Forming Wage Expectations through Learning: Evidence from College Major Choice*

Xiaoyu Xia

December 8, 2014

Abstract

How do college students choose their major and what role does family play? I use data from two major longitudinal surveys to develop and estimate a model in which students learn about earning opportunities associated with different majors through earning realizations of older siblings and parents. Reduced-form models show that the probability of choosing a major that corresponds to the occupation of an older sibling or parent is strongly affected by whether the family member is experiencing a positive or negative earnings change at the time the major choice is made. Building on this finding, I estimate a model of major choice that incorporates learning from family-based information sources. The results confirm that students use family members' earnings experience to form own expectations about the rate of return to a major choice, however they overestimate the predictive power of family members' earnings.

JEL Code: D84, I21, J24, J31.

Keywords: Wage Expectation, Major Choice, Intergenerational Transmission.

*I thank David Card for invaluable advice and support. For insightful suggestions, I thank Stefano DellaVigna, Enrico Moretti, Matthew Rabin, as well as Patrick Kline, Christopher Pissarides, Jesse Rothstein and Christopher Walters. I also thank Willa Friedman, Tristan Gagnon-Bartsch, Albrecht Glitz, Sheng Li, Andrej Milivojevic, Carl Nadler, Ana Rocca, Jin Yan and various seminar participants at University of California, Berkeley, Chinese University of Hong Kong and Lingnan University for their help. All errors are my own.

1 Introduction

Young people often have to make important decisions — such as which college to attend or which job to take — in the presence of substantial uncertainty about the future consequences of their choices. How do they form expectations about the payoffs to different options? A long line of social science research, dating at least from Hyman (1942), has argued that people learn from the choices and outcomes of "reference groups."[1] Perhaps the most important reference group for many decisions is the family.[2] There are strong family correlations in many socioeconomic outcomes (see Black and Devereux, 2011 for a recent survey), including occupational status. However, whether this correlation arises though learning opportunities or through other potential channels, such as correlated abilities or tastes, remains unclear.

In this paper I use data from two major longitudinal surveys to directly examine the role of learning from the labor market experiences of close family members (older siblings and parents) in the choice of college major. Choice of major is an important first step in the careers of many college students and is a significant determinant of their subsequent earnings and occupational status (Arcidiacono, 2004; Hamermesh and Donald, 2008; Altonji *et al.*, 2012). According to human capital theory (Becker 1962), young people choose their major and future career by calculating the internal rate of return in each occupation. While inherited abilities or tastes are key components in such calculation, the focus of this paper is testing whether the information based on the wage realizations of close family members exerts an additional effect on the choice of major.

Simple reduced-form models show that college students are more likely to choose a major associated with the occupation of a family member who is experiencing a wage increase, after controlling for family member's long-term average wage and the occupational wage trend. This result suggests that even short-run wage fluctuations affect students' major choices. Estimates from a nested multinomial logit model confirm that a student places significant weight on wage information based on the contemporaneous wage of her family member. Adding this a family-based information component leads to an improvement in model fit relative to a benchmark "rational expectations" model used in previous studies.[3]

[1] A recent example in economics field is the survey in Dominican Republic conducted by Jensen (2010). His data shows that students' main source of information about earnings were the people they knew in their community.

[2] Studies across social science fields have documented the influence of family members on youth behavior in a broad range, such as Weast (1956), Bank *et al.* (1990), Brody (1998) and Duncan *et al.* (2001).

[3] Imposing the assumption that students have rational expectations that equal to the realized

Although there could be many competing explanations to above results, two direct ones are probably family-correlated ability and that students can obtain help from family members who work in relevant occupations. To rule out these explanations, I examine the correlations of students' post-college earnings with the family wage outcomes observed at the time of their initial decision for a college major. The weight students placed on family wage realizations when choosing a major is much larger than can be justified by the empirical correlation between their own earnings and their family members' earnings. Moreover, students who choose a major corresponding to a positive wage change of family members perform worse in the labor market later than others whose family members received a negative wage change. These results appear consistent with the hypothesis that students overestimate the predictive power of family members' earnings but not with the hypothesis that students choose a major based on family-correlated ability or direct help from family ties.

My main data source is the National Longitudinal Survey of Youth 1979 (NLSY79). It includes a relatively large number of sibling pairs, allowing me to link the college major choice of the younger sibling to the earnings of the older sibling. I use an additional data set, the National Education Longitudinal Study of 1988 (NELS88), to link major choice to changes in parental income.

I begin by establishing the importance of family based labor market information flows using a reduced-form model of the probability that a student adopts a major choice corresponding to the occupation of her older sibling or parent. I find that a student is more likely to pursue a major that is related to the occupation of her family member when the family member earns a higher than occupation-average wage, and especially, when the family member experiences a wage increase during a student's time in high school or college. There is no such correlation between a student's major choice and the wage changes experienced by her siblings or parents in later years. The marginal elasticity of a student's major choice with respect to a family member's wage change is 0.77, which is about seven times larger than the elasticity with respect to average earnings in the occupation estimated by previous literature (e.g., Blom, 2012, Zafar and Wiswall, 2012, Beffy *et al.*, 2012). The comparison suggests that earlier studies may have substantially under-estimated the sensitivity of student major choice to perceived earnings opportunities.

To understand the role of family information in a student's career planning, I go on to develop a nested multinomial logit model of major choice in which students

occupation average wage conditional on observables is a common approach taken by previous major-choice studies, for example, Siow (1984), Berger (1988), Keane and Wolpin (1997), Rosen and Ryoo (2004).

update their expected earnings in response to their family members' wage outcomes. This choice model also generates an empirical prediction that can test whether students are using the optimal weight when updating expectations. I assume a student believes that the future wages associated with a given major can be decomposed into two components: a predictable wage component that is known before she starts working and an unknown wage component. To form her belief about the unknown component, the student relies on the realized wage of her family member, who works in an occupation related to the major under consideration.

The effect of family wage outcomes on students' major choices therefore identifies to what extent students learn from family-based information. My point estimate for the weight students place on family member's wage when they update their expectations of future earnings is 0.39. However, by examining the empirical wage correlation between siblings in the NLSY79, I find the upper bound of the empirical correlation is between 0 and 0.1, much lower than my point estimate of students' perceived correlation. This comparison suggests that students overestimate the correlation of wage outcomes between close family members. Consistent with this result, I also find that students who choose a major corresponding to the occupation of a family member who received a positive wage signal at the time of the choice suffer a greater incidence of changing college majors and have lower earnings in later years, indicative of the lower match quality that would be expected if students are overestimating the informativeness of the wage realizations of close family members.

This paper calls attention to an information channel for how family background affects students' educational decisions. If students rely on parental earnings to learn about the returns to education or training, this channel can provide one causal mechanism that explains intergenerational persistence in education attainment and occupational choice.[4] It also connects this paper with recent literature on how family background affects college attendance or female laborforce participation, and understanding this information channel has direct policy implications.[5] If students from disadvantageous families not only suffer from credit constraint but also from insufficient information in human capital investment, utilizing the precise wage information that available from school career centers or public agencies becomes extremely important for them to make informed decisions.

This paper provides a testable learning mechanism that sheds new light on

[4]For example, studies by Hellerstein and Morrill (2011), Corak and Piraino (2011), show that a large portion of children in recent cohorts work in the same occupation as their parents.

[5]Cameron & Heckman (2001) find that parental income has its greatest influence on their children's college attendance by enhancing the abilities and attitudes required for entering college rather than through actual financing. In addition, a recent study by Fernandez (2013) emphasizes the intergenerational learning process in explaining the trend of female labor participation rate.

how young people form their expectations. Motivated by earlier thinking on how students forecast their potential earnings (Freeman, 1975b, 1975a, 1976; Willis and Rosen, 1979; Manski and Wise, 1983), my model incorporates both the conventional rational expectation and the new within-family "cob-web" expectation. This generalized model predicts students' major choices more accurately, and it can provide a micro-foundation for heterogeneity in expectations. While there is a perennial debate on expectations assumptions in economic thinking, the findings in this paper favor models of individual-specific adaptive expectations in decision-making.

The rest of the paper is laid out as follows. Section 2 provides the motivating evidence that family members' wage changes affect students' career planning. Section 3 develops and estimates a nested multinomial logit model in which students update beliefs about future earnings based on family member's wage changes. Section 4 shows the correlation between a student's labor market performance and her family member's wage shocks. Section 5 concludes.

2 Motivating Evidence

This section establishes the importance of family-based information using a reduced-form model of the probability that a student adopts a major choice corresponding to the occupation of their older sibling or parent. It provides motivating evidence that family members' wage changes affect students' career planning. Section 2.1 introduces the data sources. Section 2.2 presents the evidence that a student is more likely to choose a college major related to her older sibling's or parent's occupation if the family member has received a recent wage increase. Moreover, a student reports that her "ideal" occupation is the occupation that her older sibling works in more often when the latter has experienced a recent wage increase.

2.1 Data: Linking Students with Family Members

My main data source is the National Longitudinal Survey of Youth 1979 (NLSY79), which connects students' educational choices with their older siblings' wages. The Bureau of Labor Statistics has collected the NLSY79 since 1979 using a sample of 12,686 men and women born between 1957 and 1964. This survey first interviewed all individuals aged between 15 and 22 in a household in 1979, and then follows them with annual interviews until 1994, and continues on a bi-annual basis.

