

Evolution of Gender Differences in Post-Secondary Human Capital Investments: College Majors

June 27, 2011

Ahu Gemici

New York University
Department of Economics
ahu.gemici@nyu.edu

Matthew Wiswall

New York University
Department of Economics
mwiswall@nyu.edu ¹

Keywords: human capital, gender differences, college majors

JEL Codes: J16, J24, J71

¹We thank Jennifer Sallman for excellent research assistance. All errors are our own.

Evolution of Gender Differences in Post-Secondary Human Capital Investments: College Majors

Abstract

Over the past 40 years, the level of human capital investments has changed substantially for men and women. Changes in the intensive margin of college major selection have been also been substantial, as the number of graduates in humanities, social science, and teaching has declined, and the number in science, engineering, and business has increased, especially for women. However, while women are now more likely to complete a college degree than men, the distribution of college majors among college graduates remains unequal with women about 2/3 as likely as men to major in a business or science field. In this paper, we develop and estimate a dynamic overlapping generations model of human capital investment and employment decisions to understand these long-term changes in human capital investments. Our departure from the previous literature is that we separately examine college major choices, rather than aggregating these choices to the education level (e.g. college or no college). We overcome the absence of field of study information in the CPS and Census data by combining these data with auxiliary data sources which characterize the changes in field of study composition across a large number of birth cohorts. Results from counterfactual experiments show that changes in skill prices, higher schooling costs, and a reduction in the value of home for women all played an important role in the educational attainment and college major composition trends.

1 Introduction

One of the starkest changes in developed economies over the past several decades has been the increase in women’s educational attainment. In the United States, the proportion of women obtaining a college degree has increased more than 4 fold, from about 8 percent for the cohorts born in the 1920s (graduating from college in the 1940s) to about 35 percent for cohorts born in the 1960s (graduating from college in the 1980s). The rapid rise in college attainment for women has reached the point where women are now more likely than men to graduate from college.²

Less widely known is that accompanying this change in the *extensive* margin of college attendance and graduation, there were also substantial changes in the *intensive* margin of college major choice. For the cohorts born in the 1920s, the women who graduated from college obtained about 80 percent of their degrees in the humanities, social sciences, or teaching fields, and only 14 percent in science, mathematics, or engineering and 6 percent in business or economics. In contrast, college educated men born in the same years had around 50 percent of their degrees in science, mathematics, or engineering, and 15 percent in business. For the cohorts born in 1960s, the proportion of women earning degrees in science fields doubled to about 28 percent and the proportion in business tripled to about 20 percent.³

As Figure 4 shows these changes have resulted in an increase in the female-male ratio of the proportion of degrees in science and business from the 1920s to 1960s birth cohorts. But unlike the female-male ratio in college attainment, the gender gap in college major composition is still far less than parity for these recent cohorts, with women about 2/3 as likely as men to earn a degree in a science or business field than men. Incorporating this information on college major choice, we then have a more nuanced picture of the gender differences in educational attainment: while women have reached parity with men in rates of college graduation, there remains a substantial gender difference in college major choices.

To understand the evolution of these educational choices, this paper develops and estimates a dynamic overlapping generations model of human capital investment and labor supply. Our main departure from the previous literature is the way we measure

²Calculated from Census and CPS data, discussed below. See Figure 1. As Goldin, Katz and Kuziemko (2006) point out, this more recent trend represents a “homecoming” of women to college as the earlier cohorts of women, who graduated from college in the 1900s-1930s (born approximately in the 1880s-1910s), actually attended college at the same rate of men.

³See Figures 2 and 3.

human capital, making a distinction between college degrees with different majors. We define human capital skill classes by schooling years and degree, including specific college fields of study, rather than by schooling years only (as in e.g. Heckman, Lochner, and Taber (1998) or Heathcote, Storesletten, and Violante 2010), or years of schooling combined with white, blue, and pink collar occupation categories (as in Lee 2005 and Lee and Wolpin 2006). Our model explicitly incorporates college major selection as a distinct choice and allows for heterogeneity in major specific skills and tastes. Our overlapping generations model allows for non-stationary college major specific rental rates, allowing the returns to science degrees relative to humanities degrees to vary over time.

Due to data limitations, most notably that the Current Population Survey (CPS) and Decennial Census do not record college major information, economists studying long-term trends in human capital investments in the United States typically use years of completed schooling as their measure of human capital. For the college educated population, years of schooling is a substantially incomplete measure of their human capital as the various college majors chosen by college graduates represent substantial investments in specific human capital, as suggested by the large average earnings differences between individuals with different majors (compare the average earnings of an individual with a degree in the humanities versus one with a degree in engineering). To overcome the lack of long-term time series data on college majors, we turn to auxiliary data from 1993 and 2003 National Survey of College Graduates (NSCG). With the retrospective questions on college majors, the NSCG data allows us to reconstruct the date of completion and specific major of the college degrees earned for a large sample of US residents born from the 1920s to the 1960s. This dataset offers the most extensive historical coverage of trends in college major composition by birth cohort.⁴

We combine the NSCG data with the CPS, Census, and other datasets, and use the combined data to provide a fuller picture of the trends in human capital investments and as the basis of our estimation framework for the choice model. Identification of the time series for major specific skill rental rates is a key issue here given that we do

⁴Other data exists to track the college major composition of degrees earned, using administrative counts from each US college and university, collected by the HEGIS and IPEDS surveys since the mid 1960s (for graduates born approximately since the mid 1940s). However, the NSCG data has several advantages: i) it provides college major composition by birth cohort rather than for graduating classes, and therefore provides information on the lifecycle timing of college decisions, ii) the data is available for a longer span of cohorts allowing greater historical coverage, and iii) the NSCG data provides contemporaneous earnings and labor supply information linked to college major.

not observe the long-term major specific wage rates in the CPS or Census and have only a limited number of years of earnings by major from the NSCG. We show how one can use cohort differences in average wages for each calendar year (from the CPS and Census), combined with the proportion of each cohort graduating with each major (from NSCG), to identify major specific skill rental rates. Unlike previous studies that explicitly specify an aggregate production technology and use equilibrium supply and demand conditions to identify skill rental rates (e.g. Lee 2005, Lee and Wolpin 2006, 2009), we side-step the issue of specifying the technology by treating the skill prices as unknown parameters and directly estimating the non-stationary sequence of prices along with the other model parameters. This procedure avoids the considerable computational cost of computing the equilibrium for each trial vector of parameters and allows us to more robustly estimate other model parameters by avoiding mis-specifying the technology.

We decompose the across cohort changes in educational attainment and major selection into three channels: i) changes in gender neutral relative major specific skill rental rates, ii) changes in gender and major neutral post-secondary tuition rates, and iii) changes in the gender specific value of home/leisure. We find that all three channels played a quantitatively important role in the determining male and female human capital investments.

Our estimates indicate that the rental rate of science and business major specific skills increased relative to humanities skills during the 1980s and 1990s, and this shift caused higher college attendance and a shift toward science and business degrees for both men and women. Both men and women responded to these changes in skill rental rates, but, because of their lower level of home utility and higher expected future labor supply, men were more responsive than women. An increase in the cost of tuition during this period discouraged college attendance and partially offset the change in skill prices. The effect of higher schooling costs on college major composition is theoretically ambiguous, but given the distribution of skills and taste we estimate, we find that higher tuition reduced the proportion of individuals who would have completed science and business majors from entering college, which militated against the changes in skill prices favoring these fields. An important factor in the increase in female college graduation and the shift toward science and business fields was a reduction in the value of time in the home for women and higher expected future labor supply. We do not model the explicit mechanisms of the changes in home value and

higher female labor supply, and the current literature offers several possible candidate explanations, including changes in the price of home goods (Greenwood, Seshadri and Yorukoglu 2005), increase in the availability of oral contraceptive (Goldin and Katz 2002; Baily 2006), and changes in cultural norms with regard to women's participation in the labor force (Fernandez, Fogli, and Olivetti 2004).⁵ Our estimates are in line with these findings and we show that these types of mechanisms can also account for a shift in the college major composition for women.

Our research builds on previous studies that model college major choices. A number of papers have examined field of study choices in equilibrium models, focusing on particular fields such as engineers, lawyers, or teachers (Freeman 1971, 1976a, 1976b; Siow 1984; Zarkin 1985). Our framework generalizes these studies by jointly modeling the lifetime sequence of education and labor supply choices, examining multiple fields rather than one field in isolation, and incorporating heterogeneity in skills and tastes. Later work has studied field of choices using single cohort, partial equilibrium models, incorporating such factors as heterogeneity in earnings and tastes, lifecycle earnings growth, earnings risk, and learning about abilities in college (Blakemore and Low 1984; Berger 1988; Eide and Waehrer 1998; Arcidiacono 2004; Shore and Saks 2005; Montmarquette, Cannings, and Mahseredjian 2005).⁶ We complement these studies, which focus on a particular point in time estimate of the college major choice process, by estimating a multiple cohort model which allows us to study the non-stationary features of the economy that can explain the long-term changes in college major composition.

Our paper proceeds as follows. First we provide some descriptive evidence on trends in college major composition and earnings and labor supply differences for individuals with different majors. We then outline our choice model and layout the identification and estimation strategy. The final sections discuss the model estimates and conclude with a decomposition of the trends in educational attainment for men and women.

⁵See Goldin (2005) and Goldin, Katz, and Kuziemko (2009) for a summary of other possible factors behind the growth in women's labor force participation and educational attainment, including changes in divorce rates and workplace discrimination law. In addition, Charles and Luoh (2003) argue for the importance of earnings risk differences.

⁶More recent work studies uses expectations data to study college major choices (Zafar 2011; Arcidiacono, Hotz, and Kang 2011; Wiswall and Zafar 2011). Other research, noting the large difference in average earnings across fields, studies to what extent college majors can help explain gender gaps in earnings (Rumberger and Thomas 1993; Paglin and Rufolo 1990; Eide 1994; Brown and Corcoran 1997; Weinberger 1998; Black, Haviland, Sanders, and Taylor 2000; Machin and Paglin 2003).

2 Descriptive Evidence

This section presents some descriptive information on trends in human capital compositional, differences in labor supply and earnings by field of study, and the relationship between the gender gap in earnings and field of study composition.

2.1 Trends in Educational Attainment

Figure 1 presents college graduation trends for men and women for birth cohorts born in 1920 through 1969. We restrict the sample to individuals aged 25-60. This figure uses the years of schooling variable available in the Census and CPS and examines the fraction of the population with more than 16 or more years of schooling, which we define as a college degree following the previous literature. Data description details are below. For men, the fraction with a college degree increased rapidly, from about 13 percent with a college degree in the 1920 cohort to 31 percent for the 1950 cohort. The rate of college attainment by birth cohort slowed through the 1950s cohorts but recovered to nearly the 1950 peak by the 1969 cohort. For women, the trend in post-secondary education mirrors the male trend, but with a more steady rise. The proportion of women with 16 or more years of schooling increased from 7 percent for the 1920 cohort to 33 percent for the 1969 cohort.

2.2 Trends in Field of Study Composition

Because the CPS and Census lack field of study information, we turn to the National Survey of College Graduates (NSCG) to document the trends by birth cohort in the underlying college major composition of the college graduate population. Figures 2 and 3 provide the field of study composition for men and women, respectively. We aggregate the college majors into three bachelor degree categories: i) science, mathematics, and engineering (“science”), ii) business and economics (“business”), and iii) humanities, social sciences, and teaching (“humanities”). Data construction and field aggregation details are provided below. For the 1920 birth cohort, most of which graduated from college in the 1940s, 84 percent of all college degrees earned by women were in humanities, social sciences, and teaching fields, 11 percent in science, and 5 percent in business. For the men born in the same year, the field of study composition was very different, with 51 percent of degrees earned by men in science, engineering, and mathematics, 19 percent in business and economics, and the remaining 29 percent in

humanities, arts, and education fields. College educated men born in 1920 were nearly 5 times more likely to major in a science, mathematics, or engineering field than a college educated women born the same year, and nearly 4 times more likely to major in a business or economics field.

