delta(s, t) = E[y(t)] - E[y(s)].$$

### 10.1 General results

We consider $\Delta(s, t)$ under the assumption of monotone treatment response (MTR) and the assumption of monotone treatment selection (MTS). Adopting from Manski and Pepper (2000), we first derive a generally applicable bound on the average treatment effect from Assumption MTS and MTR.

Assumption MTR states that: For all individual $j$ and all treatment pairs $(s, t)$,

If $t \geq s$, then $y_j(t) \geq y_j(s)$.

where $y_j$ is the response function of individual $j$ such that $y_j : T \to Y$.\(^{26}\) Observe that $T$ is the set of treatments and $Y$ is the outcome space.

---

<table>
<thead>
<tr>
<th>Panel A: Lower or lower-middle class</th>
<th>Mixed-logit model</th>
<th>Cross-sectional pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Maximum Likelihood</td>
<td>OLS</td>
</tr>
<tr>
<td>Log self-earnings belief</td>
<td>0.692</td>
<td>0.409**</td>
</tr>
<tr>
<td></td>
<td>(0.584)</td>
<td>(0.201)</td>
</tr>
<tr>
<td>Log perceived-ability</td>
<td>11.118**</td>
<td>1.373***</td>
</tr>
<tr>
<td></td>
<td>(4.402)</td>
<td>(0.354)</td>
</tr>
<tr>
<td>Total observations</td>
<td>311</td>
<td>232</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Upper or upper-middle class</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log self-earnings belief</td>
<td>1.165***</td>
<td>0.358***</td>
</tr>
<tr>
<td></td>
<td>(0.358)</td>
<td>(0.104)</td>
</tr>
<tr>
<td>Log perceived-ability</td>
<td>9.511***</td>
<td>1.425***</td>
</tr>
<tr>
<td></td>
<td>(2.015)</td>
<td>(0.208)</td>
</tr>
<tr>
<td>Total observations</td>
<td>877</td>
<td>691</td>
</tr>
</tbody>
</table>

Notes:

Column (1): The independent variables whose coefficients are random are (log)-self-earnings belief and (log)-perceived-ability in the percentile-rank; the number of Halton draws used for simulation is 50,000; identifier variable was included as we consider a panel data.

Column (2): Heteroskedastic cluster robust standard error in parentheses; standard errors were adjusted for clustering at the individual level; no other control variable was included.

***, **, * indicate significance at 1%, 5%, 10% respectively.
Assumption MTR allows each individual $j$ to have unique response function. Note also that Assumption MTR is not refutable. We assume, in what follows, the outcome space $Y$ is a closed subset of $[-\infty, \infty]$.

Assumption MTS states that: For each $t \in T$, 

If $s' \geq s$, then $E[y(t)|z = s] \geq E[y(t)|z = s]$.

Note Assumption MTS is also not refutable.

With the combined MTR-MTS assumption, from Manski and Pepper (2000), we have the following bound on the average treatment effect:

$$0 \leq E[y(t)] - E[y(s)]$$

$$\leq \sum_{t' \leq t} E(y|z = t')P(z = t') + E(y|z = t)P(z \leq t)$$

$$- \sum_{s' < s} E(y|z = s')P(z = s') - E(y|z = s)P(z \geq s).$$

A notable feature of the combined MTR-MTS assumption is its refutability; Manski and Pepper (2000) show that if $E(y|z = t)$ does not increase in $t$, then either MTR or MTS assumption is incorrect.

### 10.2 Application

Considering college major as treatment and the ordering of the major choice set as above, Assumption MTR can be motivated in the context of present section. We interpret Assumption MTR as that each student's wage function is weakly increasing in the order of perceived lucrative of major. Assumption MTR implies—for instance—as majoring in Economics results in higher expected earnings than Humanities, the expected wage for having majored in Economics is higher than that of having majored in Humanities.