The NLSY79 surveys the baby-boom cohort, and therefore a large faction of families in the survey has records for multiple siblings. There are 3,448 sibling pairs in the NLSY79. Older siblings are defined as siblings entering the labor market first. After dropping sibling pairs in which younger siblings have not attended a college or declared a college major, 1,639 siblings pairs remain for use in this study.[6] The survey includes the 1970 census 3-digit occupation code to record the occupation of all respondents. I group all the professional occupations (1970 Census occupation codes 001-245) into 23 categories that can be directly mapped into college majors listed in the NLSY79.[7] In my analysis, the occupation variable for an older sibling is thus defined as the first full-time professional occupation during 1979-1992.[8] By excluding the students with older siblings who do not have records of working in a professional occupation, I construct a Student-Sibling sample (S1) with 1,004 sibling pairs. Table 1 Panel A summarizes the mean and standard deviations of key variables in the NLSY79 and my S1 sample.

The definition of the time period of interest is crucial for my analysis. Students in the NLSY79 declare their first college major in a certain year between 1979 and 1992. As the focus is on the effect of family wage outcomes on a student's major choice, the time window of interest is a few years before and after a student declares her major. In the S1 sample, around 25% of the older siblings' earnings records are missing in any given year. The issue of missing data in combination with the concern that the NLSY79 switched to a bi-annual survey after 1994 explains why I use a 4-year time window as the relevant time period. For example, if a student declares her first major in year 1985, the pre-choice time window is 1982-1985 and the post-choice one is 1986-1989. In this way, I can construct a balanced average wage variable in the pre-choice time window and the post-choice window for most students in my sample.[9] In particular, the average wage in pre-choice window is referred to as the "contemporaneous wage" of a family member.[10]

[6] An alternative way to link students and their family member is to match them by their college majors. Given the purpose of this study is to see how students learn from family member's labor market experience, I believe the wage outcome of a close family member tells a students more about the earnings associated with the family member's occupation rather than the family member's choice of college major.

[7] Table A1 lists the 23 majors in the NLSY79, and Table A3 maps each professional occupations to an associated college major. Which major an occupation matches to is determined by the college major held by the majority of college educated workers in that occupation. I also examine the major-occupation match-matrix from American Community Survey (ACS) 2009. It has very similar pattern to the NLSY79.

[8] All students in the NLSY79 have declared their first college major during 1979-1992.

[9] The regression results are robust to changing time window length to 3-year span or 5-year span. Yet, the number of observations would decrease significantly if the length of the time window changes to 2-year span.

[10] There are three reasons why I focus on the wage in pre-choice window. First, this time period is likely to be the critical learning time for students to form wage expectations. Second, the wage records in this time period are available for most students, while earlier wage records are incomplete. Third, around 90% siblings in S1 have a age difference smaller than 4-year, therefore

The key variables for my reduced-form models are a student's first declared major and her older sibling's wage at the time the student was making her decision.[11] A student's first major choice is the main dependent variable. A sibling's wage is measured in log hourly rate that is normalized to 2010 dollars. I use the hourly wage rate because it tells a student more about the net payoff for a certain occupation compared to the annual earnings.[12] By calculating a sibling's average log wage in the pre-choice time window and the post-choice one, I construct the main explanatory variables as "Pre-Choice Sibling's Wage" and "Post-Choice Sibling's Wage".

I use the National Education Longitudinal Study of 1988 (NELS88) as an additional data set. It includes a nationally representative sample of eighth graders first surveyed in 1988 then re-surveyed through four follow-ups in 1990, 1992, 1994, and 2000. There are around 20,000 students who have completed all follow-up surveys. Among those, 7,299 students had declared a college major by 1992 and have records of their parents' occupations. The NELS88 provides an opportunity to link students' education choices with their parents' labor market experiences. However, the NELS88 codes a parent's occupation in an aggregated way that there were only 16 different occupation categories.[13] Among the 16 occupations, the only professional occupations that can be directly mapped to a college major are those of a manager and a school teacher. Thus I use 1,093 students whose parent works as a manager to form the first Student-Parent sample (S2-Manager), and 705 students with a parent who works as a school teacher to form the second Student-Parent sample (S2-Teacher). The time of period of interest in the NELS88 is the time between 1988 and 1992, when students were attending high school. I construct the change of family income between 1988 and 1992 as the proxy for the change in parental income. Table 1 Panel B lists the summary statistics for the NELS88 and my S2 sample.

According to Table 1, students in sample S1 and S2 are similar to the population of college students in the NLSY79 and the NELS88. There is slightly positive selection based on AFQT scores or after-college wages in my sample, which is likely because all students in sample S1 and S2 have at least an older sibling or a parent working in a professional occupation. Panel A of Table 1 also shows that older

the contemporaneous wage captures most wage information students received from their older siblings.

[11] Approximately 30% of students have changed their majors during college years, but family backgrounds may affect their initial college major most.

[12] In robustness checks, I show the regression results in the same specification but with a worker's wage measured by annual income.

[13] The 16 categories are: Clerical, Craftsperson, Farmer, Homemaker, Laborer, Manager, Military, Operative, Account/Artist/Nurse, Dentist/Lawyer, Proprietor, Protective Service, Sales, School Teacher, Service, Technical.

siblings have lower AFQT scores and earn lower wages compared to their younger siblings. This is because in my sample the older siblings might not have attended any college, while all younger siblings have received at least some college education.

2.2 Major Choice and Contemporaneous Family Wages

I begin my analysis by examining the determinants of major choices in a descriptive way. Specifically, I estimate the correlation between a student's major choice and her older sibling's wages. The outcome variable is a binary indicator for whether a student's college major matches her sibling's occupation. One important control variable is an older sibling's permanent wage, which is the average log wage from 1979 to 1992. I use this permanent wage to capture the base level of a sibling's earning. By taking the difference between a sibling's pre-choice wage and a siblings's permanent wage, I construct a wage shock variable "Sibling's Wage Pre-Choice - Permanent". Similarly, I construct a post-choice wage shock variable "Sibling's Wage Post-Choice - Permanent". Other control variables include occupation average wage in pre-choice and post-choice window, a student's AFQT score, demographic characteristics, and a student's pre-determined taste for certain occupations, captured by a variable that records a student's ideal occupation in 1979.[14]

Table 2 shows that the wage change an older sibling received in the pre-choice window strongly correlates with a student's major choice, while there is no such correlation between a sibling's wage change in the post-choice window and a student's choice. The coefficient of "occupation permanent wage" is strong and significant across all specifications, suggesting that a student is more likely to follow his older sibling's footstep when the older sibling is working in a occupation with high earning. The coefficient on a sibling's permanent wage is around 0.09 without controlling for occupation average wage, and this coefficient decreases as I add the control of occupation average wage. As a robustness check, Table A6 runs a regression with same dependant variable but a set of independent variables with the level of wage in pre-choice window. It shows that a sibling's "pre-choice" wage has additional influence on a student's major choice after controlling for occupation and individual permanent wage. Both results indicate that students are responding to recent wage fluctuations of their older siblings in addition to any change in the occupation average wage.

The difference between Column 1 and Column 3 provides a first identification

[14]The demographic characteristics include a student and her sibling's gender, age, race, region, year of major choice and years of education.

of which channel is more likely to explain the influence on students' major choices. If a sibling's wage in both time windows is associated with a student's major choice similarly, family-correlated preferences would be able to explain the observed correlation. Instead, if only the pre-choice sibling's wage affects a student's choice, learning from family-based information is more likely to be the underlying mechanism, as the pattern in Table 2 cannot be easily explained by unobservable family characteristics. Suppose there is an component of unobserved family characteristic that leads a student more likely to choose a given major and this component leads her sibling more likely to receive higher wage in a related occupation. To explain the influence of family member's wage on a student's major choice as shown in Table 2, this component cannot be time invariable (otherwise its effect is absorbed by the "permanent wage" regressor); and this component has to change right around the time when the student is choosing a college major.

Based on Column 1, a 10% increase in a sibling's pre-choice wage results in a 1.23% increase in a student's likelihood to choose the same major as the older sibling's occupation. Given the average match-ratio of 16%, the marginal elasticity of a student's major choice with respect to a change in her sibling's wage is 0.77. The value is much larger than previous estimates of the elasticity with respect to the change in occupation wage. For example, Wiswall and Zafar (2012), Beffy *et al.* (2012) find the elasticity with respect to the changes in occupation wage is around 0.1. The difference between my estimate and their estimates imply that though students respond little to changes in occupation average wage, they strongly respond to perceived earning opportunities based on their family wage outcomes.

Columns 2 and 4 add controls for the average wage of the sibling's occupation. The coefficient on the "Permanent Occupational Average Wage" suggests that a 10% increase in occupation long-term average wage results in a 1.5% increase in the student's likelihood to choose the major associated with her older sibling's occupation. The effect of "Pre-Choice Sibling's Wage" is still in similar magnitude after controlling for any change in the occupation average wage. The sign of coefficients for other control variables is as predicted: a student's ideal occupation before going to college strongly predicts her major choice, and she is more likely to choose a major that matches the occupation of a sibling of the same gender. Appendix Table A7 lists similar results in a specification with wages measured by log annual income.

I find similar results in the NELS88. Table 3 displays the positive correlation between a student's family income and the probability that she chooses the major associated with her parent's occupation. The dependent variable is a binary indicator for whether a student's major matches her parent's occupation. The key explanatory variable is a dummy variable recording whether a household income increases between 1988 and 1992.[15] Other control variables include a student's ideal occupation, parental years of education, student test scores, as well as other demographic characteristics. Column 1 of Table 3 shows the impact of changes in family income in the S2-Manager sample, and Column 2 shows the effect in the S2-Teacher sample.