Examining the Figures 2 and 3 we see that up until the 1950s cohort there was a decline in the fraction of men graduating with science, mathematics, or engineering degrees, and a slight rise in the fraction of men graduating in humanities, social sciences, and teaching. For women, there is an opposite pattern during this period, with a decline in the fraction of degrees in humanities from 84 percent to 74 percent from the 1920 to 1950s cohort.

The largest trend break, for both men and women, occurred for the cohorts born in the 1950s (who mainly graduated in the 1970s). For men, from the 1950 to the 1969 birth cohort, the proportion of degrees in science, mathematics, or engineering grew from 43 percent to 55 percent. During this period there was a concomitant fall in the proportion of men who graduated with humanities degrees, from 38 percent to 25 percent. The fraction in business also increased, but only slightly from 18 to 20 percent. It appears that beginning in the 1970s, male undergraduate students began to switch their majors away from humanities fields to science fields, and to a lesser extent business fields.

The trend for women is similar to that for men, but the decline in the proportion of humanities majors, which were the predominant degree fields for women graduates prior to the 1970s, is even larger. For the cohorts born starting from the 1950s to the 1960s, the proportion of science graduates increased from 18 to 28 percent and the number of graduates in business increased from 5 percent to 20 percent, a level approaching that of men. It is important to note that while the field of study composition among college degree holders changed considerably, at the same time there were more female and male college graduates, hence the *number* of humanities degrees actually increased over this period.

2.3 Gender Differentials in Human Capital Investments

Figure 4 directly compares the gender differences in the trend in the extensive margin of college completion and the intensive margin of the ratio of female-to-male proportions of undergraduate degrees in non-humanities fields (i.e. science and business fields). In the Figure, there are three salient patterns. First, field of study composition gender

differences are larger (more unequal) than for college attainment. Examining only extensive margin human capital differences greatly understates the gender differences in field of study composition. Second, the trend for the female-to-male ratio for both the extensive and intensive education margins are generally increasing across birth cohort, with the steepest rise for the cohorts born in the 1950s and mid-1960s, and a flattening, and even declining ratio for the later 1960s cohorts. Third, while the ratio of female-to-male college graduation has reached parity, and even exceeded parity by the mid 1950s cohort, the gender ratio in field of study composition is still far below 1 for even the most recent cohorts. Even for the most recent cohorts in our data, women are only about $2/3$ as likely as men to earn a college major in a science or business field.

2.4 Labor Supply by Field

How much does labor supply differ at the intensive margin of field of study? Table 1 uses the NSCG to document these differences by pooling labor supply and earnings information for the college graduate population surveyed in 1993 and 2003. These statistics are for the full sample of college graduate respondents, aged 25-59 and born between 1920-69. The data on labor supply and earnings is for the reference years 1989 (for the 1993 NSCG) and 2002 (for the 2003 NSCG), as detailed below in the data section.

As Table 1 demonstrates, labor supply and earnings vary widely across fields. We use a simple indicator for full time/full year status of working more than 2000 hours per year. For men, the fraction working full time/full year is 83 percent for those with degrees in science, engineering, or mathematics, 84 percent with a degree in business or economics, and 73 percent in humanities, social sciences, and teaching. For women, the fraction working full time/full year varies from 51 percent with a science degree, 54 percent with a business degree, and 41 with a humanities degree. While college educated men are more likely to work full time/full year in all fields, there is still considerable variation in labor supply across fields.

Panel A of Table 2 provides age adjusted differences in labor supply across fields, using a regression of full time/full year status on field of study indicators for science and business (omitted category is humanities) and a full set of indicators for each age (by year). Recall that the cohort pattern in educational attainment is quite different between men and women, and this can lead to different age distributions for men and

women across fields at a given calendar period. The regression results reveal that the relative differences in labor supply across fields are quite similar for college educated men and women. Men are 9.2 percent more likely to work full time/full year with a degree in science (relative to humanities), and 10 percent more likely with a degree in business. Women, on the other hand, are 8.7 percent and 11.9 percent more likely to work full time/full year with a science or business degree (relative to humanities), respectively.

2.5 Earnings by Field

Table 1 documents differences in earnings by field, where we use annual earnings for all workers aged 25-59 who worked full time/full year. Earnings are in 2002 USD. Further details are in the data section.

Average annual earnings vary considerably across fields. For full time/full year men whose college degree is in a science and business fields, average annual earnings are \$86,789 and \$81,963, respectively, while the average earnings for men with a degree in a humanities, social sciences, or teaching field is \$72,198. On average, women earn less overall than men in each of the three field of study categories. But, as for men, there are considerably differences in annual earnings across field. For full time/full year women, average annual earnings are \$63,873 with a degree in science, \$57,675 with a degree in business, and \$52,590 with a degree in humanities.

To control for age differences, Panel B of Table 2 reports regression results from a regression of log annual earnings on field of study and a full set of age indicators. Men have larger differences in age adjusted earnings across fields than women. Average male earnings are 24 percent higher in a science field, and 16 percent higher in a business field, relative to humanities. For women the differences are 19 percent and 10 percent. For comparison, note that the return to a year of schooling, estimated from a log wage regression on years of schooling using CPS data, is typically about 6 to 8 percent during this period. For men, a field of study earnings difference of 25 percent in science (relative to humanities) is therefore similar in magnitude to about 3 to 4 additional years of schooling.⁷

⁷These figures are of course descriptive in nature, as selection based on unobserved skills can bias the estimated returns to years of schooling and fields of study. Our estimated model discussed below directly incorporates selection based on unobserved skills.

2.6 Fields of Study and the Gender Earnings Gap

The NSCG data demonstrate that there were significant changes in the composition of college majors among those with post-secondary degrees. In general, for the most recent generation the shift was from lower earning fields (humanities, social sciences, teaching) to higher earning fields (science, engineering, and business), for both men and women, although the change was substantially larger for women. During this same period the gender gap in earnings closed as well. We next examine how closely related the change in field of study composition is to the change in the gender gap in earnings.

To examine birth cohort level changes in earnings, we first estimate the birth cohort effects from the the following log wage regression:

$$\ln w_{itc} = \gamma_t + \delta_c + \epsilon_{itc}$$

where the γ_t are time specific intercepts and the δ_c are cohort specific intercepts. We estimate this regression for college educated men and women separately using the combined CPS and Census data for the years 1949-2008, with the sample restricted to cohorts born 1920-1969 and individuals aged 25-59. We then construct a female-male log wage ratio in earnings cohort effects from the estimated δ_c cohort intercepts: $\delta_c(\text{female}) - \delta_c(\text{male})$.

Figure 5 graphs the female-male log wage ratio against the female-male ratio in the proportion of science and business degrees. There is a clear positive and significant relationship between the college educated earnings gender gap by birth cohort and the proportion of degrees in higher earning science and business majors. The regression line for this relationship has a slope of 0.55 (0.041 standard error) and R-squared value of 0.79. While this relationship cannot be given a causal interpretation, we take this correlation as suggestive that college major composition is strongly related to gender earnings gaps.

3 Model

3.1 Overview

The economy consists of a single sector and overlapping generations. Time is discrete and individuals make decisions over a finite horizon. Each period or age for an indi-

vidual is indexed $a = 16, 17, \dots, A$, where the initial age of decision making is age 16 and age A is an exogenous retirement age.

At each age, individuals make decisions regarding labor supply and human capital investments based on the expected future labor market returns and their own heterogeneous preferences for working in the labor market and schooling. Our major point of departure from the existing literature is that our formulation of human capital skill classes is by schooling years and degree, including specific college fields of study, rather than by schooling years only (as in e.g. Heckman, Lochner, and Taber 1998 or Heathcote, Storesletten, and Violante 2010), or years of schooling combined with white, blue, and pink collar occupation categories (as in Lee 2005 and Lee and Wolpin 2006).

3.2 Choice Set and Preferences

At each age a , individuals choose from a set of mutually exclusive activities: enroll in school to obtain degree d from the set $d = 1, \dots, D$, work in the labor market, or stay at home. The degree set includes high school drop-outs ($d = 0$), high school degrees ($d = 1$), 2 year college degrees ($d = 2$), and specific college degrees defined in 1 of 3 college major categories ($d = 3, 4, 5$). We aggregate college majors into 3 categories: i) science, mathematics, and engineering, ii) business and economics, and iii) humanities, social sciences, and teaching, where the last category encompasses all remaining fields. Individuals at age 16 start as high school drop-outs (with degree $d = 0$) and then decide whether to finish high school and earn any post-secondary degrees.

The flow utility of an individual of gender g , type k , and age a in period t from choosing each alternative j is:

$$u_t(a) = \begin{cases} \gamma_d(k) + \gamma_6\tau_{d,t} + \varepsilon_{d,t}^1(a) & \text{if go to school for degree } d \\ \gamma_7w_{d,t}(k, g, a) + \varepsilon_t^2(a) & \text{if work} \\ \gamma_8(k) + \gamma_9(g) + \gamma_{10}(g)c_t(g, a) + \gamma_{11}(g)(t - a) + \varepsilon_t^3(a) & \text{if stay at home} \end{cases}$$

The consumption value of going to school for a particular degree d has a time invariant and type specific component $\gamma_d(k)$, $d = 1, \dots, D$, a component that depends on a time varying tuition cost, $\tau_{d,t}$, as well as a stochastic component, $\varepsilon_{d,t}^1(a)$. Tuition costs are constructed using data on average costs of schooling for 4 year and 2 year degrees in the United States, as detailed below. We assume tuition costs are the same for all individuals who attend school in year t , but allow the consumption value of

school attendance for each degree to be individual type specific. The type specific component reflects the cost of study effort an individual incurs when completing a degree and the consumption value individuals receive because they enjoy studying a particular subject.

For an agent whose highest degree obtained at age a is d , the utility from working is determined by the wage rate, $w_{d,t}(k, g, a)$, and a stochastic component $\varepsilon_t^2(a)$, which reflects idiosyncratic (dis-) utility of working. As explained below, wages are degree, calendar time, age, and type specific, reflecting heterogeneity in skills across types and age, and calendar time varying rental rates of skill for each degree.

The consumption value of staying at home has a type specific and time invariant component, denoted by $\gamma_8(k)$; a gender specific intercept, $\gamma_9(g)$; and a component that depends on the average cohort and age specific fertility rates for men and women, given by $c_t(g, a)$. $c_t(g, a)$ is estimated from the CPS and Census data using the average number of children under 5 years old at a at period t . We allow the value of staying at home to depend on the number of children under 5 years old in order to reflect the possibility that the value of home production changes with the presence of young children. In this way, we allow the home value to be age and cohort varying due to changes in cohort specific fertility rates. Note that this term varies by the individual's age, as well as by cohort, reflecting how the value of home changes through the life-cycle because of birth timing and spacing. We also allow the extent to which fertility changes the value of leisure to differ by gender, denoted by the coefficient $\gamma_{10}(g)$. We take the fertility rate changes to be exogenous to the model and our model examines labor supply and human capital decisions relative to these changes. In addition to the changes in value of leisure induced by changes in fertility rates, we also allow the value of leisure to change by year of birth, where birth cohorts are indexed by $t - a$. $\gamma_{11}(g)$ reflects the extent to which the value of leisure changes by birth cohort, and this trend slope is allowed to vary by gender.