We interpret Assumption MTS to be that students who self-select into a more lucrative major have weakly higher mean wage functions than those who do not. Here, Assumption MTS is unsupported by economic models and empirical findings. Yet, it may be supposed that those who self-select into lucrative majors (such as Economics and Kellogg Certificate combination) are those who place greater value on wealth acquisition and thus would be more willing to optimize their life choices to maximize earnings.

---

27 For each individual, we only observe a single point of $y_j(\cdot)$, namely, $y_j = y_j(z_j)$ where $z_j$ is the treatment that $j$ received.

28 The ordering of $T$ is derived from Table 2 ranking of expected earnings associated with each major.

29 Suppose $A$ is a humanities student who desires to lead an extravagant lifestyle. He would have liked to be accepted to the Kellogg Certificate Program as he believes that it would help him launch his consulting career. Suppose another humanities student $B$ who desires to pursue PhD in classics and who has no penchant for luxury. Then, without much difficulty, one can imagine, even when both were to major in Humanities, the student $A$ with desire for luxury would likely earn (weakly) more than the other all else equal.
10.2.1 Data

The survey elicits each student’s subjective probabilities for graduating major and expected earnings at age 30 for each major. To bound the average treatment effect, \( \Delta(s,t) \), we need to estimate the probabilities \( P(z) \) of realizing major \( z \) and the expectations \( E[y|z] \) of log-wage conditional on major \( z \). For \( P(z) \),

\[ z \in \{ \text{Engineering, Natural Sciences, Humanities, Arts, Drop-out} \} \]

we take the sample average of reported subjective probability \( P(z) \). When the subjective probabilities were elicited, the survey does not consider Economics, Economics/MMSS, and Economics/Kellogg as separate major categories. Thus, for Economics, Economics/MMSS, and Economics/Kellogg, we use the empirical distribution of currently pursued major to derive \( P(z) \). To estimate \( E[y|z] \), we similarly take the sample average of log-expected-earnings for each major. See Table 10 for the estimates of \( E[y|z] \) and \( P(z) \).

| Table 10: Mean log-expected-earnings and distribution of major |
|-----------------|----------------|----------------|
| \( z \)         | (1) \( E[y|z] \) | (2) \( P(z) \) | (3) Current major |
| Economics       | 3.343           | 0.136          | 12              |
| Engineering     | 3.403           | 0.262          | 27              |
| Natural Sciences| 3.171           | 0.198          | 16              |
| Humanities      | 2.886           | 0.206          | 16              |
| Arts            | 2.388           | 0.051          | 3               |
| Drop-out        | 1.822           | 0.035          | 0               |
| Econ/MMSS       | 3.307           | 0.136          | 12              |
| Econ/Kellogg    | 3.419           | 0.023          | 2               |
| Total           |                 |                | 88              |

Notes: Column (3) counts the number of students who declared \( z \) as primary major

Lastly, we derive asymptotically valid confidence intervals for the upper-bound by bootstrapping. The validity of such bootstrap approach is granted with the assumption that the unknown distribution \( P(y,z) \) equals the empirical distribution of the variables in the sample. We report the 0.95-quantile of the bootstrap sampling distribution.

10.2.2 Findings

We note from Table 10 estimates of \( E[y|z] \) and their 95-percent confidence band that there exists an everywhere increasing function that lie within the confidence band. Thus we cannot refute that the combined MTR-MTS assumption is incorrect.

\[ ^{30} \text{Of 106 students surveyed, 18 pursue Economics/MMSS/Kellogg triple-major combination, on which we do not have } E[y|z] \text{ data. To divide the sample into disjoint major categories, we were thus forced to drop the eighteen observations. Using two different methods to estimate } P(z) \text{ may be problematic. Yet, the sample average of the reported subjective probability for Economics is 0.286 and the probabilities for Economics, Economics/MMSS, Economics/Kellogg derived from empirical distribution are 0.13, 0.13 and 0.02 respectively (which sum to 0.262).} \]

\[ ^{31} \text{We perform 50,000 bootstrap replications.} \]
Table 11 reports the estimates of average treatment effect $\Delta(s,t)$. We compare the returns to Economics/MMSS combination with the returns to other major categories. It may be useful to compare the reported estimates with other estimates of returns to schooling. Card (1993) reports a point estimate of 0.132 for an additional year of schooling. Manski and Pepper (2000) reports a point estimate of 0.397 as the upper-bound for mean log-wage increase for completion of a four-year college relative to completion of high school only. Thus, we observe from Table 11 that the average treatment effect of major choice is indeed significant.