Older siblings' wage outcomes not only affect students' choices of major, but also influence their earlier career planning. In 1979 and 1982, the NLSY79 surveyed students who plan to work at age 35 with a question "what kind of work (1970 Census 3-digit occupation code) would you like to be doing when you are 35 years old?". A student's "ideal job" is a proxy for her career plans, and it can be used as an outcome variable to test how an older sibling's wages affect a student's career planning. The dependent variable records whether a student's ideal job is the same as her sibling's occupation (3 digit occupation code level), and the key explanatory variable is a sibling's average wage during 1979-1982. Control variables are a sibling's permanent wage, whether a sibling's occupation is the same as her parent's occupation and other background characteristics.[16]

In Table 4, I find that a student is more likely to plan to work in the same occupation as a sibling if the latter earned a higher wage during 1979-1982. The coefficient on the "Sibling's Wage during Survey Years" indicates that a 10% increase of hourly wage of a sibling increases the probability of a student wishing to work in sibling's occupation by 0.2%. Among 2,396 respondents, 1.59% of the students wished to work in the exact occupation of their siblings. Thus the corresponding elasticity of a student's ideal occupation with respect to the change in her sibling's hourly wage is around 1.[17] The effect of a sibling's wage on "ideal occupation" is robust to controlling the match between the sibling's occupation with a parent's occupation, and a student is more likely to follow an older sibling's footsteps if the sibling chooses the same occupation as their parents. Again, only a sibling's wage received during the survey years but not the permanent wage is associated with a student's ideal occupation, which is consistent with the learning mechanism

[15] The household income distribution is listed in Appendix Table A2. Instead of using the dummy variable to record whether a student's household incomes increases, I can use an indicator to record whether the income increases above a given cutoff (e.g. $10,000). The regression results are the same.

[16] Define a parent's occupation as the father's longest occupation when available. When a father's occupation information is missing, I use a mother's longest occupation as parent's occupation.

[17] Sample size in Table 4 is larger than that in Table 2 because the sample used in Table 4 is not restricted by college attendance.

hypothesis.

3 A Model of Learning from Family Members

Previous section demonstrates the correlation between students' major choices and the recent wage fluctuations of their family members. Though the results provide suggestive evidence that students are learning from family-based wage information, it is difficult to infer the exact learning process or test whether students are using the information optimally based only on the reduced form models. To explain how recent wages of family members affect a student's choice of major, I propose a nested multinomial logit model that embeds how students adjust beliefs of future earnings after observing a family member's wage realizations. Section 3.1 introduces the setting. Sections 3.2 - 3.4 describe how a student updates her belief in detail. Sections 3.5 - 3.7 develop an estimation strategy and show the structural estimations.[18]

3.1 Setting

Preferences

Assume that a college student chooses a major in a two-stage process. First she decides whether she wants to choose a broad field of studies $i \in I = \{1, 2, ..., 5\}$ [19], then she chooses a major j_i within the field i, $j_i \in J_i = \{1, 2, ..., J_i\}$. This nested structure can relax the restriction of IIA assumption in a typical multinomial logit model. I only assume the idiosyncratic preference ξ^{j_i} is drawn from a Type I Extreme Value (Gumbel) distribution within a field. To simplify the notation, I present the model under the condition that i is given, such that I can use j to replace j_i in following equations.

As a student decides which major to declare, she has a family member f already working in the occupation associated with one specific major k. Again to simplify the notation, assume there is one-to-one mapping from majors to occupations this section, and the return to major j can be captured by the average wage in occupation j.[20]

[18] The structural estimation is only based on the S1 sample from the NLSY79, because only the NLSY79 has the detailed annual wage information for a student's family member.

[19] Fields of studies include five fields: "humanities", "social sciences", "health", "STEM"(science, technology, engineering, and mathematics) and "other" (Appendix Table 3).

[20] In the estimation section (Section 3.6), I will introduce the details about how to map average occupation wage to average return to each major.

The value function of the student when she considers choosing major j includes three components: flow utility while attending college, u^j; utility from expected future earnings as a linear function of log wage, E^j; and an idiosyncratic preference, ξ^j. Specifically,

$$V^j = u^j + \theta E^j + \xi^j \tag{1}$$

where θ is the utility weight on future earnings.

Flow utility. A student's flow utility from studying major j can be divided into two components. The first component is a population taste shared by every student, c_0^j. The second is a proxy for individual-specific taste T^j.

The flow utility from studying major j is

$$u^j = c_0^j + c_1 T^j. \tag{2}$$

Utility from Future Wage. A student does not know her future earnings E^j, so she updates her belief through a learning process. By imposing the assumption that students are risk-neutral (linear consumption utility function), only the expected value of future wages enters her utility function. A more general model may add risk aversion in students' preferences, then a student can update her belief about the mean, the variance and other moments of the future wages.

3.2 A Student's Expected Return to College Majors

Suppose a student realized earning in a given occupation j, w^j, can be decomposed as

$$w^j = A^j + \eta^j \tag{3}$$

A^j stands for the ex-ante predictable component for the wage, which comes from observable characteristics such as age, gender, education level and average wage in each occupation. I will refer to A^j as predictable wage, and each student knows it before they start working. η^j represents the unknown wage component, which may include the match quality of personal skills to occupation k for a student, and it is unknown to the student till he starts working.[21]

In addition to major-specific predictable wage, this student also observes her

[21] η^j represents the generic ex-ante unknown wage determinant to each worker, but an intuitive interpretation of η^j is the match quality component (Jovanovic, 1979).

family member's wage realization, w_f^k, and she will use this information to update her own earnings in occupation k. Importantly, when student considers major k, she believes own future earning is correlated with her family member f's w_f^k.

Assumption 1. *Assume that the perceived correlation coefficient between unobservable wage component η^j and w_f^k is such that*

$$corr(\eta^j, w_f^k) = \begin{cases} \lambda & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases} \tag{4}$$

Therefore for any $j \in J_i$, a student believes her expected future earnings will be in the form of

$$E^j = \begin{cases} A^k + \lambda w_f^k & \text{if } j = k \\ A^j & \text{if } j \neq k \end{cases} \tag{5}$$

Suppose this student observes w_f^k, A_f^k, A^j. How would she update her belief about future earnings in occupation k? It depends the correlation between the student's unobservable wage component and her family's contemporaneous wage. If w_f^k does not correlate with η^j, the student should use $\lambda = 0$. If w_f^k correlates with η^j, the optimal weight of family member's wage is $\lambda \neq 0$ and $corr(w^k - A^k, w_f^k) \neq 0$. Therefore λ is the weight students place on family member's wage signals when they update their beliefs about own future earnings. The optimal weight students can use is that $\lambda = corr(w^k - A^k, w_f^k) \in [0, 1]$. The estimation of λ tells how much additional information about future earnings a student receives from her family member's wage realizations, which will be the key finding of this study.

3.3 An Econometrician's Knowledge of Wage Determinants

As an econometrician cannot directly observe A^j, I use observable characteristics of a student and her family member to estimate their predictable wages. These observable variables are summarized as X for the student and $\tilde{X}_f$ for her family member, which may include gender, race, birth cohorts, years of education, test score and working experience.

$$A^j = X'\Pi_j + \epsilon^j \tag{6}$$

$X'\Pi_j$ represents the average wage in occupation j conditioned on pre-determined observable characteristics, and ϵ^j is the measurement error in A^j.[22]

The student's updated value function can be rewritten as:

$$V^j = \begin{cases} u^k + \theta(X'\Pi_k + \epsilon^k + \lambda w_f^k) + \xi^k & \text{if } j = k \\ u^j + \theta(X'\Pi_j + \epsilon^j) + \xi^j & \text{if } j \neq k \end{cases} \tag{7}$$

where ξ^j is drawn from a Type I Extreme Value distribution for a given $i \in I$.

With the measurement error terms ϵ^j , the econometrician has to estimates the following model without knowing the exact distribution of the noise term ζ^j

$$V^{j_i} = \begin{cases} u^k + \theta(X'\Pi_k + \lambda w_f^k) + \zeta^k & \text{if } j = k \\ u^j + \theta X'\Pi_j + \zeta^j & \text{if } j \neq k \end{cases} \tag{8}$$

where $\zeta^j = \theta\epsilon^j + \xi^j$.

I estimate the above model using quasi-maximum-likelihood method as if ζ^j still follows a Type I Extreme Value distribution for the given i. Previous studies, such as Lee (1982), find that estimating a multinomial logit model with independent omitted variables does not generate biased estimators, and the direction of potential bias can be analyzed based on the covariances between the the omitted variables and explanatory variables.[23]

In this model, ϵ^j is known to the students and they use it when they calibrate the expected earnings. An econometrician, however, cannot incorporate ϵ^j in the estimation. A family member's realized wage w_f^k can correlate with the omitted variable — measurement error term ϵ^j, and such correlation causes bias in the estimation of λ.

A family member's wage can be decomposed as:

$$w_f^k = \tilde{X}_f'\tilde{\Pi}_k + \tilde{\epsilon}_f^k \tag{9}$$

$\tilde{X}_f'\tilde{\Pi}_k$ is the estimate for the predictable wage of family member f based on observable characteristics, and $\tilde{\epsilon}_f^k$ represents the wage residual of a family member.

[22]In previous studies on educational decisions, $X'\Pi_j$ usually represents a student's rational expectation for future returns, such as in Rosen and Willis (1979), Siow (1984), Berger (1988), Keane and Wolpin (1997), Rosen and Ryoo (2004). In these studies, students have common knowledge of the actual process generating life-cycle incomes conditional on personal variables, and they apply such knowledge to forecast future personal income should he or she choose a major.

[23]An alternative estimation strategy is to use a multinomial probit model. Given the choice-set is 22 different majors, the computation burden is not trivial.

According to Equation (7), the estimators of θ and λ can be biased if the omitted variable ϵ^k is correlated with the sum of $\tilde{X}_f'\tilde{\Pi}_k$ and $\tilde{\epsilon}_f^k$.

- $cov(\tilde{X}_f'\tilde{\Pi}_k, \epsilon^k) \sim 0$ A family member's predictable wage is unlikely to correlate with a student's unobserved expected earnings, because $\tilde{X}_f'\tilde{\Pi}_k$ is population average wage based on pre-determined observables.