3.3 Wages and Skill Production Technology

An individual of type k and gender g , who has obtained degree d supplies $s_{d,t}(k, g, a)$ to the labor market if he/she decides to work. The skill supply of a type k individual at age a and time t is given by:

$$s_{d,t}(k, g, a) = \exp(\alpha_d(k) + \beta_1 x_t(a) + \beta_2 x_t(a)^2) \quad (1)$$

where $\alpha_{1,d}(k)$ is the degree specific intercept. $x_t(a)$ is the total labor market experience at age a and period t . The skill level of an individual is determined by the highest degree the individual currently holds. Therefore as individuals earn college degrees, they switch from supplying high school labor to college labor in a specific field. For an individual of type k , whose highest degree is d , the wage offer $w_{d,t}(k, g, a)$ at period t and age a is given by:

$$w_{d,t}(k, g, a) = r_{d,t}s_{d,t}(k, g, a) \quad (2)$$

where $r_{d,t}$ is the period t rental rate of skill degree class d , and $s_{d,t}(k, g, a)$ is the level of accumulated degree d skill.⁸

3.4 Household's Problem

The decision model starts at age ($a = 16$). There are three non-stationary elements to the model: skill prices, tuition costs, and home values. Individuals are assumed to have perfect foresight regarding the future evolution of these components but there is uncertainty about the future realizations of the stochastic shocks $\varepsilon_t(a) = [\varepsilon_{1,t}^1(a), \dots, \varepsilon_{D,t}^1(a), \varepsilon_t^2(a), \varepsilon_t^3(a)]$. At each age, after realization of the current period shocks, individuals choose between going to school to earn degree d , working in the labor market, or staying at home. The state space of an agent at age a in period t includes the present and future sequence of degree-specific rental rates, $R_t, R_{t+1}, \dots, R_{t+A-a}$, where $R_t = \{r_{1t}, \dots, r_{Dt}\}$, the present and future cost of schooling T_t, T_{t+1}, T_{t+A-a} , where $T_t = \{\tau_{1,t}, \dots, \tau_{D,t}\}$, and the present and future value of home, which depends on the age and cohort specific fertility rates $\{c_t(a), c_{t+1}(a+1), \dots, c_{t+A-a}(A)\}$, and an additional cohort trend. Other state variables include the individual's type k (which determines her degree specific skill endowments schooling and leisure tastes), total labor market experience at t $x_t(a)$, highest degree already obtained $d_t(a) \in \{0, 1, \dots, D\}$, and whether the agent was in school for degree d in previous period, $y_{d,t}(a)$.⁹

The vector of state variables $\Omega_t(a)$ is then,

⁸Gender based discrimination in the labor market would be reflected in gender differences in skills. Since skills are not separable identified from wages, we cannot distinguish between differences in skills and discrimination conditional on skill.

⁹Note that the entire sequence of future sequences of skill prices, home utility, and schooling costs are not relevant to decision making in period t ; only these sequences from age a until retirement at age A .

$$\Omega_t(a) = [k, x_t(a), y_{d,t}(a), R_t, R_{t+1}, \dots, R_{t+A-a}, \\ c_t(a), c_{t+1}(a+1), \dots, c_{t+A-a}(A), T_t, T_{t+1}, \dots, T_{t+A-a}, \varepsilon_t(a)].$$

where $\varepsilon_t(a)$ is the vector of idiosyncratic shocks to preferences. The Bellman equation formulation of the dynamic problem is then

$$V(\Omega_t(a), a) = \max_{h_t(a), q_t(a)} u_t(a) + \delta E[V(\Omega_{t+1}(a+1), a+1) | h_t(a), q_t(a)], \quad (3)$$

where $\Omega_{t+1}(a+1)$ is updated according to the current period labor choice $h_t(a)^*(\Omega_t(a)) \in \{0, 1\}$ and schooling choice $q_t(a)^*(\Omega_t(a)) \in \{0, 1, \dots, D\}$. $h_t(a)^*(\Omega_t(a)) = 0$ and $q_t(a)^*(\Omega_t(a)) = 0$ indicates optimal choices. Agent expectations are with respect to the future distribution of preference shocks: $F(\varepsilon_t(a))$. We assume the random shocks to alternative specific utilities are independently and identically distributed across individuals and over time:

$$E[V(\Omega_t(a), a) | h_t(a), q_t(a)] = \int V(\Omega_t(a), a) dF(\varepsilon_t(a))$$

Initial conditions are set such that individuals have not obtained any high school, college, or graduate degree at age $a = 16$. Labor market experience is 0 at age 16, $x_t(16) = 0$. Labor market experience is accumulated as $x_{t+1}(a+1) = x_t(a) + h_t(a)^*(\Omega_t(a))$. We assume that individuals cannot simultaneously work and attend school, and we rule out part-time schooling or part-time working during a calendar period. Given that the data limits us to examining completed degrees, we assume that individuals who enter school for an undergraduate and graduate degree finish these degrees in a given number of periods. The decision to enter school to obtain a degree therefore becomes a temporary absorbing state. Also, individuals cannot study for more than one degree simultaneously. We do allow for individuals to enter school at any age (up to a maximum age of 40) to complete a degree. This flexibility allows the model to capture older individuals returning to school and completing degrees when the expected returns are higher.

3.5 Equilibrium Conditions

An important difference between our model solution and previous research is that we do not model the equilibrium determination of skill prices but instead treat skill

prices as unknown parameters. We discuss identification of skill prices below. Our treatment of skill prices as free parameters differentiates this study from two types of previous research. Unlike previous research (e.g. Katz and Murphy 1992 and Krusell, Ohanian, Rios-Rull, and Violante 2000), we do not assume skill prices are exactly equal to observed wages, but instead allow for heterogeneity and endogeneity in skill acquisition, which creates a “wedge” between observed wages and skill prices. On the other hand, unlike other studies which explicitly model the equilibrium in skill prices (e.g. Lee and Wolpin 2005; Heathcote, Storesletten, and Violante 2010), we do not model the equilibrium explicitly. As in these studies, modeling the equilibrium in the market for skills would require specifying an aggregate production technology and its evolution over time. We side-step the assumptions required to close the model in this fashion by treating the equilibrium prices as parameters directly and focusing on the identification of these parameters, along with the other model parameters, using the existing data. The advantage of our method is then two-fold: i) we can estimate the model parameters more robustly by avoiding mis-specifying the technology and equilibrium of the market, and ii) we avoid the typically computationally costly calculations involved in computing the equilibrium. The disadvantage of our approach is that we cannot compute equilibrium counterfactual experiments directly since we do not model or estimate the necessary components this would require. However, our model estimates do allow us to meet our main goal: decomposing the changes in human capital investments into components by prices and other non-stationary features of the economy.

3.6 Sources of Changes Gender Differences in Human Capital Investments

Using the modeling framework we have laid out above, we can now discuss the various avenues the model allows for gender differences in human capital investments. Men and women in the model differ on several dimensions: skills, tastes for schooling, and the value of the home alternative. Among the non-stationary elements of the model, skill prices and tuition costs are gender neutral, while the value of home production, which depends on gender, age, and birth cohort specific fertility rates, is gender specific. We discuss the role of each of these modeling elements in turn:

1) Skill prices: Changes in relative skill prices alter the relative return to various human

capital investments. However, because men and women can have different levels of skill endowments, changes in the price of skill per unit of skill can affect the labor market return to human capital differentially by gender. In addition, because women have a higher utility from home, and therefore expect to work less over their lifetime, any wage gains or losses from choosing a particular degree do not translate into as large total lifetime utility gains or losses as it would for men.

2) Tuition costs: Higher tuition costs for post-secondary education reduces college graduation rates. However, the effect on college major composition is ambiguous given that the tuition is field of study neutral, with a single cost for all degrees. On the one hand, higher tuition costs can cause individuals to forgo completing majors that do not offer a high enough labor market return to justify the upfront cost. This would predict that higher tuition costs would shift the composition of college majors toward science and business majors, and away from humanities majors, because the science and business majors have higher pecuniary returns. However, on the other hand, the effect of a higher tuition on the overall composition of college majors depends on the distribution of skills and taste that comprise the selection into college. Higher costs of college can disproportionately increase the opportunity cost of college for individuals with an absolute advantage in all degrees (e.g. individuals with high innate general ability). If the remaining individuals have a high comparative advantage in humanities majors (or high comparative tastes for humanities), then these types would be less sensitive to increases in college costs. Higher college costs would shift the college graduate population toward a higher proportion of individuals who would major in humanities as the individuals who would major in science or business forgo college altogether. The effect of tuition costs on the gender composition of college majors then depends on the differences in male and female distributions of major specific skills and tastes.

3) Home value: A reduction in the value of home has both short-term and long-term implications for human capital investments. In the short-term, a reduction in the value of home lowers the opportunity cost of schooling. In the long-term, a reduction in the value of home increases expected labor supply and the lifetime gains from human capital investment. On the one hand, as the home value for women declines relative to that for men, women become more similar to men in terms of the importance they put on the monetary returns to each degree. This can cause more women to complete college degrees and to earn majors in higher earning fields such as science and business as they

anticipate a longer working career in which to reap the returns from these investments. However, a lower home value need not necessarily increase the proportion of college graduate women in science and business fields if the marginal woman has higher tastes or skills in humanities fields. In this case, a reduction in the value of home could shift the composition of women graduating from college toward women who have a comparative advantage in humanities fields, offsetting or even reversing the effect of higher expected labor supply on major choice.

4 Data

4.1 CPS

We use the 1968-2003 March Current Population Survey (CPS) data, covering educational attainment, labor supply, and earnings for the 1967-2001 period. We use the sample of individuals aged 16-65 in all years. To create a common educational attainment measure, we use years of schooling (prior to the 1992 CPS) and degrees obtained (1992 CPS and after) to classify individuals into 4 degree groups: high school drop-out (less than 12 years of schooling or no high school diploma), high school graduate (12 years of schooling or high school diploma), some college (13-15 years of schooling or some associate level degree), college graduate (16 or more years of schooling or at least a college degree).

To maintain comparability across years, we use the intervalled weeks worked variable. Weeks worked during the previous year were reported in 7 categories, which we recode to a single measure of weeks: 0 weeks (0), 1-13 weeks (10), 14-26 (20), 27-39 (33), 40-47 (44), 48-49 (48.5), 50-52 (52). We use hours worked last week as our measure of weeks worked. Observations with missing weeks worked last year or hours worked last week are dropped. Annual hours are calculated as hours x weeks. We define full time/full year status as individuals who worked at least 2,000 annual hours.

Annual income is taken from annual pre-tax wage and salary income. These values are topcoded, and the topcode varies across years: topcode value is \$50,000 until 1981 CPS, topcode of \$75,000 for 1982-84 CPS, and topcode of \$99,999 for later CPS. We assign a value of 1.5 times the topcoded value for any observations topcoded. Annual earnings are deflated using the CPI-U and reported in 2002 USD. We exclude all observations who report working positive annual hours but have hourly earnings

(constructed from annual earnings / annual hours) less than \$3 per hour or more than \$200 per hour in 2003 \$.

4.2 Census

To extend the information provided by the CPS further back in time, we use the 1940, 50, and 60 US Decennial Censuses, which provide information on the years 1939, 49, and 59. The variable and sample construction is the same as used with the March CPS data. One exception is that hours worked last week in the 1960 Census is given in interval values, unlike the remaining data. We use the average hours worked in the 1950 Census by these same interval categories to impute hours worked in the 1960 Census.

4.3 NSCG

Because the main source of labor force data for the United States, the Census and the CPS, do not ask respondents for information on fields of study in college, we supplement these datasets with the National Survey of College Graduates (NSCG) for 1993 and 2003, which provides information on the field of study of degrees acquired.¹⁰ The 1993 and 2003 NSCG samples were taken from the 1990 and 2000 Census samples, respectively. The NSCG samples were limited to respondents who reported in the Census having earned at least a baccalaureate or higher degree and were age 72 or younger by the time of the Census. The data collections were intended to be nationally representative of all college graduates currently residing in the United States, regardless of citizenship.

For both the 1993 and 2003 NSCG surveys, the survey instrument asks respondents to list up to three baccalaureate or higher degrees in the following categories: i) their most recent degree, ii) their second most recent degree, and iii) their first bachelor degree, if not previously reported. For each degree, respondents were instructed to record the month and year when the degree was earned, and the first and second major field for the degree. To record their major fields, respondents were instructed to write

¹⁰Several data sets, such as the National Longitudinal Surveys (Original Cohorts), the National Longitudinal Survey of the Class of 1972, the National Longitudinal Survey of Youth, and the High School and Beyond surveys include detailed information on college graduates as part of a representative sample of an entire cohort of Americans. However each of these surveys have too few college graduates to analyze college majors in any detail as the sample of college graduates numbers at most a few thousand.

the major and record one of about 150 different field of study codes included with the survey. These codes were identical across the 1993 and 2003 surveys. We aggregate degrees into 3 categories: i) science, mathematics, and engineering, ii) business and economics, and iii) humanities, social sciences, and teaching, which encompasses all remaining fields. The Data Appendix describes the aggregation of fields of study used in the analysis below.