<table>
<thead>
<tr>
<th>s</th>
<th>t</th>
<th>(1) Estimate</th>
<th>(2) Bootstrap 0.95-quantile$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Econ/MMSS</td>
<td>Econ/Kellogg</td>
<td>0.326</td>
<td>0.430</td>
</tr>
<tr>
<td>Econ/MMSS</td>
<td>Engineering</td>
<td>0.310</td>
<td>0.422</td>
</tr>
<tr>
<td>Economics</td>
<td>Econ/MMSS</td>
<td>0.221</td>
<td>0.349</td>
</tr>
<tr>
<td>Natural Sciences</td>
<td>Econ/MMSS</td>
<td>0.316</td>
<td>0.458</td>
</tr>
<tr>
<td>Humanities</td>
<td>Econ/MMSS</td>
<td>0.532</td>
<td>0.728</td>
</tr>
<tr>
<td>Arts</td>
<td>Econ/MMSS</td>
<td>1.011</td>
<td>1.237</td>
</tr>
<tr>
<td>Drop-out</td>
<td>Econ/MMSS</td>
<td>1.583</td>
<td>1.806</td>
</tr>
</tbody>
</table>

Notes: $^a$Bootstrap 0.95-quantiles (normal-based) obtained by performing 50,000 replications.

11 Conclusion

This study investigates determinants of college major choice. In so doing, we utilize a panel data of subjective expectations obtained from Northwestern students. With the novel data, we estimate a choice model that robustly identifies each utility component. From the main choice model, we find that major-specific earnings expectation indeed plays a significant role in major choice—especially for underclassmen. The model suggests also perceived ability influences major choice. Despite the strong correlation between perceived ability and major choice likelihood, given the limited data, it is speculative to conclude that the perceived ability is a deciding factor determining the choice of major. Moreover, we find that students revise their beliefs in a rational manner, incorporating newly gained information into their belief system so that their expectations become, in general, more realistic.$^{32}$

This study makes the following contributions to the literature on schooling behavior. First, this study adds to the growing literature that uses subjective expectations to shed light on schooling choices. Indeed, the inferential problem as faced by college students in determining their major is analogous to that which is faced by econometrician. And, given the nature of problem, consideration of subjective expectation is indispensable. This study confirms advantage of the subjective probability model by explicit comparison to the classical

$^{32}$We observe, in the case of the salary-treatment group, that the absolute deviations from the realized salary-outcomes become less severe in post-treatment expectations
choice model. Second, by adopting the single-survey panel study design, we identify the model parameters robustly, resulting in more conservative estimates of the choice elasticities with respect to expected earnings. Using only readily available information, we perturb ability and earnings expectations. Third, we confirm our conclusion by semiparametric estimation, which is robust under weaker assumptions. Fourth, we consider gender, racial, and socioeconomic heterogeneity in major choice and provide an evidence against exiting literature. Fifth, the panel data original to this study not only allows us to identify the model but also gives insight into how Northwestern students perceive programs particular to Northwestern, such as the Kellogg Certificate Program and MMSS. The data shows that earning a second major in MMSS or a certificate from Kellogg results in a higher earnings expectation. Yet, it is difficult to analyze the trade-offs associated with earning the Kellogg certificate/MMSS additionally as, albeit students expect higher future earnings, they also expect to perform worse if they were to participate. Lastly, we provide upper-bounds to the returns to Economics/MMSS major combination relative to alternative majors; in particular, we estimate the upper-bound for mean log-wage increase for majoring in Economics/MMSS relative to Humanities to be 0.532.