- $cov(\tilde{\epsilon}_f^k, \epsilon^k) > 0$ $\tilde{\epsilon}_f^k$ is family member's wage residual. It is plausible that there exists a positive correlation between ϵ^k and $\tilde{\epsilon}_f^k$. This positive correlation can cause an upward biased estimator of λ when directly estimating Equation (7).

Therefore, I will present two sets of estimation results based on w_f^k and $\tilde{X}_f'\tilde{\Pi}_k$ respectively. Using $\tilde{X}_f'\tilde{\Pi}_k$ can help to correct the bias in estimation of λ.

3.4 Estimation Strategy

Estimating θ and λ requires the specification of T^j, $X'\Pi_j$ and $\tilde{X}_f'\tilde{\Pi}_k$. A student's individual taste proxy, T^j, is the number of classes she has taken in high school related to major j, adjusted by the population average and standard deviation. I also add a dummy variable recording if a student's family member works in occupation j as the robustness check specification for the individual taste proxy.

I estimate $X'\Pi_j$ using a Mincerian wage regression with 1990 census micro data.[24] In the 1990 census sample, a representative worker n's wage in occupation j can be characterized as

$$w_n^j = X_n'\Pi_j + \varepsilon_n^j \tag{10}$$

where X_n include gender, race, birth cohorts, years of education, a quadratic function of working experience, and ε_n^j is the noise term.

This regression gives occupation-specific wage coefficients Π_j. Thus, the predictable wage to major j for a student is $X'\Pi_j$ — the average starting wage of a student conditional on pre-determined observables.[25] However, since students enrolled in major j can potentially work in occupations other than j, the value of

[24] I use the census data to estimate Π_j, because it gives very small standard error for the estimate of Π_j. There are 1,272,594 individuals born between 1957 and 1964 with a professional occupation that can be categorized into 22 college majors. The observable variables X shared by NLSY79 and 1990 Census 5% Micro Sample include gender, race, years of education, and birth year dummies.

[25] I use the starting wage as a proxy of life earnings. Flyer (1997) finds that the correlation coefficient between projected occupational starting wage and projected occupational life-cycle earnings (with mobility) is over 0.5 in all six occupations. In five of the six occupations the simple correlation coefficient is greater than 0.65.

$X'\Pi_j$ is the weighted average of an occupational starting wage multiplied the match-probability between any occupation and major j. The match-probability between professional occupation j_1 and major j is calculated by the fraction of workers who graduated with major j working in occupation j_1, coded as Pr^{jj_1}. The predicted return to major j is $X'\Pi_j = \sum_{j_1=1}^{22} Pr^{jj_1} X'\Pi_{j_1}$.

To estimate $\tilde{\Pi}_k$, recall Equation (10)

$$w_f^k = \tilde{X}_f' \tilde{\Pi}_k + \tilde{\epsilon}_f^k$$

I regress a family member's contemporaneous wage w_f^k (the average wage in the pre-choice window) on the observable characteristics of family member f, $\tilde{X}_f$. $\tilde{X}_f$ include the pre-determined observables that shared with the student, X_f, f's AFQT score, and year fixed-effect dummies DY_{time} — capturing the population average wage of all workers at the time when student declares her major.[26] By using the term $\tilde{X}_f' \tilde{\Pi}_k$, I can exclude the wage residual in Equation (11) that contains the unobserved wage a student observes but not the econometrician.

$$w_f^k = X_f \hat{\Pi}_k + b_1 AFQT_f + b_2 DY_{time} + \tilde{\epsilon}_f^k \tag{11}$$

where X_f include a family member's gender, race, birth cohort, region, and years of education.

3.5 Structural Parameter Estimation

The parameters of interest include the weight on predictable wage, θ; the weight on family member's wage, $\theta \cdot \lambda$; and flow-utility function parameters c_0^j and c_1. Table 5 lists the estimation for these parameters. (Table A5 compares the estimation result in a classic multinomial logit model.)

Table 5 presents three different specifications. Column 1 estimates a major-choice model with $\lambda = 0$. Column 2-3 present the model that incorporating wage information based on family member's wages $\lambda \neq 0$. Column 2 uses the specification with a family member's realized wage w_f^k as the wage signal to students and Column 3 uses the predicted wage component $\tilde{X}_f' \tilde{\Pi}_k$ to correct potential omitted-variable bias.

According to Table 5, $\hat{\theta}$ is positive and significant, ranging from 1.45 to 1.89 across all three specifications. This shows that monetary payoff is definitely an

[26] w_f^k equals 4-year-average log hourly wage before the student declares her major, defined in Section 2.1.

important factor in determining student's major choice. My estimation of $\hat{\theta}$ is very close to the comparable parameter in previous studies. For example, Arcidiacono *et al.* (2012) estimate a similar multinomial logit model of choice, though they directly elicited students' expectations of future wages with on-campus survey. The $\hat{\theta}$ in that paper ranges from 1.46 to 1.69, which cross-validates my estimates.

Parameter $\hat{\theta} \cdot \hat{\lambda}$ tells the importance of a close family member's wage in determining students' major choices. Column 2 in Table 5 shows that $\hat{\theta} \cdot \hat{\lambda}$ is around 0.58, indicating that a close family member's wage has positive and significant influence. Column 3 in Table 5 uses family member's predictable wage to correct the potential bias in the estimation. After controlling for the family member's wage residual, $\hat{\theta} \cdot \hat{\lambda}$ is still as high as 0.57. This result confirms that students adjust their major choices based on family wage outcomes, not only because these wage outcomes reflect family-correlated unobserved ability. In summary, adding information on contemporaneous wage realizations of siblings leads to an improvement in model fit relative to a benchmark major choice model that assumes students' expectations are just based on occupation average wage.[27]

Dividing $\hat{\theta} \cdot \hat{\lambda}$ by $\hat{\theta}$ gives the estimator $\hat{\lambda}$, and its standard error can be calculated by the Delta method. According to Table 5, the point estimate of λ falls in the range between 0.36 and 0.39. The standard error of $\hat{\lambda}$ is around 0.1, so students perceive the wage correlation between siblings ($\hat{\lambda}$) to be positive and significant.

The estimates of other parameters in utility functions are all in the sign as predicted. For example, personal taste proxy (normalized number of classes taken in high school related to each college major) is strong and robust predictor of students' major choices. Appendix Table A4 lists all the estimates for demographic characteristics and location parameter for each field of study.

4 Over-learning from Family Experience?

Can the estimations in Table 5 tell whether students use family-based wage information optimally?

This section takes two different approaches to answer this question. I first compare my estimate of perceived correlation by students, $\hat{\lambda}$, with the actual correlation between siblings' wages found in the data. The comparison shows that the decision weight students placed on family wage is much larger than the empirical correlation between own earnings and their siblings' earnings. Another approach is to compare

[27]For example, see Siow (1984), Berger (1988), Keane & Wolpin (1997), and Ryoo & Rosen (2004).

the labor market performance between students who received positive wage signals to those who received negative signals.[28] I find that students who received positive wage signals perform worse later in the labor market than those who received negative wage signals, a result that is a hard to be explained by family correlated ability.

4.1 Perceived v.s. Empirical Wage Correlation

Recall the wage determination process defined in Equation (3), a student's potential wage realization in occupation k can be written as

$$w^j = A^j + \eta^j$$

To estimate $\lambda = corr(\eta^k, w_f^k)$, I run the following two regressions conditional on the match between students' majors and their siblings' occupations:

$$w^k = \underline{\lambda} w_f^k + X'\underline{\Pi} + \varepsilon_s \tag{12}$$

where ε_s is a noise term.

The ideal dependent variable for above regressions is the potential wage w^k for every student. As we only observe the realized wage of students who actually declared major k, above regression needs selection correction. The instrument variables I use is an indicator variable recording whether a student's pre-determined ideal occupation match with her sibling's ideal occupation to conduct the selection correction. This indicator strongly correlates with whether a student chooses a major matched with her sibling's occupation, but not correlate with a student's wage outcome.

Table 6 presents the results for above regressions both with and without selection correction. Columns 1 and 2 estimate the correlation coefficient in a sub-sample of students who chose the major matched to their siblings' occupation. The correlation coefficient estimates $\underline{\lambda}$ is negative though not significant. Columns 3 and 4 demonstrate the same regression using the whole sample of students, and the unconditional correlation between students realized wage and their siblings' wages is larger. The correlation coefficient between a student's starting wage and her older sibling's wage without selection correction is around 0.09 (Column 3), but such correlation disappeared in Column 4 after adding selection correction.

[28] A wage signal is positive if the family member earns a higher than average wage in a given occupation.

As shown in Table 5, the point estimate of λ is 0.39, thus students' perceived correlation is larger than the correlation coefficient in Table 6. The relationship implies that students overestimate the predictive power of family members earnings, and such overweighing may lead students form biased belief in future earnings. To further test whether students place too much weight on family wage outcomes, I take an alternative approach by comparing the labor market performance between students who received different wage signals.

4.2 Students' Labor Market Outcomes

My model predicts that students sort into or out of a major in response to family wage signals. If a family member's wage is higher than the average wage among all the workers in that occupation, I define that the wage signal a student receives from this family member is positive. A positive wage signal from close family member lead students with less competitive characteristics in occupation k (lower A^k) more likely choose major k. Despite such sorting process, as long as students interpret the wage signals correctly, those who received positive wage signals should still have higher realized wage compared to others who received negative signals.