With non-response rates of about 73 and 63 percent, the initial sample sizes are 148,905 and 100,402 individual-level observations for the 1993 NSCG and 2003 NSCG surveys, respectively. After excluding about 4.5 percent of observations with missing and nonsensical information (see Appendix), the usable combined sample is 238,344 observations.

We use the NSCG data in two ways: i) to provide a long-term dataset by birth cohort to track field of study composition of college degrees, and ii) to provide contemporaneous earnings and labor supply information by age and field of study for the reference dates of the NSCG. For the first purpose, we use the year of birth of respondents to create representative samples of the proportion of each birth cohort that completed a college degree in each of the college majors. We select cohorts that were between the ages 35 and 65 in the reference years 1989 (1990 Census and 1993 NSCG) and 2002 (2003 NSCG).¹¹ Our total sample has respondents born between 1924 (aged 65 in 1989) and 1967 (aged 35 in 2002). Each cohort sample has between 1,000 and 8,000 individual observations, where we have more observations for cohorts that appear in both NSCG surveys. As long as there are no differential death rates by field of study, each birth cohort sample can be used to consistently estimate the proportion of the cohort that completed a particular degree. From the information on the date at which each degree was earned, we are also able to construct a panel at the birth cohort level which tracks the age at which each undergraduate degree was earned by each respondent. This allows us to document undergraduate degrees earned at later ages by individuals returning to college or entering college later in life.

We also use the NSCG data as a standard cross-sectional dataset on earnings and labor supply. For the 1993 NSCG, we have available the respondent's long-form earnings and labor supply Census information linked to the NSCG questions about field of study. For the 2003 NSCG, we have similar information for 2002 reference year col-

¹¹Unlike the 1993 NSCG, the publicly available 2003 NSCG files do not contain 2000 Census information. Instead we rely on the 2003 NSCG survey information for reference year 2002.

lected internally as part of the NSCG survey. We construct earnings and labor supply information following the same procedures as with the CPS data. Because the surveys were not intended to be used in a retrospective fashion, as we do here, the survey instrument does not ask respondents to report retrospective information about past employment, wages, and the timing of major life course events such as marriage and children.

4.4 Fertility

We construct a measure of fertility by using the reported number of children under 5 years of age in the CPS and Census. We construct the average number of children under 5 by birth cohort, age, and gender to parameterize the non-stationary home value term discussed above. Since some cohorts are missing this variable for some years (for years between Census years), we construct a smoothed fertility measure using a regression of the log number of children under 5 year of age on i) a linear spline in birth cohort with nodes at 10 year intervals from 1910 to 1970, ii) a linear spline in age with nodes in 4 year intervals from age 16 to 40, iii) all interactions of the birth cohort and age splines. The predicted values of this regression form our estimate of age and cohort specific fertility rates. To eliminate some outliers, we additionally impose the restriction that the number of children under 5 years of age is 0 for men and women 47 years of age or older.

4.5 Tuition

We construct a measure of the annual tuition cost for 2 year (some college) and 4 year (college undergraduate) degrees using average tuition rates collected by the National Center for Education Statistics (NCES). This data derives from two surveys of the population of colleges and universities in the United States, the Higher Education General Information Survey (HEGIS) for academic years 1965-66 through 1985-86, and the Integrated Postsecondary Education Data System (IPEDS) for later years. These data provide average tuition for public colleges and universities for the years 1965-, and for private and public colleges and universities for 1977-, where years in the surveys refer to the previous year. “Tuition” refers to total educational expenses, including “required fees” and living costs (“room and board”).

To construct the tuition series, we first take the individual component tuition series

for undergraduate degrees (public and private 4 year) and project each series backward using the trend linear slope in (log) tuition for the last 10 years of data. With this series in hand, we then create an average tuition level across all types of college and universities by weighting the public and private tuition levels for each year by the fraction of degrees earned in public and private institutions for that year. We created the tuition series for 2 year degrees by multiplying the 4 year time series by the ratio of average tuition for 2 year to 4 year institutions across the 1990-2002 period.

5 Econometrics Issues

5.1 Identification

One of the main identification challenges is identifying skill rental prices for specific degree skill classes when information on fields of study is not collected in the CPS or Census data, the main sources of long term earnings information for the US. We approach this issue by combining CPS and Census data with additional data from the NSCG, where the NSCG provides information on the year when specific degrees were earned for many birth cohorts. However, this data provides degree specific wage information for only a limited number of years (just 2 cross-sectional years). We therefore cannot form a long time series on wages by post-secondary degree field.¹²

We show that the combination of birth cohort specific wages from the CPS and Census with the major composition of the college graduates in these cohorts from the NSCG identifies the time series of average wages by college major category. To illustrate our identification approach, we analyze a simplified version of the model. Our model of wages posits that a wage offer for an individual belonging to birth cohort c is $w_{dt}(c) = r_{dt}s_{dt}(c)$, where r_{dt} is the degree d specific skill rental rate and $s_{dt}(c)$ is the level of the individual's skill in degree d . For convenience, we index wages and skill by cohort, rather than age, but for a fixed calendar period t , birth cohort defines age at t . For our simplified identification analysis, we set $s_{dt}(c) = 1$ for all individuals and all degrees, and ignore differences in the distribution of skills within cohorts who complete each degree that arise from endogenously accumulated labor market experience and self-selection into the degree. We further assume all individuals

¹²Other US surveys, such as various surveys from the National Longitudinal Surveys (NLYS 1979 cohort, NLS Original Cohorts) and the Recent College Graduate surveys, provide some information on earnings by degree field, but for much smaller samples and for a limited range of cohorts and time periods.

work, and hence the observed distribution of earnings reflects the actual distribution of wage offers. All of these assumptions are relaxed in our more general model in which we estimate a model for labor supply and human capital investments where individuals endogenously accumulate degrees and experience. Identification of the skill levels, up to a normalization discussed below, is a straightforward application of sample selection methods.¹³

The following Lemma establishes the identification result in our simplified modeling setup:

Lemma 1 *If we observe i) average wages $E[w_t(c)]$ for each cohort, ii) the proportion working with degree d $p_{dt}(c) > 0$ for cohorts $c = 1, \dots, C$ and $d = 1, \dots, D$, where $\sum_d p_{dt}(c) = 1$, and iii) the number of non-retired cohorts in period t (C) is at least as many as the number of degrees (D), $C \geq D$, then we can identify rental rates, r_{1t}, \dots, r_{Dt} .*

Proof Average wages for cohort c in period t is given by

$$E[w_t(c)] = \sum_{d=1}^D p_{dt}(c) r_{dt}$$

For C non-retired cohorts in period t , we then have the following system of C equations:

$$\begin{aligned} E[w_t(1)] &= \sum_{d=1}^D p_{dt}(1) r_{dt}, \\ &\vdots \\ E[w_t(C)] &= \sum_{d=1}^D p_{dt}(C) r_{dt}. \end{aligned}$$

With $C \geq D$, there is at least one sequence of rental prices r_{1t}, \dots, r_{Dt} that solves this system of equations. We can identify a unique sequence of rental rates by considering only $C = D$ wage moments. QED

¹³In this simplified setup, the skill rental rates are equal to (unconditional) average wages: $r_{dt} = E[w_{dt}]$, where $E[w_{dt}]$ is the average wage in period t for degree d across all birth cohorts. Hence, identification of skill rental rates by college major is equivalent to identification of average wages by college major.

A necessary condition for identification of the prices is the availability of the data from the NSCG on each cohort's college major composition. This data allows us to vary the known composition of each cohort relative to cohort specific wages, identifying the across cohort common price in a given year.

Example As an example, consider the case of two degree skill groups ($d = \{A, B\}$), e.g. humanities and science degrees, and two birth cohorts ($C = \{1, 2\}$). In this case, we have a system of two equations:

$$E[w_t(1)] = p_{At}(1)r_{At} + p_{Bt}(1)r_{Bt},$$

$$E[w_t(2)] = p_{At}(2)r_{At} + p_{Bt}(2)r_{Bt},$$

and two unknown skill prices r_{At} and r_{Bt} . Solving this system of equations, the ratio of degree specific rental prices is given by

$$\tilde{r}_t = \frac{r_{At}}{r_{Bt}} = \frac{p_{Bt}(1)E[w_t(2)] - p_{Bt}(2)E[w_t(1)]}{p_{At}(2)E[w_t(1)] - p_{At}(1)E[w_t(2)]},$$

The level of skill rental price r_{Bt} can then be identified as

$$r_{Bt} = \frac{E[w_t(1)]}{p_{At}(1)E[w_{At'}(1)]\tilde{r}_t + p_{Bt}(1)E[w_{Bt'}(1)]},$$

and the level of skill rental price $r_{At} = \tilde{r}_t r_{Bt}$. For this case, with $C = D$, we have a unique solution.

5.2 Empirical Model and Model Solution

The model is defined up to the distribution of types and the distribution of preference shocks. We assume there is a finite number of types, which we set at 5 types in the estimation given the lack of substantial improvement in within sample fit with the addition of a sixth type. The distribution of types is stationary but differs by gender. The probability masses are given by $\pi(k, g)$, where k indexes type and g indexes gender. We assume the type distribution support is the same for men and women, but allow the probability masses for men and women to differ. For tractability given the large choice set and multiple cohorts, we assume the distribution of preference shocks is Type 1 Extreme Value. We assume the last period before all cohorts retire is $A = 65$, although

since labor supply is endogenous at all periods, an individual can “retire” and choose not to work before age 65. The discount rate is set at $\delta = 0.95$.

Given the finite horizon, we utilize a backwards recursion solution, detailed in the Appendix. The economy consists of overlapping generations of individuals age $a = 16, \dots, A$. We solve the model for the lifetime sequence of choices and wages (if working) for each birth cohort, type, and gender separately. The distribution of human capital investments in the economy at any calendar period t then depends on the degree choices of each non-retired birth cohort up to that point. We solve the model for all cohorts less than aged 65 in the period 1970-2002.

In addition to the main model behavioral parameters, we also allow a measurement error process in earnings. We assume (log) earnings are measured with error such that the measure of earnings in data is $\ln w_{dt}^*(k, g, a) = \ln w_{d,t}(k, g, a) + \omega_d$, where ω_d is a mean zero measurement error term, with an unknown variance σ_d^2 which we estimate. Due to the partial observability of earnings by college major, we constrain the measurement error in earnings to be the same for all college majors: $\sigma_3 = \sigma_4 = \sigma_5$.

5.3 Estimation

The full set of parameters in the model consists of i) γ parameters that determine utility flows from each of the decisions from (1), ii) α and β skill function parameters from (1, iii) gender specific type distribution parameters $\pi(k, g)$, iv) degrees specific skill rental rate parameters r_{dt} for each year and degree, and v) earnings measurement error parameters σ_d .

While the rest of the model is relatively parsimoniously parameterized, there are many skill rental rates with 6 total degrees and over 80 years of choices. To reduce the dimensionality of the parameter space, we use a spline approximation for the rental rate series for each degree assuming a constant slope in 5 year intervals from 1948-2002. Prior to 1948 and after 2002, we assume a constant skill rental rate for each degree at the level of the last skill price in 1948 or 2002.

We use a method of moments (minimum distance) estimator. Let θ denote the full set of parameters, including the skill price parameters. The vector of parameter estimates $\hat{\theta}$ is given by

$$\hat{\theta} = \operatorname{argmin}(\hat{M} - M(\theta))'W(\hat{M} - M(\theta)),$$

where \hat{M} is the sample analog moments corresponding to the model moments $M(\theta)$. Note that we exploit the structure of the model to avoid simulation and have exact analytic expressions for $M(\theta)$, as explained in the Appendix. This has the advantage that the objective function of our estimator is smooth with respect to the parameters, up to machine precision. Since we avoid simulation, we also avoid any associated “simulation noise” which would otherwise inflate the variance of our estimator.