It is important to note that the major choice behavior of Northwestern students is largely heterogeneous and the unobserved taste component plays a substantial role in major choice. Although this study is an attempt to learn the extent of influence that the taste component has on students decision of major, it falls short of providing a definitive answer. Furthermore, it can be critiqued that the information treatment setting is artificial and that the effects of information treatment would not persist over time. To address such concerns, a compilation of larger data set with higher frequency is necessary. This study does not posit that the amount of data collected is even remotely close to address such critiques. Yet, I believe with a better data set—perhaps complemented with the revealed choice data—can allow us to form and confirm a more vivid understanding of major choice behavior of Northwestern students.

Albeit imperfect, the findings of this study have policy implications. Having found that major choice becomes more inelastic overtime, it may be advisable to provide freshmen with opportunities to explore both in terms of post-graduation career paths and in terms of courses. It may be crucial for students to learn about their natural academic propensity early on through short-term low-intensity survey courses so that they may be introduced to core concepts of various disciplines. And, for the departments having trouble recruiting new students, it may be recommended that they place their most popular professors to intro-level classes. The departments may even consider restructuring introductory classes so as to entice and cater to incoming students.

This study investigates a choice that is perhaps un-economic in nature with economic methodology. From introspection, I concede that such economic analysis perhaps fails to do

---

33 Indeed, the data shows that the students expectation of academic performance is lower for Economics and MMSS combination than for Economics alone. The result is included in Appendix.

34 A game theoretic model of major choice—which would model such trade-off—suggests an avenue for future study.

35 Low-intensity (i.e. generous grading) is recommended because students are responsive to grade feedback.
justice to the intricate problem of major choice where the youth balance among numerous factors of life. Yet, acceding to vagaries of youth, perhaps this study is illuminating on a bigger picture painted by each student over hours of rumination and dreaming.
12 References

## Appendix A

Table A-1: Major choice likelihood and expected earnings  
(sample restricted to salary treatment group)

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Log-odds revision (post-pre)</th>
<th>p-value</th>
<th>OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fixed-effect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel A: Full sample</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log self-earnings revision</td>
<td>-0.0124</td>
<td>0.846</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0634)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log perceived-ability revision</td>
<td>-0.0637</td>
<td>0.396</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0783)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.444</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total observations</td>
<td>94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of respondents</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel B: Underclassmen (Freshmen and sophomore)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log self-earnings revision</td>
<td>-0.0863</td>
<td>0.253</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0701)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log perceived-ability revision</td>
<td>-0.228</td>
<td>0.449</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.286)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.770</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total observations</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of respondents</td>
<td>9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Individual covariates include an indicator for gender; indicators for race; GPA indicators (0-3.0, 3.0-3.5, 3.5-3.8, 3.8-4.0); SAT math/verbal scores; indicators for parents’ education level (BA or more); indicator for non-reported SAT; indicators for socioeconomic status; major dummies; standard errors in parenthesis. ***,**, * indicate significance at 1%, 5%, 10% respectively; heteroskedastic cluster robust standard error in parentheses; SE were adjusted for clustering at the individual level.


Appendix B

<table>
<thead>
<tr>
<th>Table A-2: Academic performance expectation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Econ</td>
</tr>
<tr>
<td>Econ+Kellogg</td>
</tr>
<tr>
<td>Econ+MMSS</td>
</tr>
<tr>
<td>Total observations</td>
</tr>
</tbody>
</table>

Notes: Only pre-treatment observations were considered
Appendix C

Table A-3: Densities of expected earnings of Economics

Notes: Observations with $500,000 expectation were dropped; pre-treatment expected earnings in blue post-treatment expected earnings in red.
Appendix D

Table A-4: Densities of expected earnings across majors

Notes: Observations with $500,000 expectation were dropped; pre-treatment expected earnings considered only.
Belief Survey v.6