To investigate whether this intuition is consistent with my model, I run following thought experiment. Suppose a student is considering whether to choose major k when her family member works in occupation k. Equation (1) and (3) suggest that a student's decision is determined by the expected value of her wages in occupation k (E^k) and the best alternative option she has ($\bar{O}$).[29] A student chooses major k if and only if $E^k > \bar{O}$ and her expected wage becomes,

$$E^k = A^k + \lambda w_f^k$$

For the students who actually choose major k, their earnings in occupation k (w^k) are observable.

$$w^k = A^k + \eta^k$$

These students who choose major k also have older siblings working in occupation k. The wages of older siblings are w_f^k

$$w_f^k = \tilde{X}_f' \tilde{\Pi}_k + \tilde{\epsilon}_f^k$$

[29] I excluding the difference in flow utility function here. The simplification holds if students choose majors based on the expected wages after controlling for the compensating differentials.

Students know their predictable wage component A^k and their siblings' wages in occupation k. When students are using family wage information optimally, I want to know whether the average wage of all students whose family members received positive wage signals is larger than the average wage of students whose family members only experienced negative wage shocks for the given λ. I use following calculation to answer this question.

Among students whose family members are working in occupation k, divide them to two groups **1** and **2**. The family member of each student in group **1** has a positive wage signal $M_1 = \lambda(w_f^k - \bar{w^k}) > 0$; while students in group **2** receive a negative wage signal $M_2 = \lambda(w_f^k - \bar{w^k}) < 0$. I further assume M_1 is drawn from the positive half of a normal distribution $N(0, \sigma_k^2)$, and M_2 is drawn from the negative half of that distribution. $\lambda = 0.39$.

The selection rule for each student to choose major k in group **1** is $w_1^k > \bar{O}_1$, similarly group **2** students choose major k if and only if $w_2^k > \bar{O}_2$. Suppose a student's best alternative option is not correlated with the student's family wage signal, then the distribution of $\bar{O}_1$ and $\bar{O}_2$ is the same across the two groups and its c.d.f. is denoted as $H(o)$.

The wage determination process in Equation (3) indicates that a student's predictable wage component A is not correlate with the wage signal M by construction. The distribution of A is similar across the two groups and I assume A across students in each group is from a normal distribution $N(\mu, \sigma_\alpha^2)$.

The average realized wage in group **1** is $\bar{w}_1$

$$\begin{aligned}\bar{w}_1 =& \frac{\iiint (A+M)\mathbf{1}(A+M>o)\mathbf{1}(M>0)f(M)h(o)g(A)dodAdM}{\iiint \mathbf{1}(A+M>o)\mathbf{1}(M>0)f(M)h(o)g(A)dodAdM} \\ =& \frac{\iint (A+M)H(A+M)\mathbf{1}(M>0)f(M)g(A)dAdM}{\iint H(A+M)\mathbf{1}(M>0)f(M)g(A)dAdM}\end{aligned}$$

The average realized wage in group **2** is $\bar{w}_2$

$$\begin{aligned}\bar{w}_2 =& \frac{\iiint (A+M)\mathbf{1}(A+M>o)\mathbf{1}(M<0)f(M)h(o)g(A)dodAdM}{\iiint \mathbf{1}(A+M>o)\mathbf{1}(M<0)f(M)h(o)g(A)dodAdM} \\ =& \frac{\iint (A+M)H(A+M)\mathbf{1}(M<0)f(M)g(A)dAdM}{\iint H(A+M)\mathbf{1}(M<0)f(M)g(A)dAdM}\end{aligned}$$

Appendix Section A5 shows that $\bar{w}_1 > \bar{w}_2$ for students in my sample.

Putting optimal weight on family member's wage outcomes, the students whose family members receive positive wage signals are predicted to do better than others who receive negative signals. However, my regressions on a student's labor market outcome and her family wage shocks show that students who received positive wage signals are actually doing *worse* than those who did receive negative wage signals, which suggesting that students overreact to family wage changes instead of using such information optimally.

The empirical specification I use is in Equation (13). LM represents the labor market outcome for a student; Y is an indicator for that a student works in the same occupation as her close family member, and $S_f^k = \mathbf{1}\{w_f^k - \bar{w}\}$ is the indicator for positive wage signal. The labor market outcome measurements include: a student's starting wages after college graduation, whether a student changes major during college years, and whether a student becomes unemployed within five years after graduation.

r_2 measures the effect of choosing a major matched with her family member's occupation, and r_3 records the effect on a student's labor market performance when her family member received a positive wage signal. $r_4 = (\bar{w}_1 - \bar{w}_2)$ actually measures the differential effect between a student who chooses a matched major after her family member received higher than occupation average wages and other students who choose a matched major when their family members received lower than average wage.

$$LM = r_0 + X'r_1 + r_2Y + r_3S_f^k + r_4Y \cdot S_f^k + +\nu \tag{13}$$

Table 7 presents the regression results from the above specification in the NLSY79. Compared to students who choose a major that is not related to an older sibling's occupation, a student who chooses a major matched with her sibling's occupation while the sibling earns lower than average wage actually earns a higher wage and he is less likely to change college major. However, when a student chooses a matched major with positive wage signal compare to others who received negative signals, this student is 21% more likely to switch majors in college and earns a 17% lower wage.

Table 8 shows the similar results for students in S2 of the NELS88. A student earns about 10% lower wage when she chooses a matched major after her parent's income rises during high school years, relative to other students whose family income did not increase during high school. This wage difference is only significant for S1-Manager sample, probably because the wages of teachers vary little over time.

These results indicate that worse-matched students sort into a major associated with family's occupation after observing positive wage signals from their family members. If the wage signals are informative and students place the appropriate weight on the wage shocks when they update their beliefs, there will not be such a huge difference in overall labor market performances between students whose siblings received positive wage signals and other students whose siblings received negative wage signals. The results in this section point in direction that students overreact to wage shocks experienced by their family members.

5 Conclusion

This paper studies how students adjust career path based on the wage changes of their family members. Students interpret the family wage changes as additional information about their future earnings in a certain occupation. However, they are likely to overestimate the predictive power of family members earnings. The decision weight students place on family wage signals is larger than the empirical correlation between their own earnings and their family members' earnings, and there is strong evidence that students who receive positive wage signals perform worse later in the labor market than those who receive negative signals.

Previous studies have found large wage premiums for business, engineering and science majors, suggesting many students could earn higher wages if they choose alternative majors.[30] And yet, enrollment for many high-wage college majors stays low while the enrollment for low-wage majors remains high.[31] For this reason, the President's Council of Advisors on Science and Technology thus calls for a big increase in college graduates in science, technology, engineering and mathematics (STEM majors).

How can we inform students that there is an increasing market demand for STEM majors relative to some currently popular majors, such as business, social sciences, history or education? Previous survey studies have shown students have very limited information about occupation wage differentials.[32] However, estimations in this paper reveal that young people strongly respond to perceived earnings opportunities. If their choices are restricted by their limited information about labor market conditions, they could not respond to the increasing market demand

[30] Hamermesh & Donald (2008) estimate that there is a 40% gap in annual earnings between college graduates who majored in business and those who majored in humanities after controlling for hours of work, academic performance, and other background characteristics. Similar findings are shown in Arcidiacono (2004) and Altonji *et al.* (2012).

[31] The potential supply of students in STEM majors are usually not restricted by program size, though the early-decision deadline and college preparatory requirements may limit the size of the applicant pool.

[32] Betts (1996) first documents that college students have limited knowledge of salaries by fields of major. Arcidiacono *et al.* (2012) and Wiswall & Zafar (2012) find similar results.

even though they prefer higher earnings. The fact that students have limited information may also be the reason that they overly rely on the wage information coming from close family members. Considering the availability of precise wage information from school career centers or public agencies like the Bureau of Labor Statistics, students can potentially improve their predictions of future earnings by utilizing these sources of outside information.

The heterogeneity in learning by family backgrounds found in this paper also has direct policy implications for promoting intergenerational mobility. If family socio-economic status strongly influences students' expectations on future earnings, students from disadvantaged families may never have the opportunity to know the actual return to higher education or the wages in certain professions, their career planning is constrained by the insufficient information. For example, a recent study by Hoxby and Avery (2012) finds that students with high-achievement but from low-income families misunderstand the actual cost of prestigious private universities. The empirical results in this paper suggest that providing information on career prospects to students from disadvantaged families may help them expand their choice sets of education and occupation options.

Table 1: Summary Statistics

Panel A: NLSY79	College Sample	Student-Sibling (S1)	
		Student	Siblings
Male	45.9%	48.3%	48.8%
Female	54.1%	51.7%	51.2%
Non-Hispanic White	62.8%	62.9%	62.8%
African American	23.5%	21.2%	21.2%
Hispanic	13.7%	15.9%	16.0%
AFQT Score	58.0	60.9	55.0
Average Hourly Wage during Age 30-35 (2010 $)	22.2	23.5	21.5
Average Annual Labor Income during Age 30-35 (2010 $)	43k	48k	42k
Total Weeks Unemployed during Age 30-35	12.5	13.3	10.1
Years of Completed Education	15.0	15.6	13.9
N	5585	1004	1004
Student's Major = Sibling's Occupation		16.2%	

Panel B: NELS88	Whole Sample	Student-Parent (S2)	
		Manager	Teacher
Male	45.5%	45.9%	49.5%
Female	54.5%	54.1%	50.5%
Non-Hispanic White	70.7%	76.0%	80.1%
African American	9.6%	6.0%	7.9%
Hispanic	8.3%	7.5%	6.1%
Other Race	11.4%	10.5%	5.8%
1999 Annual Labor Income (1999 $)	23k	27k	26k
N	7335	1094	705
Student's Major = Parent's Occupation		16.5%	13.5%

Table 2: The Impact of a Sibling's Wages on a Student's Major Choice

Dependent Variable: **Student's Major = Sibling's Occupation**	(1)	(2)	(3)	(4)
Sibling's Wage Pre-Choice - Permanent	0.123***	0.110***		
	[0.0364]	[0.0365]		
Occupation Average Wage Pre-Choice - Permanent		−0.373***		
		[0.125]		
Sibling's Wage Post-Choice - Permanent			0.0424	0.0359
			[0.0406]	[0.0398]
Occupation Average Wage Post-Choice - Permanent				−0.191
				[0.180]
Sibling's Permanent Wage	0.0872**	0.0644*	0.0419	0.0191
	[0.0347]	[0.0358]	[0.0299]	[0.0311]
Occupation Permanent Wage		0.150**		0.232***
		[0.0659]		[0.0608]
Ideal Occupation = Sibling's Occupation	0.437***	0.445***	0.418***	0.420***
	[0.0549]	[0.0537]	[0.0544]	[0.0536]
Same Gender with Sibling	0.0638***	0.060**	0.0645***	0.0668***
	[0.0243]	[0.0241]	[0.0231]	[0.02311]
Education & Demographics	X	X	X	X
Observations	853	852	908	908
R-squared	0.167	0.181	0.153	0.164
Point Elasticity	0.765	0.735	0.265	0.224

Note:

1. Clustered standard errors by household in brackets, *** p<0.01, ** p<0.05, * p<0.1
2. Siblings' wages are measured in logged hourly wage in 2010 Dollars
3. Pre-choice window: 0-3 years before the major choice; Post-choice window: 1-4 years after the choice.
4. Control variables include gender, race, region, birth year dummies, own and sibling's education, own and sibling's AFQT score, highest education of parents, and the age when declaring the major.