We use the following set of moments from each data set:

i) CPS:

a) Fraction of population in one of 4 year of schooling categories (less than 12, exactly 12, 13-15, and 16 or more) by year (1963-2002), conditional on birth cohort and gender.

b) Employment rates, average annual wages and standard deviation of annual wages in one of 4 year of schooling categories (less than 12, exactly 12, 13-15, and 16 or more) by year (1963-2002), conditional on birth cohort and gender.

ii) Census:

a) Fraction of population in one of 4 year of schooling categories (less than 12, exactly 12, 13-15, and 16 or more) by year (1949 and 1959), conditional on birth cohort and gender.

b) Employment Rates, average annual wages and standard deviation of annual wages in one of 4 year of schooling categories (less than 12, exactly 12, 13-15, and 16 or more) by year (1949 and 1959), conditional on birth cohort and gender.

iii) NSCG:

a) Fraction of college educated population holding each degree for all cohorts who are non-retired and aged 35 older in years 1993 and 2002, conditional on cohort and gender.

b) Employment rates, average annual wages and standard deviation of annual wages for post-secondary degree holders in 1989 and 2002, conditional on college major, age, and gender.

6 Results

6.1 Parameter Estimates

6.1.1 Wages and Skills

Panel A of Table 3 displays the parameter estimates for the experience component of the skill production technology (1). As is typical, we estimate a concave experience profile, with a positive linear term and a negative experience squared term. Our estimate of the linear component at 0.01 is considerably smaller than the estimates using OLS regressions of log wages on potential experience (age-years of school-6), typically around 0.03 to 0.04. Unlike these estimates, our estimate of experience take into account selection into different experience profiles, given that we jointly estimate the return to experience along with our other model parameters using our model of endogenous labor supply. Our lower estimate of experience suggests that the OLS estimates are upwardly biased due to positive selection into the full time/full year labor force of higher productivity types.

The estimates of our 5 point distribution for the $\alpha_d(k)$ type specific skill function intercept terms are presented in the Appendix, Tables C-1, C-2, and C-3. To interpret the estimates, one has to take into account the particular skill price normalization. For any particular degree skill, the price r_{dt} (for degree d in calendar period t) and skill level $s_{dt}(k, g, a)$ (for type k , gender g , and age a) are not separately identifiable since skill and prices are not directly observed, and we infer skill prices from wage and choice data. We normalize prices relative to the type 1's skill intercept: $\alpha_d(1) = 0$. This implies that with no experience at some age, the level of skill for type 1 agents is $s_{dt}(1, g, a) = 1$ for all d, t, g . The wage for type 1 agents with no experience is then $w_{d,t}(1, g, a) = r_{d,t}$ for all d, t, g , where the $r_{d,t}$ prices are freely varying parameters we directly estimate, as discussed above.¹⁴

In order to interpret the heterogeneity in skill levels we estimate, Table 4 provides the average level of the heterogeneous skill intercepts for each of the various degree types. The average level is calculated as $\sum_k \pi(k, g) \alpha_d(k)$, where $\pi(k, g)$ is the probability mass for type k and gender g and $\alpha_d(k)$ is the type k level of skill in degree d .

¹⁴Ignoring skill differences from experience (assume $x_t(a) = 0$), the normalization implies the following for log wages: $\ln w_{1,t}(1, g, a) = \ln r_{1,t}$ for type 1 high school drop-outs, $\ln w_{1,t}(2, g, a) = \ln r_{1,t} + \alpha_1(2)$ for type 2 high school drop-outs, $\ln w_{2,t}(1, g, a) = \ln r_{2,t}$ for type 1 high school graduates, $\ln w_{2,t}(2, g, a) = \ln r_{2,t} + \alpha_2(2)$ for type 2 high school graduates, and so on.

One of the key issues for this paper is the male and female differences in skills. The difference in male and female “skills” represents both gender based discrimination in the labor market and male-female differences in levels of human capital. Given that we infer skills from wages, we cannot separately identify discrimination from “true” differences in productivity, which could arise from male-female differences in other, unmodeled human capital levels, such as differences in physical abilities (e.g. physical strength), high school curriculum, informal job training, or quality of work experience.

We estimate that the largest male advantage in “skills” is in lower education levels. The average man is estimated to have 0.026 log points and 0.065 log points higher skills than the average woman, with no high school degree and with only a high school degree, respectively. In contrast, the male-female difference in average skills is less than 0.01 log points in all of the post-secondary degrees. Our estimates of a male advantage in the labor market with a secondary or lower degree may reflect discrimination in this part of the labor market or differences in productivity, for example a natural strength advantage male workers have in low education occupations. Our estimate that in employment with a post-secondary degree the male advantage is considerably diminished could be because of lower levels of discrimination for these labor markets or because there is truly little innate skill differences between men and women in jobs requiring higher education. An important caveat in interpreting the results is that because of self-selection into human capital investments based on comparative advantage, these differences in average skills do not necessarily translate into differences in average realized wages.

Another important issue in interpreting the estimated skill heterogeneity is that small mean differences do not necessarily indicate that there is no difference in the distribution of male and female skills. For example, the Appendix tables show that the type with the highest level of science/math/engineering skill (Type 4) is estimated to comprise just 3 percent of women but 18 percent of men. On the other hand, men have a higher proportion of the lowest science skill type (Type 2), which comprises 29 percent of men and 17 percent of women. We estimate a distribution of science/math/engineering skills that is more dispersed for men than women, although the means are quite similar, with only a small male advantage. These estimates are remarkably similar to the general pattern found in past research when examining gender differences in cognitive test scores.¹⁵

¹⁵Previous research is somewhat mixed but generally supports the conclusion of small innate differences.

Turning next to the skill rental rate component of earnings, Figure 6 displays the estimated skill rental rates for each of the specific college degrees. The general trend is a divergence in the skill rental rates as the science/mathematics/engineering and business/economics fields experienced a much larger increase in rental rates during the 1980s and 1990s than the humanities/social sciences/teaching field. Starting in the 1970s there is a decline in the skill rental rates for science and business fields. However both of these fields experienced a relative increase in the 1980s and 1990s. The rate of increase is especially high for the business field. For science, the rate of increase accelerates in the 1990s. By contrast, for the humanities/social sciences/teaching field, there is a general stagnation in the skill price throughout this period, with only a recovery toward the end of the period in the late 1990s.

6.1.2 School Cost/Degree Taste Parameters

In our model of schooling, the non-stochastic utility flow from attending school consists of two parts: i) a degree and type k specific heterogeneous taste for each degree $\gamma_d(k)$ and ii) a homogeneous non-stationary tuition cost given by $\gamma\tau_{d,t}$. The estimated type specific degree tastes are given in the Appendix, Tables C-1, C-2, and C-3. We allow separate tastes for each of the degrees, and we allow the type distribution to vary by gender incorporating gender specific tastes for different degrees.

Table 4 provides the estimated average level of tastes for each degree, computed separately for men and women. On average, women have a higher level of taste for each of the degree levels. However, the male-female ratio in average tastes varies across degrees, with a high of 0.7 for high school to 0.59-0.63 for two year and college degrees. Comparing these male-female differences in average tastes for degrees with the differences in average skills across degrees, our estimates indicate a much larger difference in tastes than skills. This finding is consistent with a large body of research which finds important gender differences in interests and tastes for science, engineering, and

In an analysis of several nationally representative data sets, Hedges and Nowell (1995) find that there is little difference in the mean abilities of boys and girls, as measured by IQ and subject exams taken by elementary and high school students. However, Hedges and Nowell do find that there is greater variance in male ability and larger numbers of very high-scoring boys. These gender differences are relatively small, however, and it is difficult to associate these differences with innate biological or genetic factors rather than differences in early socialization and schooling. Similarly, in a more recent review of the existing psychology and development literature, Hyde (2005) finds that boys and girls are similar in most psychological dimensions with the exception of some motor skills, sexual behavior, and levels of aggression. For evidence on other dimensions of skill, see Jacob (2002) who emphasizes the role of non-cognitive skills.

mathematics fields generated by the different family, school, and cultural environments young men and women face (e.g. Figlio 2005; Xie and Shauman 2003; Leslie, McClure, and Oaxaca 1998; Betz 1997).

Turning next to the non-stationary schooling cost component of the flow utility of schooling, the second part of Table 3 shows that the coefficient on the time varying tuition level variable is $\gamma_6 = -0.0005$. With the marginal utility of income set at $\gamma_7 = 0.00003$, this estimate implies that \$1 increase in tuition is equivalent to a $\gamma_6/\gamma_7 = -16.7$ reduction in income. Our estimate of the marginal utility of tuition essentially reflects the complex mapping between tuition costs and school utility flows, incorporating possible credit constraints that reduce the marginal utility of schooling more than what is reflected in the tuition menu price. Such channels are not explicitly modeled here, but their implications are captured through the parameter governing marginal utility of tuition costs. From the relationship of schooling choices to the time series of tuition levels, we identify a rather large wedge between schooling choices and tuition costs. The decomposition analysis reported below, in which we manipulate the level of tuition in the estimated model, allows us to directly analyze the relative importance of non-stationary tuition costs to the human capital investment process.

6.1.3 Home/Leisure Value

The final component of our model is the value of the non-work and non-school home alternative. Like the flow utility for schooling, there are two components to the value of home: i) a heterogeneous value of leisure which varies by type k $\gamma_8(k)$ and gender intercept $\gamma_9(g)$ and ii) a non-stationary component consisting of a term that depends on cohort, age, and gender specific fertility $\gamma_{10}(g)c_t(g, a)$ and a term that depends on cohort directly $\gamma_{11}(g)(t - a)$.

Table 3 displays the parameter estimates for the utility of leisure. The stationary home value/leisure intercepts vary substantially across types: Type 1 has the highest value at 25.14, while Type 2 has a lowest value at -1.06. Table 4 provides the average level of the home value intercepts for men and women. Women have an average value of 13.13 compared to 7.59 for men, yielding a male-female ratio of 0.58. These differences in home value indicate that even without children present, there is still a substantial difference in the value of home to women compared to men. These gender differences in the value of staying at home reflect many possible elements that may give rise to women's taste for home being higher than men. Some examples are women's

comparative advantage in home production or cultural differences, both of which are not explicitly modeled here but are subsumed in the gender specific parameters for value of staying at home.

From Table 3 we see that the number of children under 5, our measure of fertility, is estimated to increase the value of home for women, but decrease the value for men. One interpretation of this estimate is that with an increase in the number of children, men and women specialize their time, with women increasingly staying at home in home production or child rearing and men increasing their time in the labor market to generate labor income to finance child expenditures. Figure 7 displays the average value of home for men and women across birth cohorts. These are the dollar equivalent values of utility of staying at home for type 1 men and women at age 30 in each year. It can be seen here that on average women’s value of staying at home is on average \$746,000, whereas for men, it is on average \$490,000. Moreover, for women, these values change over time, especially with fertility. Mirroring the trend in fertility, the largest home value is for the mothers of the “baby boom” cohorts, born in the 1920s. Following this peak, there is a steady secular decline in the value of home for women. The pattern for men is more flat than for women. The sharp fall in fertility rates have led to a sharper changes in the utility of leisure for women than it did for men.

6.1.4 Measurement Error

We estimate the standard deviation of the measurement error process ($\sigma_m(d)$) for log annual earnings at around 0.36-0.37. With the standard deviation of log earnings varying between 0.5 and 0.6, depending on education level, our estimates of the measurement process imply that about 1/4 to 1/3 of the observed variance in log earnings is attributable to measurement error noise.¹⁶

¹⁶With our i.i.d. measurement error model, $V(\ln w^*) = V(\ln w) + V(\omega)$, where w^* is observed earnings, w is the true level of earnings determined by the model, and ω is the i.i.d. measurement error component. With the standard deviation of observed log earnings at 0.5 and $V(\omega) = 0.36^2$, the proportion of the variance in observed log earnings due to measurement error is $0.36^2/(0.5^2 + 0.36^2) = 0.34$. With the standard deviation of observed log earnings at 0.6, this proportion is $0.36^2/(0.6^2 + 0.36^2) = 0.265$. In the combined CPS and Census data for cohorts born 1920 to 1969, aged 25 to 59, who worked full time/full year, the standard deviation of log annual earnings varied from 0.51 (high school drop-outs), 0.50 (high school graduates), 0.52 (some college), 0.58 (college graduates).