Q1 <div>Thank you for participating in this survey. This survey regards <strong>your beliefs</strong>, not the accuracy of information that can be gathered from other sources. Please do not consult any outside references or discuss these questions with any other people. <br><br>This survey will last approximately 10 minutes. Please finish the survey in one sitting (your survey responses will be invalidated after 30 minutes). The contact information for financial compensation will be requested in a separate survey to which you will be automatically redirected (the responses you give in this survey will have no association with your contact information).<br><br>A portion of this survey is derived from Wiswall and Zafar (2014), "Determinants of College Major Choice."<br></div><br>

Q2 <strong>This section will ask you about your background.</strong>

Q3 Are you a student at Northwestern University?
- Yes (1)
- No (2)

Q4 What is your current year of school in the academic year 2016-2017?
- 1st year (Freshman) (1)
- 2nd year (Sophomore) (2)
- 3rd year (Junior) (3)
- 4th year (Senior) (4)
- 5th year or more (5)

Q5 Do you know about the adjunct major in Mathematical Methods in the Social Sciences (MMSS)?
- Yes (1)
- No (2)

Display This Question:
- If Do you know about the adjunct major in MMSS? Yes Is Selected

Q7 Are you pursuing the adjunct major in MMSS?
- Yes (1)
- No (2)

Q8 Do you know about the Kellogg Certificate Program?
- Yes (1)
- No (2)
Q9 Are you pursuing the Kellogg Certificate Program?

- Yes (1)
- No (2)

Q10 What is your current or intended major(s)? If you have more than two, list two (primary and secondary).

Q11 Major 1 (Primary Major):

Q12 Major 2 (Secondary Major): (NA if you are pursuing only one major)

Q13 What is your career plan after graduation?

Q16 Questions About Your Beliefs

In this section, you will be asked about the percent chance of something happening. The percent chance is a number between 0 and 100. 2 or 5% indicates "almost no chance." 19% or so indicates "not much chance." 47 or 55% indicates "about even chance." 82% or so indicates a "very good chance." 95 or 98% indicates "almost certain." If you are not sure about the notion of percent chance read the following example: Suppose you are playing baseball. When asked about the percent chance that your team will win, you answer 70. This means that you believe that your team would win 70 out of 100 games on average.

Q17 Now, the following questions ask you about your beliefs about yourself.

Q18 What do you believe is the percent chance that you would either graduate from Northwestern with a major in the following major categories (considering your PRIMARY MAJOR) or that you would never graduate/drop-out?

- Economics (1)
- Engineering/Computer Science (2)
- Humanities/Other Social Sciences/Communications/Journalism (3)
- Natural Sciences/Mathematics (4)
- Music/Visual/Performing Arts (5)
- Drop-out/Never Graduate (6)
Q19 Consider a situation in which you will graduate with a Bachelor’s degree in each of the following major categories. Think about other individuals at Northwestern who will graduate with a major in each category. In PERCENTILE, in terms of your graduating GPA, what would be your rank in each major category compared to other students at Northwestern who will graduate with a major in each category? Note: 99 percentile would mean your GPA would be in the TOP 1%; correspondingly, 1 percentile would mean that your GPA would be in the BOTTOM 1% (or equivalently 99% of the people would have higher GPA than you).

_____ Economics (1)
_____ Engineering/Computer Science (2)
_____ Humanities/Other Social Sciences/Communications/Journalism (3)
_____ Natural Sciences/Mathematics (4)
_____ Music/Visual/Performing Arts (5)

Q20 Consider a situation in which you will graduate with the following major/certificate combination. Think about other individuals at Northwestern who will graduate with the following major/certificate combination.

In PERCENTILE, where do you think you would rank in terms of your graduating GPA when compared to them?

_____ Economics (1)
_____ Economics + MMSS (2)
_____ Economics + Kellogg Certificate Program (3)

Q21 Now, look forward to when you are 30 years old and think about the kinds of jobs that will be available to you if you were to graduate in each of the following major categories.