Table 3: The Impact of a Parent's Wage on Major Choice

Dependent Variable:	S2-Manager	S2-Teacher
Student's Major = Parent's Occupation	(1)	(2)
Increase in Family Income in Pre-Choice Window	0.0196**	0.0283***
	[0.0097]	[0.0101]
Ideal Occupation = Parent's Occupation	0.243***	0.237***
	[0.0613]	[0.0683]
Average Math Test Score	0.114	−0.150*
	[0.0702]	[0.0826]
Average Reading Test Score	-0.119	0.0815
	[0.0735]	[0.0805]
Base Year Family Income	X	X
Education & Demographics	X	X
Observations	1,093	705
R-squared	0.057	0.107

Note:
1. Clustered standard errors by household in brackets, *** p<0.01, ** p<0.05, * p<0.1
2. Income in 1999 dollars. Base Year Family Income is a category variable.
3. Control variables include gender, race, region, birth year dummies, and parental education.

Table 4: The Impact of a Sibling's Wage on a Student's Ideal Job

Dependent Variable:	Hourly Rate Wage		Annual Income	
Student's Ideal Occupation = Sibling's Occupation	(1)	(2)	(3)	(4)
Sibling's Wage during Survey Years	0.0212*	0.0236*	0.0114**	0.0094*
	[0.0125]	[0.0129]	[0.0051]	[0.0056]
Sibling's Permanent Wage	-0.0073	−0.0109*	-0.0075	-0.0095
	[0.0058]	[0.0058]	[0.0066]	[0.0065]
Sibling's Occupation = Parental Occupation		0.146**		0.154**
		[0.0587]		[0.0617]
Sibling in Same Gender	0.023***	0.024***	0.023***	0.024***
	[0.0073]	[0.0074]	[0.0073]	[0.0074]
Education & Demographics	X	X	X	X
Observations	1,192	1,051	1,194	1,054
R-squared	0.020	0.059	0.021	0.061
Point Elasticity	1.33	1.48	0.717	0.591

Note:
1. Clustered standard errors by household in brackets, *** p<0.01, ** p<0.05, * p<0.1
2. All wages are measured in logged term and normalized by 2010 Dollars
3. Control variables include gender, race, region, birth year dummies, sibling's education, and AFQT score.

Table 5: Nested Logit Model Structural Parameter

Major Choice	(1)	(2)	(3)
Log-likelihood	-2729.0815	-2301.6342	-2611.4581
$\hat{\theta}$ - Predictable Wage	1.887***	1.604***	1.455***
	[0.436]	[0.459]	[0.430]
$\hat{\theta}\hat{\lambda}$ - Family's Realized Wage		0.582***	
		[0.051]	
$\hat{\theta}\hat{\lambda}$ - Family's Predicted Wage			0.572***
			[0.049]
$\hat{c}_1$ - Personal Taste Proxy	0.265***	0.288***	0.265***
	[0.032]	[0.035]	[0.033]
Demographics	X	X	X
Field of Study τ X	X	X	
# Observation	21,428	18,788	21,054
# Case	974	854	957
# Parameter	11	12	12
LR test for IIA: chi2(5)	25.03	40.33	41.90
Perceived Correlation of Match Quality		$\hat{\lambda} = 0.363^{***}$ [0.111]	$\hat{\lambda} = 0.393^{***}$ [0.124]

Table 6: Correlation between a Sibling's Wages and a Student's Wages

Dependent Variable:	Matched Sample		Whole Sample	
Student's Starting Wage	(1)	(2)	(3)	(4)
Sibling's Realized Wage w_f	-0.106	-0.172	0.0950***	-0.882^{*}
	[0.103]	[0.488]	[0.035]	[0.456]
Education & Demographics	X	X	X	X
IV 2SLS	-	X	-	X
Observations	151	148	866	827
R-squared	0.146	-	0.110	-

Note:
1. Clustered standard errors by household in brackets, *** p<0.01, ** p<0.05, * p<0.1.
2. All wages are measured in logged hourly rate 2010 Dollars
3. Control variables include gender, race, region, and birth year dummies.

Table 7: Students' Labor Market Outcomes and Siblings' Wage Outcomes

Dependent Variable:	**Changing Major**	**Unemployment**	**Starting Wage**
	(1)	(2)	(3)
Major Match & Positive Wage Signal	0.218**	0.114	−0.178*
	[0.010]	[0.077]	[0.0952]
Student's Major = Sibling's Occupation	−0.087*	0.006	0.140**
	[0.048]	[0.036]	[0.058]
Positive Wage Signal	−0.074*	−0.072***	0.065*
	[0.040]	[0.027]	[0.033]
Education & Demographics	X	X	X
Occupation Fixed Effect	X	X	X
Observations	794	794	786
R-squared	0.135	0.068	0.188

Note:

1. Clustered standard errors by household in brackets, *** p<0.01, ** p<0.05, * p<0.1
2. Wages are measured in logged hourly rate and normalized by 2010 Dollars
3. Control variables include gender, race, region, birth year dummies, years of education and AFQT score.

Table 8: Students' Starting Wages and Family Income Changes

Dependent Variable: **Student's Annual Income in 1999**	S2-Manager (1)	S2-Teacher (2)
Major Match & Positive Family Income Change	-0.103^{**}	-0.0639
	[0.0478]	[0.0988]
Student's Major = Parents's Occupation	0.269^{***}	-0.112
	[0.0536]	[0.0801]
Positive Family Income Change	0.0195	0.0235
	[0.0217]	[0.0283]
Average Math Score	0.352^{**}	0.403^{**}
	[0.140]	[0.193]
Average Reading Score	-0.347^{**}	-0.294^{*}
	[0.139]	[0.173]
Female	-0.102^{**}	-0.150^{***}
	[0.0471]	[0.0541]
Base Year Family Income	X	X
Occupation FE	X	X
Education & Demographics	X	X
Observations	1,002	643
R-squared	0.085	0.109

Note:
1. Robust standard errors, *** p<0.01, ** p<0.05, * p<0.1
2. Income in 1999 dollars. Base Year Family Income is a category variable.
3. Control variables include gender, race, region, birth year dummies, years of education.

References

Altonji, Joseph G., Blom, Erica, & Meghir, Costas. 2012. Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers. *NBER Working Paper No. 17985*, April.

Arcidiacono, Peter. 2004. Ability Sorting and the Returns to College Major. *Journal of Econometrics*, **121**(1-2), 343–375.

Arcidiacono, Peter, Hotz, Joseph, & Kang, Songman. 2012. Modeling College Major Choices Using Elicited Measures of Expectations and Counterfactuals. *Journal of Econometrics*, **166**(1), 3–16.

Bank, Barbara J., Slavings, Ricky L., & Biddle, Bruce J. 1990. Effects of Peer, Faculty, and Parental Influences on Students' Persistence. *Sociology of Education*, **63**(3), 208–225.

Becker, Gary S. 1962. Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy*, **70**(5), 9–49.

Beffy, Magali, Fougere, Denis, & Maurel, Arnaud. 2012. Choosing the Field of Study in Post-Secondary Education: Do Expected Earnings Matter? *The Review of Economics and Statistics*, **94**(1), 334–347.

Berger, Mark C. 1988. Predicted Future Earnings and Choice of College Major. *Industrial and Labor Relations Review*, **41**(3), 418–429.

Betts, Julian R. 1996. What Do Students Know about Wages? Evidence from a Survey of Undergraduates. *Journal of Human Resources*, **31**(1), 27–56.

Black, Sandra, & Devereux, Paul J. 2011. Recent Developments in Intergenerational Mobility. *Handbook of Labor Economics*, **4**(5), 1489–1541.

Blom, Erica. 2012. Labor Market Determinants of College Major. *Manuscript*, November.

Brody, Gene H. 1998. Sibling Relationship Quality: Its Causes and Consequences. *Annual Review of Psychology*, **49**(February), 1–24.

Cameron, Stephen V., & Heckman, James J. 2001. The dynamics of educational attainment for black, hispanic, and white males. *Journal of Political Economy*, **109**(3), 455–499.

Corak, Miles, & Piraino, Patrizio. 2011. The Intergenerational Transmission of Employers. *Journal of Labor Economics*, **29**(1), 37–68.

Duncan, Greg J., Boisjoly, Johanne, & Harris, Kathleen M. 2001. Sibling, Peer, Neighbor, and Schoolmate Correlations as Indicators of the Importance of Context for Adolescent Development. *Demography*, **38**(3), 437–447.