6.2 Model Fit

Next we explore how well the estimated model fits the main empirical patterns of interest.

6.2.1 Educational Attainment (Women)

Figure 8 provides the sample fit of the estimated model to the general birth cohort trends in female educational attainment. For all of the figures, we graph the educational attainment of the cohort at age 35. In terms of the extensive margin of human capital investments, the estimated model captures the rise in proportion of women graduating from college in the combined Census and CPS data.

The remaining graphs in Figure 8 plot the proportion of college graduates from each cohort that completed a college degree in each of the 3 major fields. The data trends for these graphs are taken from the NSCG data. The estimated model fits the general patterns in composition of college degrees, with the model matching the increase in the proportion of women graduating with a science/math/engineering degree and the larger increase in the proportion graduating with a business/economics degree. The model also captures the sharp fall in the proportion of women completing degrees in the humanities/social sciences/teaching field, starting with the cohorts born in the mid 1940s, and the increase in the proportion with humanities degrees for the most recent cohorts, although the model under-predicts the proportion in humanities during the middle period and over-predicts during the more recent period.

6.2.2 Earnings (Women)

Figure 9 displays the annual earnings for women at age 35 by year. Generally the model captures the level difference in earnings across degree groups. For all lower education degrees (high school drop-out and high school degree), the model estimates capture the stagnation in real earnings over this period, with the exception of an over-predicted bump in earnings for high school graduates during the 1980s. For college graduates, the model captures both the higher level of earnings and the slightly upward trending time series for earnings in the post-1980 period.

6.2.3 Educational Attainment (Men)

Figure 10 displays the sample fit for male educational attainment at age 35. In terms of the extensive margin, the model fits the non-linear pattern of college attendance, first rising through the 1940s cohorts, then stagnating and falling slightly, before finally rising again for the most recent cohorts. The estimated model also captures the general pattern in the composition of college degrees. The model captures the relatively small change in the proportion of science/math/engineering, the larger increase in proportion majoring in business/economics, and the fall in humanities/social sciences/teaching degrees, before a slight recovery for the most recent cohorts.

6.2.4 Earnings (Men)

Figure 11 displays the sample fit for average earnings for men at age 35 by year. The estimated model is generally capable of replicating the difference in levels of average earnings across degree groups. The model slightly under-predicts earnings in the early period for male high school drop-outs and men with two year college degrees. However, the sample fit for the later period is considerably better. The estimated model fits the general pattern of increasing average earnings for college education men starting in the 1980s, although here too the model under-predicts average earnings for some years.

7 Counterfactual Experiments: Determinants of Long-Run Changes in Human Capital Investments

The estimated model provides evidence on the factors that gave rise to gender differences in college attainment and college major choices. The three channels include changes in the relative prices for skill, changes in the cost of post-secondary school, and changes in the value of home. In order to assess the importance of each of these channels, we use a series of counterfactual experiments reported in Table 5.

The counterfactual experiments in Table 5 progressively add in elements of the model to explain the change in educational attainment from 1940 to 1960 (age 35 men and women in both years). Column (1) presents the full predicted model change, e.g. +12 is the increase in the percentage of women graduating from college, from 14 percent in 1940 to 26 percent in 1960. Column (2) presents the baseline level of no change in which we keep all the non-stationary elements of the model fixed at the 1940

values. At this baseline there is no change in any aspect of behavior over this period. In Columns (3)-(5) we progressively add in the non-stationary elements of the model and report the 1940 to 1960 *marginal* change in education attainment from each of these elements. The marginal changes in columns (3)-(5) sum to the total predicted change in Column (1), e.g. the change in the percentage of women in college is $+12 = +6 - 10 + 16$.

7.1 Extensive Margin: College Graduation

Focusing first on the extensive margin of college graduation, in Column (3) of Table 5 we see that allowing the skill prices to change as estimated from 1940 to 1960 increases the proportion of women completing college degrees by 6 percentage points, from 14 to 20 percent. The net effect of the various changes in skill rental rates we estimate is to increase the labor market return to a college degree more than the opportunity cost of college attendance. Next, adding in the increase in tuition rates during this period raises the cost of college and reduces the percentage of women choosing to obtain a college degree by 10 percentage points. This can be seen in Column (4) of Table 5. The total effect of skill prices changes and tuition is a net $(+6 - 10) = -4$ percentage change in college graduation. Finally, we add-in all model elements: skill prices, tuition, and home value. The marginal effect from the lower home value for women increases the fraction of women with a college degree by 16 percentage points. From this decomposition, we learn that each of the 3 non-stationary elements of the model played a role in the increase in the percentage of women graduating from college, with the change in home value playing the largest quantitative role. However, changes in the value of home were not the only factor as changes in skill prices favoring college attainment amplified the reduction in home value, and increases in tuition costs dampened this factor.

For men, the across cohort trend in college graduation rates was much less stark than for women, with a 3 percentage point drop between the cohort born in 1940 and 1960. However, as the decomposition in Table 5 shows this *net* effect hides the composite countervailing factors. Just as for women, the change in relative skill prices increased the return to a college degree, and this factor alone would have increased the proportion of males with a college degree by 16 percentage points. Compared to the equivalent figure for women, the men's counterfactual response to changes in relative skill prices is much larger, reflecting the importance of gender differences in home values

(at the 1940 level) and stationary tastes for degrees. In addition, the change in tuition had a larger negative effect on men than for women, a 21 percentage point fall for men versus a 10 percentage point fall for women. We estimate that men are more “tuition sensitive” because men’s skills are less specialized across degrees and hence there are more men on the margin between entering college and working without a college degree. This can be seen in Tables C-1, C-2, and C-3. Type 2 and Type 5 individuals have skills that are less specialized across degrees in that they face smaller skill differentials across different education levels and degrees. The proportion of men who are these types is 38 percent, whereas it is 24 percent for women. For these reasons, a larger proportion of men face a smaller differential in their potential wages across different education levels and different degrees, compared to women. The small change in home value for men was a relatively minor factor in their college attendance trend.

7.2 Intensive Margin: College Major Composition

Turning next to the intensive margin of college major composition, we see that the effect of changes in skill prices alone shifts the degree composition choices for women and men away from humanities, social sciences, and teaching fields and toward science, mathematics, and engineering and business and economics fields. This reflects the divergence in the estimated skill rental rates as the science/mathematics/engineering and business/economics fields experienced a much larger increase in rental rates during the 1980s and 1990s than the humanities/social sciences/teaching field. The responses to skill price changes at the intensive margin can be seen in Column (3) of Table 5. The proportion of women college graduates who study science/mathematics/engineering increases by 6 percentage points (from 19 percent to 25 percent), and the proportion who studied business/economics increases by 31 percentage points (from 6 percent to 37 percent). Men experience a less pronounced, but similar pattern in response to skill price changes. As mentioned in the preceding section, the estimated model shows that at the intensive margin, men are less responsive to changes in relative degree specific skill prices, because a larger proportion of men face a lower differential in their degree specific skills and tastes across college majors relative to women.

We next add the increase in tuition rates over this time period, so that Column (4) of Table 5 shows the intensive margin responses to skill rental rate changes and tuition rate increases. Interestingly, the increase in tuition rates during this period has the opposite effect compared to skill rental rates. In response to tuition rate

increases, Type 2 and Type 5 agents, who have lower labor market return to college due to the lower skill differential across different degrees, stop going to college. These agents at the margin are predominantly agents who study science/math/engineering or business/economics when they go to college. Hence, with the outflow of this group, the proportion who studies these degrees falls.

Finally, the change in home value had a large effect on field of study composition for women. The decrease in home value for women reduced the percentage of women majoring in humanities by 20 percentage points, and increased the percentage in science and business fields by 5 and 14 percentage points, respectively. For men the change in home value had a relatively minor effect, although the effect on the college major composition was larger than the extensive margin change in proportion in college.

8 Conclusion

This paper documents the changes in educational attainment and college major composition for men and women, and develops and estimates a dynamic overlapping generations model of schooling, college major, and labor supply decisions to explain these trends in attainment. Parameter estimates provide evidence for the structural differences between men and women that give rise to gender differentials in college attainment levels and college major composition. The estimated model allows us to analyze the determinants of the long-term changes in human capital investments over the past 40 years in the United States.

We estimate a distribution of science/math/engineering skills that is more dispersed for men than women. These estimates are remarkably similar to the general pattern found in past research when examining gender differences in cognitive test scores. However, we estimate that the factor that is more important in giving rise to gender differences in college major choices is tastes rather than skill differentials between men and women. Our estimates indicate a much larger gender difference in tastes than skills. Second, we find that women have a higher utility from staying at home, which can be interpreted as a comparative advantage in home production or cultural differences. This difference in value of staying at home manifests itself as higher opportunity cost of going to college for women. As value of staying at home falls across cohorts, women complete more college degrees and switch their major composition to science and business majors with higher skill prices. Third, we find that men's college attainment decisions are much

more sensitive to rise in tuition costs compared to women. This is because men's skills are less specialized across degrees hence there are more men on the margin between entering college and working without a college degree. Rise in tuition costs decrease the proportion of women with a college degree and change their college major composition considerably, whereas it has a smaller effect on the proportion of men with a college degree.

While our analysis connects important aggregate features of changing human capital investment process, there are several important areas for future research. One key area is understanding the early life formation of gender differences in skills, and in particular, tastes for different college fields. While we can identify a large role for taste differences, our data does not allow us to investigate the source of these differences in earlier life, or track any long-term changes. In addition, several studies have argued that colleges and universities affect the tastes for different fields directly through the gender composition of the students and faculty and other aspects of the post-secondary schooling environment (Solnick 1995 and Bank, Slavings; Biddle 1990). Another important area for future research is understanding the flows of men and women across occupations, in particular the higher exit of women than men from science and technical occupations (Hunt 2010). Connecting these patterns to the aggregate human capital investments and life-cycle labor supply behavior we study could provide important insights into the connection between early life investments and later life occupational choices.

References

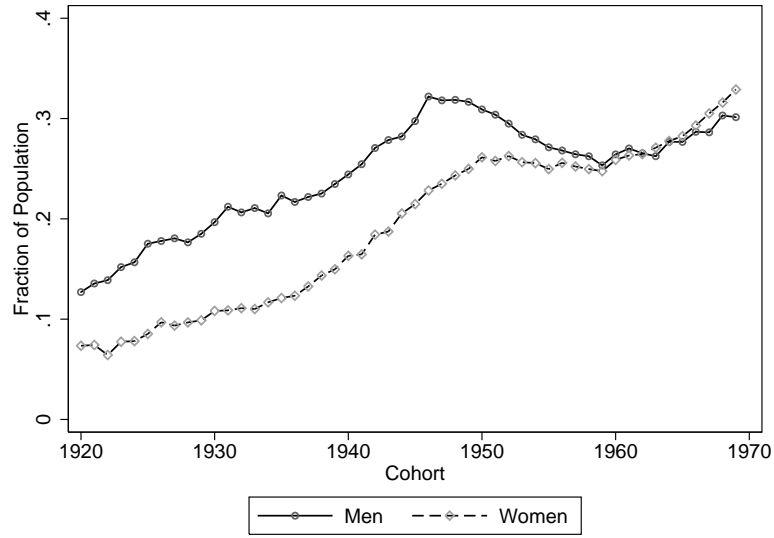
- ARCIDIACONO, P. (2004): "Ability Sorting and the Returns to College Major," *Journal of Econometrics*, 121(1-2), 343–75.
- ARCIDIACONO, P., V. J. HOTZ, AND S. KANG (2011): "Modeling College Major Choice using Elicited Measures of Expectations and Counterfactuals," *Journal of Econometrics*.
- BAILEY, M. J. (2006): "More Power to the Pill: The Impact of Contraceptive Freedom on Women's Life Cycle Labor Force Participation," *Quarterly Journal of Economics*, 121(1), 289–320.
- BANK, B. J., R. L. SLAVINGS, AND B. J. BIDDLE (1990): "Effects of Peer, Faculty, and Parental Influences on Students' Persistence," *Sociology of Education*, 63(3), 208–225.
- BERGER, M. C. (1988): "Predicted Future Earnings and Choice of College Major," *Industrial and Labor Relations Review*, 41(3), 418–29.
- BETZ, N. (1997): "What Stops Women and Minorities from Choosing and Completing Majors in Science and Engineering," in *Minorities and Girls in School: Effects on Achievement and Performance*, ed. by D. Johnson. Russell Sage Publications.
- BLACK, D., A. HAVILAND, S. SANDERS, AND L. TAYLOR (2004): "Gender Wage Disparities among the Highly Educated," *working paper*.
- BLACK, D., S. SANDERS, AND L. TAYLOR (2003): "Measurement of Higher Education in the Census and CPS," *Journal of American Statistical Association*, 98, 545–54.
- BLAKEMORE, A. E., AND S. A. LOW (1984): "Sex Differences in Occupational Selection: The Case of College Majors," *Review of Economics and Statistics*, 66(1), 157–163.
- BROWN, C., AND M. CORCORAN (1997): "Sex Based Differences in School Content and the Male-Female Wage Gap," *Journal of Labor Economics*, 15(3), 431–65.
- CARD, D., AND T. LEMIEUX (2001): "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis," *Quarterly Journal of Economics*, 116(2), 705–46.
- CHARLES, K. K., AND M.-C. LUOH (2003): "Gender Differences in Completed Schooling," *Review of Economics and Statistics*, 85(3), 559–77.
- EIDE, E. (1994): "College major choice and changes in the gender wage gap," *Contemporary Economic Policy*, 12, 55–64.
- EIDE, E., AND G. WAEHRER (1998): "The Role of the Option Value of College Attendance in College Major Choice," *Economics of Education Review*, 17(1), 73–82.
- FERNANDEZ, R., A. FOGLI, AND C. OLIVETTI (2004): "Mothers and Sons: Preference Formation and Female Labor Force Dynamics," *Quarterly Journal of Economics*, 119(4), 1249–1299.