Q26 When you are 30 YEARS OLD, if you are working FULL-TIME and you received a Bachelor’s in each of the following major categories, what do you believe is the percent chance that you would earn:

At least $35,000 per year
At least $85,000 per year

Note: (i) Your percent chance of earning at least $35,000 per year should be at least as large as the answer for percent chance of earning at least $85,000 per year. (ii) If you believe plan to get an advanced degree after Bachelor’s, assume you would have completed the degree by age 30. (iii) Ignore the effects of price inflation on earnings, that is assume that $1 at your age 30 would worth the same as $1 to you now.

Note:
(i) Your percent chance of earning at least $35,000 per year should be at least as large as the answer for percent chance of earning at least $85,000 per year. (ii) If you believe plan to get an advanced degree after Bachelor’s, assume you would have completed the degree by age 30. (iii) Ignore the effects of price inflation on earnings, that is assume that $1 at your age 30 would worth the same as $1 to you now.
Q23 <div>When you are 30 YEARS OLD and you are working FULL-TIME, what do you believe is the percent chance that you would earn at least $35,000 per year?</div>

- Economics (1)
- Engineering/Computer Science (2)
- Humanities/Other Social Sciences/Communications/Journalism (3)
- Natural Sciences/Mathematics (4)
- Music/Visual/Performing Arts (5)
- Drop-out/ Never Graduate (6)
- Economics + MMSS (7)
- Economics + Kellogg Certificate Program (8)

Q28 <div>When you are 30 YEARS OLD and you are working FULL TIME, what do you believe is the percent chance that you would earn at least $85,000 per year?</div>  

IMPORTANT: Recall, the percent chance of earning at least $85,000 should be LESS than the percent chance of earning at least $35,000.

- Economics (1)
- Engineering/Computer Science (2)
- Humanities/Other Social Sciences/Communications/Journalism (3)
- Natural Sciences/Mathematics (4)
- Music/Visual/Performing Arts (5)
- Drop-out/ Never Graduate (6)
- Economics + MMSS (7)
- Economics + Kellogg Certificate Program (8)

Q29 <em>In this part of the survey, we will provide you with some information based on actual data about employment and earnings of Northwestern graduates.</em><br />

Q31 The following information is from an annual study by Northwestern Career Advancement regarding employment outcome of the Northwestern undergraduate class of 2015:
Q61 Consider the above information carefully (next button will pop up momentarily).

Q71 Timing
   First Click (1)
   Last Click (2)
   Page Submit (3)
   Click Count (4)

Q69 In this section, you will be asked to consider a hypothetical situation in which you take a course you have not taken and receive a grade feedback.

Q62 Have you taken ECON 201: Introduction to Microeconomics or equivalent?
   ✗ Yes (1)
   ✗ No (2)