Fernandez, Raquel. 2013. Cultural Change as Learning: The Evolution of Female Labor Force Participation over a Century. *The American Economic Review*, **103**(1), 472–500.

Flyer, Fredrick A. 1997. The Influence of Higher Moments of Earnings Distributions on Career Decisions. *Journal of Labor Economics*, **15**(4), 689–713.

Freeman, Richard B. 1975a. Legal Cobwebs: a Recursive Model of the Labor Market for New Lawyers. *The Review of Economics and Statistics*, **57**(2), 171–179.

Freeman, Richard B. 1975b. Supply and Salary Adjustments to the Changing Science Manpower Market: Physics, 1948-1973. *The American Economic Review*, **65**(1), 27–39.

Freeman, Richard B. 1976. A Cobweb Model of the Supply and Starting Salary of New Engineers. *Industrial and Labor Relations Review*, **29**(2), 236–248.

Hamermesh, Daniel S., & Donald, Stephen G. 2008. The Effect of College Curriculum on Earnings: An Affinity Identifier for Non-ignorable Non-response Bias. *Journal of Econometrics*, **144**(2), 479–491.

Hellerstein, Judith K., & Morrill, Melinda S. 2011. Dads and Daughters The Changing Impact of Fathers on Women's Occupational Choices. *Journal of Human Resources*, **46**(2), 333–372.

Hoxby, Caroline M., & Avery, Christopher. 2012. The Missing One-Offs: The Hidden Supply of High-Achieving, Low Income Students. *NBER Working Paper*, **18586**(December).

Hyman, Herbert H. 1942. The Psychology of Status. *Archives of Psychology (Columbia University)*, **269**, 94.

Jensen, Robert. 2010. The (Perceived) Returns to Education and the Demand for Schooling. *The Quarterly Journal of Economics*, **125**(2), 515–548.

Jovanovic, Boyan. 1979. Job Matching and the Theory of Turnover. *Journal of Political Economy*, **87**(5), 972–990.

Keane, Michael P., & Wolpin, Kenneth I. 1997. The Career Decisions of Young Men. *Journal of Political Economy*, **105**(3), 473–522.

Lee, Lung-Fei. 1982. Specification Error in Multinomial Logit Models: Analysis of the Omitted Variable Bias. *Journal of Econometrics*, **20**(2), 197–209.

Manski, Charles F., & Wise, David A. 1983. *College Choice in America*. Cambridge, Mass.: Harvard University Press.

Ryoo, Jaewoo, & Rosen, Sherwin. 2004. The Engineering Labor Market. *Journal of Political Economy*, **112**(S1), S110–S140.

Siow, Aloysius. 1984. Occupational Choice under Uncertainty. *Econometrica*, **52**(3), 631–645.

Weast, Harry P. 1956. The Influence of Older Siblings. *The High School Journal*, **39**(6), 340–343.

Willis, Robert J., & Rosen, Sherwin. 1979. Education and Self-Selection. *Journal of Political Economy*, **87**(5), S7–S36.

Wiswall, Matthew, & Zafar, Basit. 2012. Determinants of College Major Choice: Identification Using an Information Experiment. *Working Paper*, March.

1 Appendix

A1 Summary Statistics

Table A1: Distribution of Major and Occupation in NLSY79

	College Major		Occupation	
FIELD OF STUDIES/PROFESSION	Whole	S1	Whole	S1
0100 Agriculture and Natural Resources	1.5%	1.7%	10.4%	5.3%
0200 Architecture and Environmental Design	1.5%	1.3%	0.3%	0.6%
0300 Area Studies	0.2%	0.1%	0.0%	0.0%
0400 Biological Sciences	3.7%	4.7%	1.3%	1.5%
0500 Business and Management	26.5%	27.2%	27.8%	28.4%
0600 Communications	3.1%	2.2%	0.5%	0.9%
0700 Computer and Information Sciences	7.3%	8.5%	8.2%	9.0%
0800 Education	9.5%	9.8%	6.5%	10.4%
0900 Engineering	9.2%	10.0%	6.9%	6.3%
1000 Fine and Applied Arts	5.2%	4.7%	4.7%	4.8%
1100 Foreign Languages	0.7%	0.6%	0.1%	0.2%
1200 Health Professions	10.8%	10.8%	19.1%	18.9%
1300 Home Economics	0.9%	1.3%	1.9%	1.4%
1400 Law	1.4%	1.1%	0.4%	1.0%
1500 Letters	2.1%	1.3%	0.4%	0.6%
1600 Library Science	0.0%	0.0%	0.3%	0.3%
1700 Mathematics	1.4%	1.3%	0.1%	0.2%
1800 Military Sciences	0.2%	0.2%	0.5%	0.2%
1900 Physical Sciences	2.0%	1.8%	0.7%	0.6%
2000 Psychology	4.1%	3.2%	0.1%	0.2%
2100 Public Affairs and Services	3.8%	3.8%	9.6%	8.2%
2200 Social Sciences	4.9%	4.5%	0.2%	0.9%
2300 Theology	0.0%	0.0%	0.3%	0.5%

A2 Mapping occupations to college major

The NLSY79 code students' college majors in twenty-three categories. Each individual's occupation is recorded in 1970 Census Occupational Code. To link a sibling's occupation with a student's college major, I map all professional occupations in 1970 Census to a certain major by the college major held by the majority of college educated workers of that occupation. The mapping between major and occupations are in Table A3.

Table A2: Family Income Distribution in NELS88

NELS88 Parent-Child Sample	Manager-Business	Teacher-Education
1988 Family Income		
\$1,000 - \$2,999	0.09%	0.14%
\$3,000 - \$4,999	0.09%	
\$5,000 - \$7,499	0.37%	0.57%
\$7,500 - \$9,999	0.82%	0.14%
\$10,000 - \$14,999	2.65%	1.28%
\$15,000 - \$19,999	3.38%	1.42%
\$20,000 - \$24,999	4.30%	6.10%
\$25,000 - \$34,999	14.72%	15.04%
\$35,000 - \$49,999	25.05%	29.79%
\$50,000 - \$74,999	28.88%	31.35%
\$75,000 - \$99,999	9.23%	7.94%
\$100,000 - \$199,999	8.32%	4.54%
\$200,000 OR MORE	2.01%	1.28%
1992 Family Income		
\$1,000-\$2,999	0.09%	
\$3,000-\$4,999	0.18%	
\$5,000-\$7,499	0.37%	
\$7,500-\$9,999	0.82%	0.85%
\$10,000-\$14,999	2.10%	1.13%
\$15,000-\$19,999	3.38%	1.42%
\$20,000-\$24,999	4.11%	4.11%
\$25,000-\$34,999	8.96%	8.94%
\$35,000-\$49,999	17.55%	20.43%
\$50,000-\$74,999	31.54%	36.74%
\$75,000-\$99,999	13.25%	13.76%
\$100,000-199,999	14.81%	10.50%
\$200,000 OR MORE	2.83%	2.13%

Table A3: Mapping Occupations to College Majors

College Major	1970 Census Occupation Code							
0100 Agriculture and Natural Resources	25	801	802	821	822	*24*	*102*	*42*
0200 Architecture and Environmental Design	2	*213*	*95*					
0300 Area Studies	*135*							
0400 Biological Sciences	44	74	105	140	150	195	104	
0500 Business and Management	1	56	91	*116*	202	205	216 - 225	230 - 235
0600 Communications	184	192	193					
0700 Computer and Information Sciences	3	4	5	482	343	*156*	*172*	*342*
0800 Education	115-145	240	382	*952*				
0900 Engineering	006-023	*53*	55	112	151-162	*111*		
1000 Fine and Applied Arts	175	182-191	194	260	*181*			
1100 Foreign Languages	130							
1200 Health Professions	61-85	113	212	921-926	426			
1300 Home Economics	131	402	*26*	425	613			
1400 Law	*30*	31	*132*					
1500 Letters	130	*120*	126	362				
1600 Library Science	32	33						
1700 Mathematics	34	*35*	*36*	112	*156*			
1800 Military Sciences	471							
1900 Physical Sciences	43	45	51-54	151	*103*	*110*		
2000 Psychology	93	*114*						
2100 Public Affairs and Services	100	101	961-965					
2200 Social Sciences	94-96	91	*92*	*116*	*121*	*122*		
2300 Theology	86	90	*133*					

Note: Italic Occupation Code are not observed in NLSY79 data.