- FERNANDEZ, R., N. GUNER, AND J. KNOWLES (2005): "Love and Money: A Theoretical and Empirical Analysis of Household Sorting and Inequality," *Quarterly Journal of Economics*, 120(1), 273–344.
- FIGLIO, D. N. (2005): "Why Barbie Says 'Math is Hard'," *working paper*.
- FREEMAN, R. (1975a): "Legal Cobwebs: A Recursive Model of the Market for Lawyers," *Review of Economics and Statistics*.
- (1975b): "Supply and Salary Adjustments to the Changing Science Manpower Market: Physics, 1948–1975," *American Economic Review*, 65(1), 27–39.
- (1976): "A Cobweb Model of the Supply and Starting Salary of New Engineers," *Industrial Labor Relations Review*.
- GAO (2004): *Gender Issues: Women's Participation in the Sciences Has Increased, but Agencies Need to Do More to Ensure Compliance with Title IX*. United States Government Accountability Office.
- GOLDIN, C. (2005): "From the Valley to the Summit: A Brief History of the Quiet Revolution that Transformed Women's Work," *Regional Review*, 14(3), 5–12.
- GOLDIN, C., AND L. F. KATZ (2002): "The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions," *Journal of Political Economy*, 110(4), 730–70.
- GOLDIN, C., L. F. KATZ, AND I. KUZIEMKO (2006a): "The Homecoming of American College Women: The Reversal of the College Gender Gap," *Journal of Economic Perspectives*, 20(4), 133–56.
- (2006b): "More Power to the Pill: The Impact of Contraceptive Freedom on Women's Lifecycle Labor Supply," *Quarterly Journal of Economics*, 121(1), 289–320.
- GREENWOOD, J., A. SESHADRI, AND M. YORUKOGLU (2005): "Engines of Liberation," *Review of Economic Studies*, 72, 109–133.
- GROGGER, J., AND E. EIDE (1994): "Changes in College Skills and the Rise in the College Wage Premium," *Journal of Human Resources*, 30(2), 280–310.
- HEATHCOTE, J., K. STORESLETTEN, AND G. VIOLANTE (2010): "The Macroeconomic Implications of Rising Wage Inequality in the United States," *Journal of Political Economy*, 118(4), 681–722.
- HEDGES, L. V., AND A. NOWELL (1995): "Sex Differences in Mental Test Scores, Variability, and Numbers of High-Scoring Individuals," *Science*, 269(5220), 41–45.
- HUNT, J. (2010): "Why Do Women Leave Science and Engineering?," *NBER Working Paper 15853*.
- HYDE, J. S. (2005): "The Gender Similarities Hypothesis," *American Psychologist*, 60(6), 581–92.

- JACOB, B. A. (2002): “Where the boys aren’t: non-cognitive skills, returns to school and the gender gap in higher education,” *Economics of Education Review*, 21, 589–98.
- KATZ, L. F., AND K. M. MURPHY (1992): “Changes in Relative Wages, 1963-87: Supply and Demand Factors,” *Quarterly Journal of Economics*, 107(1), 35–78.
- KRUSELL, P., L. E. OHANIAN, J.-V. RIOS-RULL, AND G. L. VIOLANTE (2000): “Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis,” *Econometrica*, 68(5), 1029–1053.
- KRUSELL, P., AND A. SMITH (1998): “Income and Wealth Heterogeneity in the Macroeconomy,” *Journal of Political Economy*, 106, 867–896.
- LEE, D. (2005): “An Estimable Dynamic General Equilibrium Model of Work, Schooling and Occupational Choice,” *International Economic Review*, 46(1), 1–34.
- LEE, D., AND K. WOLPIN (2006): “Intersectoral Labor Mobility and the Growth of Service Sector,” *Econometrica*, 74(1), 146.
- (2009): “Accounting for Wage and Employment Changes in the U.S. from 1968-2000: A Dynamic Model of Labor Market Equilibrium,” *Journal of Econometrics*, 156(1), 68–85.
- LESLIE, L. L., G. T. MCCLURE, AND R. L. OAXACA (1998): “Women and Minorities in Science and Engineering: A Life Sequence Analysis,” *Journal of Higher Education*, 69(3), 239–76.
- MACHIN, S., AND P. PUHANI (2003): “Subject of degree and the gender wage differential: evidence from the UK and Germany,” *Economics Letters*, 79(3), 393–400.
- MONTMARQUETTE, C., K. CANNINGS, AND S. MAHSEREDJIAN (2005): “How do young people choose college majors?,” *Economics of Education Review*, 21, 543–556.
- NSF (2003): *New Formulas for America’s Workforce: Girls in Science and Engineering*. National Science Foundation.
- PAGLIN, M., AND A. M. RUFOLO (1990): “Heterogeneous Human Capital, Occupational Choice, and Male-Female Earnings Differences,” *Journal of Labor Economics*, 8(1), 123–44.
- RUMBERGER, R. W., AND S. THOMAS (1993): “The Economic Returns to College Major, Quality, and Performance: A Multilevel Analysis of Recent Graduates,” *Economics of Education Review*, 12, 1–19.
- SHORE, S., AND R. E. SAKS (2005): “Risk and Career Choice,” *Advances in Economic Analysis & Policy*, 5(1).
- SIOW, A. (1984): “Occupational Choice under Uncertainty,” *Quarterly Journal of Economics*, 52(3), 631–45.

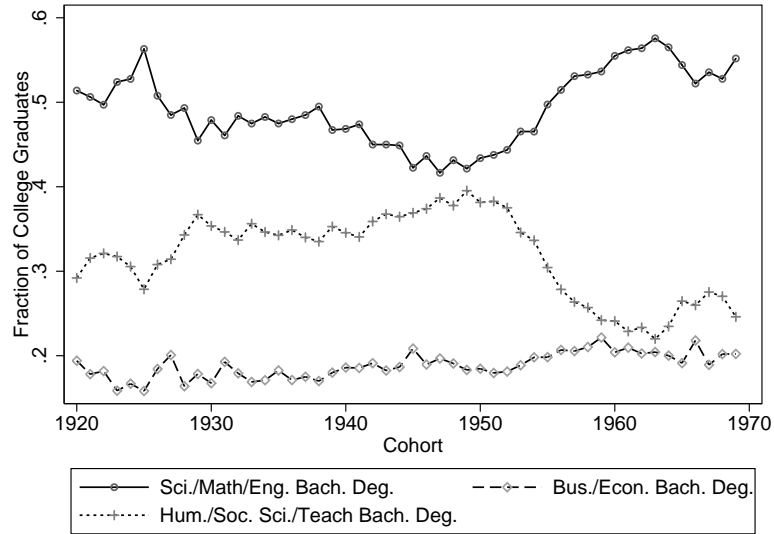
- WEINBERGER, C. J. (1998): “Race and Gender Wage Gaps in the Market for Recent College Graduates,” *Industrial Relations*, 37(1), 67–84.
- WISWALL, M., AND B. ZAFAR (2011): “Determinants of College Major Choices: Identification using an Information Experiment,” *working paper*.
- XIE, Y., AND K. SHAUMAN (2003): *Women in Science: Career Processes and Outcomes*. Harvard University Press.
- ZAFAR, B. (2011): “How do College Students Form Expectations?,” *Journal of Labor Economics*, 29(2), 301–48.
- ZARKIN, G. A. (1985): “Occupational Choice: An Application to the Market for Public School Teachers,” *Quarterly Journal of Economics*, 100(2), 409–446.

Figure 1: Fraction of Population with a College Degree



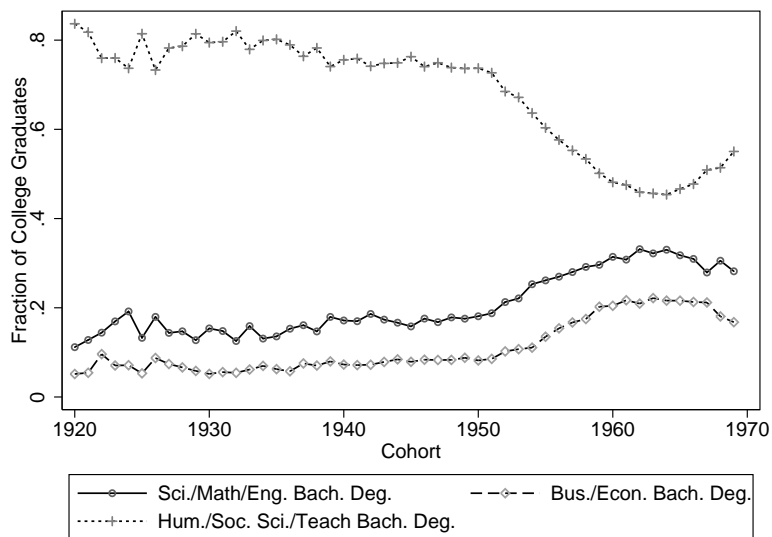
Source: 1940,50,60 US Census and 1964-2009 March CPS.

Figure 2: Field of Study Composition of Bachelor Degrees by Year of Birth (Men)



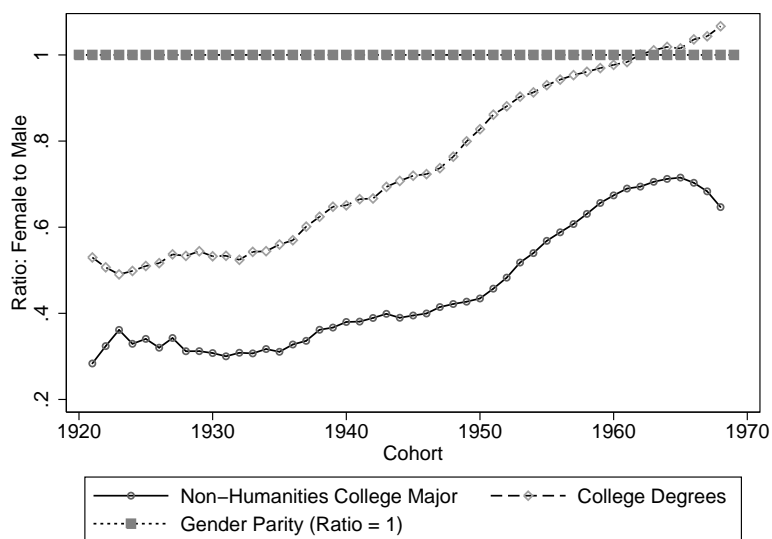
Source: 1993 and 2003 NSCG data.