Condition: No Is Selected. Skip To: Suppose you received an A in ECON 201....
<table>
<thead>
<tr>
<th>Question</th>
<th>No (1)</th>
<th>Yes (2)</th>
<th>Condition</th>
<th>Skip To:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q64 Have you taken CHEM 210-1: Organic Chemistry?</td>
<td></td>
<td></td>
<td>No Is Selected. Skip To: Suppose you received an A in CHEM 210...</td>
<td></td>
</tr>
<tr>
<td>Q65 Have you taken EECS 111: Fundamentals of Programming I?</td>
<td></td>
<td></td>
<td>Yes (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No Is Selected. Skip To: Suppose you received an A in EECS 211...</td>
<td></td>
</tr>
<tr>
<td>Q63 Have you taken JOUR 301-0: Journalism in Practice?</td>
<td></td>
<td></td>
<td>No Is Selected. Skip To: Suppose you received an A in JOUR 301...</td>
<td></td>
</tr>
<tr>
<td>Q60 Have you taken ECON 310-1: Intermediate Microeconomics or equivalent?</td>
<td></td>
<td></td>
<td>Yes (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No Is Selected. Skip To: Suppose you received an A in ECON 310...</td>
<td></td>
</tr>
<tr>
<td>Q62 Have you taken EECS 214: Data Structures?</td>
<td></td>
<td></td>
<td>No Is Selected. Skip To: Suppose you received an A in EECS 214...</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Yes (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No Is Selected. Skip To: Suppose you received an A in EECS 214...</td>
<td></td>
</tr>
<tr>
<td>Q63 Suppose you received B- in ECON 201: Introduction to Microeconomics or course equivalent.</td>
<td></td>
<td></td>
<td>No Is Selected. Skip To: Suppose you received an A in ECON 310...</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Yes (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No Is Selected. Skip To: Suppose you received an A in EECS 214...</td>
<td></td>
</tr>
<tr>
<td>Q65 Suppose you received B- in ECON 310-1: Intermediate Microeconomics.</td>
<td></td>
<td></td>
<td>No Is Selected. Skip To: Suppose you received an A in ECON 310...</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Yes (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No Is Selected. Skip To: Suppose you received an A in EECS 214...</td>
<td></td>
</tr>
<tr>
<td>Q64 Suppose you received an B- in EECS 111: Fundamentals of Programming I</td>
<td></td>
<td></td>
<td>No Is Selected. Skip To: Suppose you received an A in EECS 214...</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Yes (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No Is Selected. Skip To: Suppose you received an A in EECS 214...</td>
<td></td>
</tr>
<tr>
<td>Q67 Suppose you received B- in EECS 214: Data Structures.</td>
<td></td>
<td></td>
<td>No Is Selected. Skip To: Suppose you received an A in EECS 214...</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Yes (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No Is Selected. Skip To: Suppose you received an A in EECS 214...</td>
<td></td>
</tr>
</tbody>
</table>
Q69 <strong>Suppose you received B- in JOUR 301-0: Journalism in Practice.</strong> Condition: Suppose you received B- in ... Is Displayed. Skip To: Take a moment to think carefully about....

Q70 <strong>Suppose you received B- in CHEM 210-1: Organic Chemistry.</strong> Condition: Suppose you received B- in ... Is Displayed. Skip To: Take a moment to think carefully about....

Q68 Take a moment to think carefully about the hypothetical (e.g. Would you be sad? Terrified? Indifferent? Happy? Discouraged? Encouraged?) (Next button will pop up in ten seconds)

Q72 Timing
   First Click (1)
   Last Click (2)
   Page Submit (3)
   Click Count (4)
Q33 <strong>In this section, you will be re-asked about some of the questions.</strong>

Q35 <p>What do you believe is the percent chance that you would either graduate from Northwestern with a major in the following major categories (considering your PRIMARY MAJOR) or that you would never graduate/drop-out?</p>

- Economics (1)
- Engineering/Computer Science (2)
- Humanities/Other Social Sciences/Communications/Journalism (3)
- Natural Sciences/Mathematics (4)
- Music/Visual/Performing Arts (5)
- Drop-out/Never Graduate (6)

Q36 Consider a situation in which you will graduate with a Bachelor’s degree in each of the following major categories. Think about other individuals at Northwestern who will graduate with a major in each category. In PERCENTILE, in terms of your graduating GPA, what would be your rank in each major category compared to other students at Northwestern who will graduate with a major in each category? Note: 99 percentile would mean your GPA would be in the TOP 1%; correspondingly, 1 percentile would mean that your GPA would be in the BOTTOM 1% (or equivalently 99% of the people would have higher GPA than you).

- Economics (1)
- Engineering/Computer Science (2)
- Humanities/Other Social Sciences/Communications/Journalism (3)
- Natural Sciences/Mathematics (4)
- Music/Visual/Performing Arts (5)

Q37 Consider a situation in which you will graduate with the following major/certificate combination. Think about other individuals at Northwestern who will graduate with the following major/certificate combination. Where do you think you would rank in terms of your graduating GPA when compared to them?