A3 Structural Estimation

Table A4: Conditional Logit Model Structural Parameter

Major Choice	(1)	(2)	(3)
Log-likelihood	-2414.9136	-2298.5042	-2332.4301
$\hat{\theta}$ - Predictable Wage	1.544*	1.380	1.567*
	[0.851]	[0.873]	[0.867]
$\hat{\theta}\hat{\lambda}$ - Family's Realized Wage		0.156***	
		[0.040]	
$\hat{\theta}\hat{\lambda}$ - Family's Predicted Wage			0.123***
			[0.040]
$\hat{c_1}$ - Personal Taste T	0.283***	0.295***	0.282***
	[0.027]	[0.028]	[0.028]
$\hat{c_0}^j$ - Major Fixed Taste	X	X	X
# Observation	22,088	21,113	21,406
# Case	1004	995	1004
# Parameter	23	24	24

Table A5: Nested Logit Model First Level Choice

	Humanities	Soc-Sci & Business	STEM	Health	Other
τ	0.425	0.510	0.434	0.278	-0.696
	[0.112]	[0.110]	[0.103]	[0.158]	[0.470]
Female	0.930***	0.692***	-	2.036***	3.734**
	[0.193]	[0.160]	-	[0.286]	[1.461]
Black	-0.215	0.0064	-	-0.459	-0.9175
	[0.275]	[0.227]	-	[0.349]	[0.887]
Hispanic	0.128	0.046	-	0.180	-15.062
	[0.284]	[0.242]	-	[0.349]	[1126.14]
AFQT	-0.011^{**}	-0.004	-	-0.013^{**}	-0.005
	[0.005]	[0.004]	-	[0.006]	[0.015]

A4 Robustness Check

Table A6: The Impact of a Sibling's Wage on a Student's Major Choice Robustness

Dependent Variable: **Student's Major = Sibling's Occupation**	(1)	(2)	(3)	(4)
Sibling's Wage Pre-Choice			0.0866***	0.113***
			[0.0329]	[0.0364]
Sibling's Permanent Wage		0.0085		-0.0499
		[0.0280]		[0.0308]
Occupation Permanent Wage	0.240***	0.235***	0.193***	0.206**
	[0.0555]	[0.0574]	[0.0612]	[0.0643]
Ideal Occupation = Sibling's Occupation	0.412***	0.412***	0.441***	0.440***
	[0.0533]	[0.0534]	[0.0546]	[0.0542]
Same Gender with Sibling	0.0655***	0.0657**	0.0651***	0.0653***
	[0.0224]	[0.0224]	[0.0242]	[0.0242]
Education & Demographics	X	X	X	X
Observations	953	952	853	853
R-squared	0.152	0.152	0.173	0.175

Note:
1. Clustered standard errors by household in brackets, *** $p<0.01$, ** $p<0.05$, * $p<0.1$
2. Siblings' wages are measured in logged hourly wage in 2010 Dollars
3. Pre-choice window: 0-3 years before the major choice; Post-choice window: 1-4 years after the choice.
4. Control variables include gender, race, region, birth year dummies, own and sibling's education, own and sibling's AFQT score, highest education of parents, and the age when declaring the major.

Table A7: The Impact of a Sibling's Annual Income on a Student's Major Choice

Dependent Variable: **Student's Major = Sibling's Occupation**	(1)	(2)
Pre-Choice Sibling's Wage	0.0648***	0.0570***
	[0.0132]	[0.0131]
Pre-Choice Occupation Average Wage		0.202***
		[0.0363]
Sibling's Permanent Wage	-0.0126	-0.0223
	[0.0279]	[0.0272]
Ideal Occupation = Sibling's Occupation	0.411***	0.418***
	[0.0547]	[0.0538]
Same Gender with Sibling	0.0731***	0.0732***
	[0.0232]	[0.0229]
Education & Demographics	X	X
Observations	875	874
R-squared	0.181	0.202

Note:
1. Clustered standard errors by household in brackets.
2. Sibling's wage is measured by annual incomes in 2010 Dollars
3. Controls: gender, race, region, birth year, years of education, AFQT score, parents' education, and a student's the age when declaring the major.

A5 Discussion for Section 4.2

The average realized wage in group **1** is $\bar{w}_1$

$$\begin{aligned}\bar{w}_1 =& \frac{\iiint (A+M)\mathbf{1}(A+M>o)\mathbf{1}(M>0)f(M)h(o)g(A)dodAdM}{\iiint \mathbf{1}(A+M>o)\mathbf{1}(M>0)f(M)h(o)g(A)dodAdM} \\ =& \frac{\iint (A+M)H(A+M)\mathbf{1}(M>0)f(M)g(A)dAdM}{\iint H(A+M)\mathbf{1}(M>0)f(M)g(A)dAdM} \\ \doteq& \frac{X}{Y}\end{aligned}$$

The average realized wage in group **2** is $\bar{w}_2$

$$\begin{aligned}\bar{w}_2 =& \frac{\iiint (A+M)\mathbf{1}(A+M>o)\mathbf{1}(M<0)f(M)h(o)g(A)dodAdM}{\iiint \mathbf{1}(A+M>o)\mathbf{1}(M<0)f(M)h(o)g(A)dodAdM} \\ =& \frac{\iint (A+M)H(A+M)\mathbf{1}(M<0)f(M)g(A)dAdM}{\iint H(A+M)\mathbf{1}(M<0)f(M)g(A)dAdM} \\ \doteq& \frac{U}{V}\end{aligned}$$

$M = \lambda\eta_f$ and its p.d.f. is $f(M)$. η_f is the family wage shock with a normal distribution $N(0, \sigma_k^2)$. A is from another normal distribution $N(\mu, \sigma_\alpha^2)$, independently from M, its p.d.f. is $g(A)$. $H(\cdot)$ is the c.d.f. for Type I extreme value function with the scale parameter τ and location parameter 0.

To get some intuition for the above integral, recall that:

- A represents the predictable wage in log hourly rate, which is in the range of $[1.7, 2.7]$ for most students. Its mean $\mu = 2.2$
- $M = \lambda\eta_f$ is the correlated match quality component. $M = 0.1$ means a student's sibling receives a $\frac{10\%}{\lambda}$ increase in hourly wages. Students in my sample observe a $M \in [-0.5\lambda, 0.5\lambda]$.
- $f(M)$ is p.d.f. of a normal distribution $N(0, (\lambda\sigma_k)^2)$.
- According to the estimation in Section 3.5, $\lambda = 0.39$. σ_k in the data is 0.3335.
- $\int M \cdot \mathbf{1}(M > 0)f(M)dM = EM^+$ is the expectation of M at the right half of the distribution, $EM^+ = \frac{\sqrt{2}}{\sqrt{\pi}}\lambda\sigma_k = 0.1$
- $H(\cdot)$ is the c.d.f. of a Type I extreme value distribution, it is in the range of $[0, 1]$ and monotonically increasing.
- $A \pm M \in [1.4, 3.0]$ for most students in my data.

To prove $\frac{X}{Y} > \frac{U}{V}$, given that $X, Y, U, V > 0$, it is equivalent to prove $XV > YU$

Rewrite X as

$$
\begin{aligned}
X &= \iint (A+M)H(A+M)\mathbf{1}(M>0)f(M)g(A)dAdM \\
&> \iint (A+M)H(A)\mathbf{1}(M>0)f(M)g(A)dAdM \\
&= \int AH(A)g(A)dA \int \mathbf{1}(M>0)f(M)dM + \int [\int M\mathbf{1}(M>0)f(M)dM]H(A)g(A)dA \\
&= 0.5\int AH(A)g(A)dA + 0.1\int H(A)g(A)dA \doteq X'
\end{aligned}
$$

Rewrite V as

$$
\begin{aligned}
V &= \iint H(A+M)\mathbf{1}(M<0)f(M)g(A)dAdM \\
&> \iint H(1.4)\mathbf{1}(M<0)f(M)g(A)dAdM \\
&= H(1.4)\int [\int \mathbf{1}(M<0)f(M)dM]g(A)dA = H(1.4)\cdot\frac{1}{2} \doteq V'
\end{aligned}
$$

Rewrite U as

$$
\begin{aligned}
U &= \iint (A+M)H(A+M)\mathbf{1}(M<0)f(M)g(A)dAdM \\
&< \iint (A+M)H(A)\mathbf{1}(M<0)f(M)g(A)dAdM \\
&= \int AH(A)g(A)dA \int \mathbf{1}(M<0)f(M)dM + \int [\int M\mathbf{1}(M<0)f(M)dM]H(A)g(A)dA \\
&= 0.5 \int AH(A)g(A)dA - 0.1 \int H(A)g(A)dA \doteq U'
\end{aligned}
$$

Rewrite Y as

$$
\begin{aligned}
Y &= \iint H(A+M)\mathbf{1}(M>0)f(M)g(A)dAdM \\
&< \iint H(3.0)\mathbf{1}(M>0)f(M)g(A)dAdM \\
&= H(3.0) \int [\int \mathbf{1}(M>0)f(M)dM]g(A)dA = H(3.0) \cdot \frac{1}{2} \doteq Y'
\end{aligned}
$$

Notice that:

$$
\begin{aligned}
X > X' > 0 \\
V > V' > 0 \\
0 < Y < Y' \\
0 < U < U'
\end{aligned}
$$

Therefore, $X'V' - Y'U' > 0 \Rightarrow XV - YU > 0$. $X'V' - Y'U' > 0$ can be proved by following induction:

$$
\begin{aligned}
X'V' - Y'U' =& H(1.4)[0.5 \int AH(A)g(A)dA + 0.1 \int H(A)g(A)dA] \\
& - H(3)[0.5 \int AH(A)g(A)dA - 0.1 \int H(A)g(A)dA] \\
=& 0.5[H(1.4) - H(3)] \int AH(A)g(A)dA + 0.1[H(1.4) + H(3)] \int H(A)g(A)dA
\end{aligned}
$$

Remember that $H(1.4) - H(3) < 0$, $H(1.4) + H(3) > 0$, $A \in [1.7, 2.7]$, so

$$0.5[H(1.4) - H(3)] \int AH(A)g(A)dA > 0.5[H(1.4) - H(3)] \int AH(2.7)g(A)dA$$
$$0.15[H(1.4) + H(3)] \int H(A)g(A)dA > 0.15[H(1.4) + H(3)] \int H(1.7)g(A)dA$$

Therefore

$$\begin{aligned} X'V' - Y'U' &> 0.5[H(1.4) - H(3)]H(2.7) \int Ag(A)dA + 0.1[H(1.4) + H(3)]H(1.7) \\ &= 0.5[H(1.4) - H(3)]H(2.7)\mu + 0.1[H(1.4) + H(3)]H(1.7) \doteq \Delta \end{aligned}$$

where $\mu = 2.2$ according to my data, and $H(\cdot)$ is the c.d.f. of Type I extreme value function with location parameter 0 and scale parameter τ.

For any τ, I find

$$\Delta = 0.5[H(1.4) - H(3)]H(2.7) \cdot 2.2 + 0.1[H(1.4) + H(3)]H(1.7) > 0$$

Thus $\frac{X}{Y} > \frac{U}{V}$, $\bar{w}_1 > \bar{w}_2$.