- Economics (1)
- Economics + MMSS (2)
- Economics + Kellogg Certificate Program (3)

Q38 Now, look forward to when you are 30 years old and think about the kinds of jobs that will be available to you if you were to graduate in each of the following major categories.
Q40 When you are 30 YEARS OLD and if you are working FULL-TIME, what do you believe is the percent chance that you would earn at least $35,000 per year?

- Economics (1)
- Engineering/Computer Science (2)
- Humanities/Other Social Sciences/Communications/Journalism (3)
- Natural Sciences/Mathematics (4)
- Music/Visual/Performing Arts (5)
- Drop-out/Never Graduate (6)
- Economics + MMSS (7)
- Economics + Kellogg Certificate Program (8)

Q41 When you are 30 YEARS OLD and if you are working FULL-TIME, what do you believe is the percent chance that you would earn at least $85,000 per year?

IMPORTANT: Recall, the percent chance of earning at least $85,000 should be LESS than the percent chance of earning at least $35,000.

- Economics (1)
- Engineering/Computer Science (2)
- Humanities/Other Social Sciences/Communications/Journalism (3)
- Natural Sciences/Mathematics (4)
- Music/Visual/Performing Arts (5)
- Drop-out/Never Graduate (6)
- Economics + MMSS (7)
- Economics + Kellogg Certificate Program (8)

Q42 This section includes demographic questions.

Q43 What is your gender?
- Male (1)
- Female (2)
- Other (3) ________________

Q44 If you were to be married by the age 30, what would be the likely gender of your future spouse?
- Male (1)
- Female (2)
- Other (3) ________________
Q45 You consider yourself:
- White/Caucasian (1)
- Black/African American (2)
- American Indian (3)
- Hispanic/Latino (4)
- Asian/Pacific Islander (5)
- Other (6) ____________________

Q46 Which year were you born in?

Q47 What is your country of birth?

Q48 What is your mother's country of birth?

Q49 What is your father's country of birth?

Q50 What is the highest level of schooling completed by your mother?
- High school or below (1)
- College (2)
- Graduate school (3)

Display This Question:
If What is the highest level of schooling completed by your mother? College Is Selected
Or What is the highest level of schooling completed by your mother? Graduate school Is Selected

Q51 What was her major in college?

Q53 What is the highest level of schooling completed by your father?
- High school or below (1)
- College (2)
- Graduate school (3)

Display This Question:
If What is the highest level of schooling completed by your father? College Is Selected
Or What is the highest level of schooling completed by your father? Graduate school Is Selected

Q54 What was his major in college?
Q52 What is the best estimate of your parents' socioeconomic status?
- Upper class (1)
- Upper-middle class (2)
- Middle class (3)
- Lower-middle class (4)
- Lower class (5)

Q55 Which standardized test have you taken?
- SAT (1)
- ACT (2)

Q56 What were your scores on SAT I?
- SAT Verbal (1)
- SAT Math (2)

Q57 What was your composite score on the ACT?
- ACT Composite Score (1)

Q58 In which range bracket, do you place your current college GPA?
- 0-2.49 (1)
- 2.5-2.99 (2)
- 3.0-3.49 (3)
- 3.5-3.79 (4)
- 3.8-4.0 (5)

Q59 Thank you for completing the survey. We appreciate your responses.

If you would like to be considered for financial compensation lottery, please provide your email address in a separate survey to which you will be automatically redirected; this allows the information you provided here to be disaggregated from your contact information.

If there is anything about the study that is unclear or that you do not understand, if you have questions or wish to report a research-related problem, you may contact Henry Haksoo Lee at haklee2015@u.northwestern.edu. For questions about your rights as a research participant, you may contact the Office for Research, Institutional Review Board, (847) 467-1723, Social and Behavioral Research, 600 Foster St. Evanston, IL 60208.