Figure 3: Field of Study Composition of Bachelor Degrees by Year of Birth (Women)



Source: 1993 and 2003 NSCG data.

Figure 4: Ratio of Female-Male College Completion and Non-Humanities Fields



Source: 1993 and 2003 NSCG data.

Table 1: Descriptive Statistics for NSCG Sample of College Graduates

	All	Males	Females
Sci./Math/Eng.	0.26	0.34	0.17
Frac. Bus./Econ.	0.22	0.28	0.16
Frac. Hum./Soc. Sci./Teach	0.52	0.39	0.68
Sci./Math/Eng.			
Frac. Full Time/Full Year	0.73	0.83	0.51
Mean Wage	81780	86789.	63873
Median Wage	70328	74817	53868
Std. Wage	46613	47716	37290
Bus./Econ.			
Frac. Full Time/Full Year	0.74	0.84	0.54
Mean Wage	75860	81963	57675
Median Wage	61468	67335	49174
Std. Wage	47678	49910	34392
Hum./Soc. Sci./Teach			
Frac. Full Time/Full Year	0.53	0.73	0.41
Mean Wage	62934	72198	52590
Median Wage	51224	58907	44890
Std. Wage	41400	46967	31016
Total Observations: 181,427			

Notes:

Source: 1993 and 2003 NSCG data.

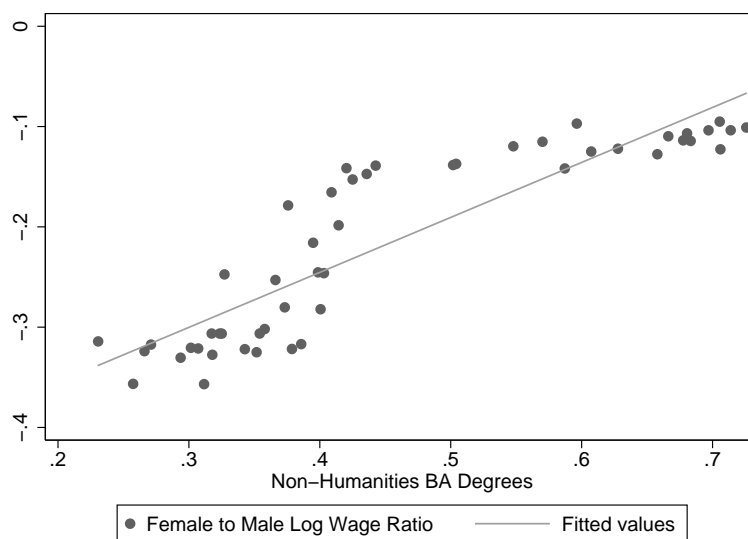
Table 2: Age Adjusted Labor Supply and Earnings Differences by College Major

Panel A: Labor Supply			
Dep. Var.: Full Time Indicator	(1) All	(2) Males	(3) Females
Sci./Math/Eng.	0.091 (0.0034)	0.092 (0.0040)	0.087 (0.0060)
Bus./Econ.	0.11 (0.0039)	0.10 (0.0045)	.1183178 (0.0072)
Male	0.32 (0.0030)	–	–
Age Fixed Effects	Yes	Yes	Yes
R-Squared	0.15	0.037	0.020
Observations	181427	104218	77209
Panel B: Earnings			
Dep. Var.: Log Annual Earnings	(1) All	(2) Males	(3) Females
Sci./Math/Eng.	0.22 (0.0053)	0.24 (0.0066)	0.19 (0.0089)
Bus./Econ.	0.15 (0.0062)	0.16 (0.0078)	0.10 (0.010)
Male	0.27 (0.0049)	–	–
Age Fixed Effects	Yes	Yes	Yes
R-Squared	0.16	0.11	0.089
Observations	120172	82772	37400

Notes: Panel A reports regression results from a regression of $\{0, 1\}$ indicator for full time status on indicators for college major. Panel B reports regression results from a regression of log annual earnings on indicators for college major. In all regressions, in all panels, the omitted category is humanities/social sciences/teaching. All regressions include a full set of indicators for respondent age and dummy variable for the survey year. Panel A is estimated on pooled sample of college graduates from the 1993 and 2003 NSCG. Panel B is estimated on the pooled sample of full time/full year college graduates from the 1993 and 2003 NSCG.

Source: 1993 and 2003 NSCG data.

Figure 5: Cohort Effect Earnings vs. Cohort Fraction in Non-Humanities Degrees



Notes: This figure plots the female-to-male ratio in age adjusted log wage cohort effects for individuals with 16 or more years of schooling (from Census and CPS) vs. the female-to-male ratio in the proportion of the cohort's college degrees in non-humanities fields (from NSCG). Non-humanities fields include science and business majors. The regression line slope estimates (standard error) is 0.55 (0.041), with an R-squared of 0.79.

Source: 1940,50,60 US Census; 1964-2009 March CPS; and 1993 and 2003 NSCG data.

Table 3: **Parameter Estimates**

		Parameter	S.E.
Panel A: Skill Production Technology			
β_1	Experience	0.01	(0.0003054)
β_2	Experience Squared	-0.000057	(0.0000077)
Panel B: Utility Function			
γ_6	Marginal utility cost of tuition	-0.0005	(0.000003)
γ_7	Marginal utility of income	0.00003	-
$\gamma_8(k), k = 1$	Intercept in value of leisure - Type 1	25.14	(0.511)
$\gamma_8(k), k = 2$	Intercept in value of leisure - Type 2	-1.06	(0.152)
$\gamma_8(k), k = 3$	Intercept in value of leisure - Type 3	7.20	(0.063)
$\gamma_8(k), k = 4$	Intercept in value of leisure - Type 4	0.48	(0.212)
$\gamma_8(k), k = 5$	Intercept in value of leisure - Type 5	0.45	(0.169)
γ_9	Female intercept in value of leisure	1.23	(0.046)
$\gamma_{10}(g), g = 0$	Degree to which children increase value of leisure - Females	10.05	(0.114)
$\gamma_{10}(g), g = 1$	Degree to which children increase value of leisure - Males	-3.86	(0.174)
$\gamma_{11}(g), g = 0$	Degree to which value of leisure changes by cohort - Females	-0.37	(0.003)
$\gamma_{11}(g), g = 1$	Degree to which value of leisure changes by cohort - Males	-0.32	(0.005)
Panel C: Earnings Measurement Error Parameters			
σ_0	No High School	0.36	(0.004)
σ_1	Only High School	0.36	(0.008)
σ_2	Two Year	0.37	(0.001)
σ_3	College	0.37	(0.003)

Notes: γ_7 is the normalized marginal utility of income. We set it at this value for computational convenience.

Figure 6: Skill Rental Rates

College to High School Skill Rental Rate Ratio Relative to 1970
 $R(\text{College})/R(\text{High School})$ Relative to $R(\text{College})/R(\text{High School})$ in 1970

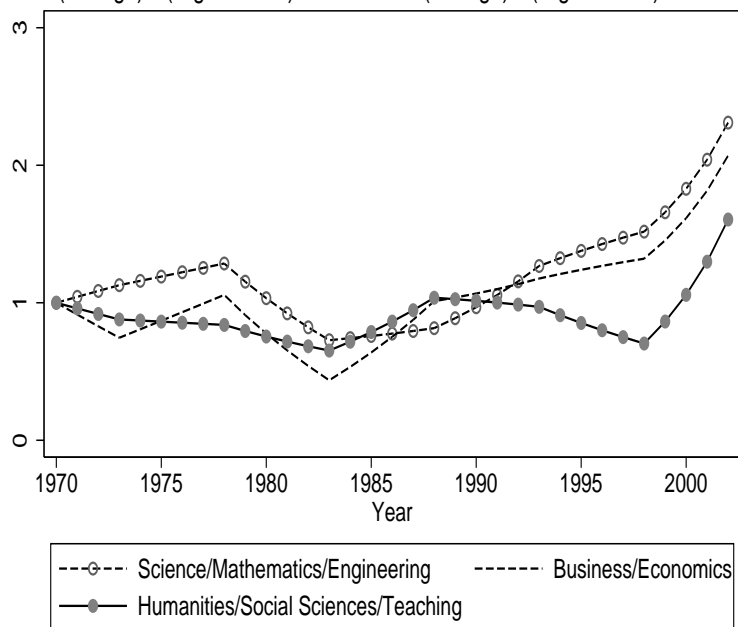


Figure 7: Average Home Value for Women

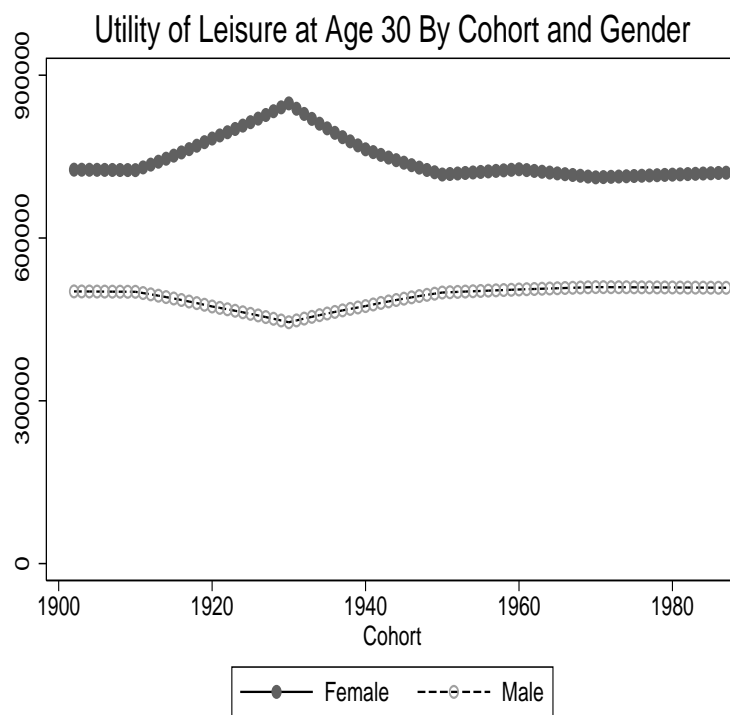


Table 4: Average Skill and Tastes by Gender

Panel A: Skill Differences			
	Female	Male	Male-Female Log Difference
Avg. Log No High School Skill (d=0)	0.044	0.070	0.026
Avg. Log High School Skill (d=1)	0.020	0.085	0.065
Avg. Log Two Year Skill (d=2)	-0.008	-0.007	0.001
Avg. Log Science Skill (d=3)	-0.032	-0.031	0.001
Avg. Log Business Skill (d=4)	-0.039	-0.044	-0.005
Avg. Log Humanities Skill (d=5)	-0.038	-0.031	0.007

Panel B: Degree Taste Differences			
	Female	Male	Male-Female Ratio
Avg. High School Taste (d=1)	7.52	5.29	0.70
Avg. Two Year Taste (d=2)	7.40	4.36	0.59
Avg. Science Taste (d=3)	7.58	4.73	0.62
Avg. Business Taste (d=4)	11.29	6.67	0.59
Avg. Humanities Taste (d=5)	9.33	5.91	0.63

Panel C: Home Value Differences			
	Female	Male	Male-Female Ratio
Avg. Utility of Leisure	13.13	7.59	0.58

Notes: Average (log) skill levels are calculated as $\sum_k \pi(k, g) \alpha_d(k)$ for each degree $d = 0, \dots, 5$ and gender, where k indexes type and g indexes gender. Average tastes are calculated as $\sum_k \pi(k, g) \gamma_d(k)$ for each degree $d = 1, \dots, 5$ and gender. Average utility of leisure (home value) is calculated as $\sum_k \pi(k, g) \gamma_8(k)$ for each gender. This table provides the actual parameter estimates. To transform the taste and leisure parameters into dollar equivalents, one must divide by the normalized value of the marginal utility of income γ_7 .

Figure 8: Model Fit

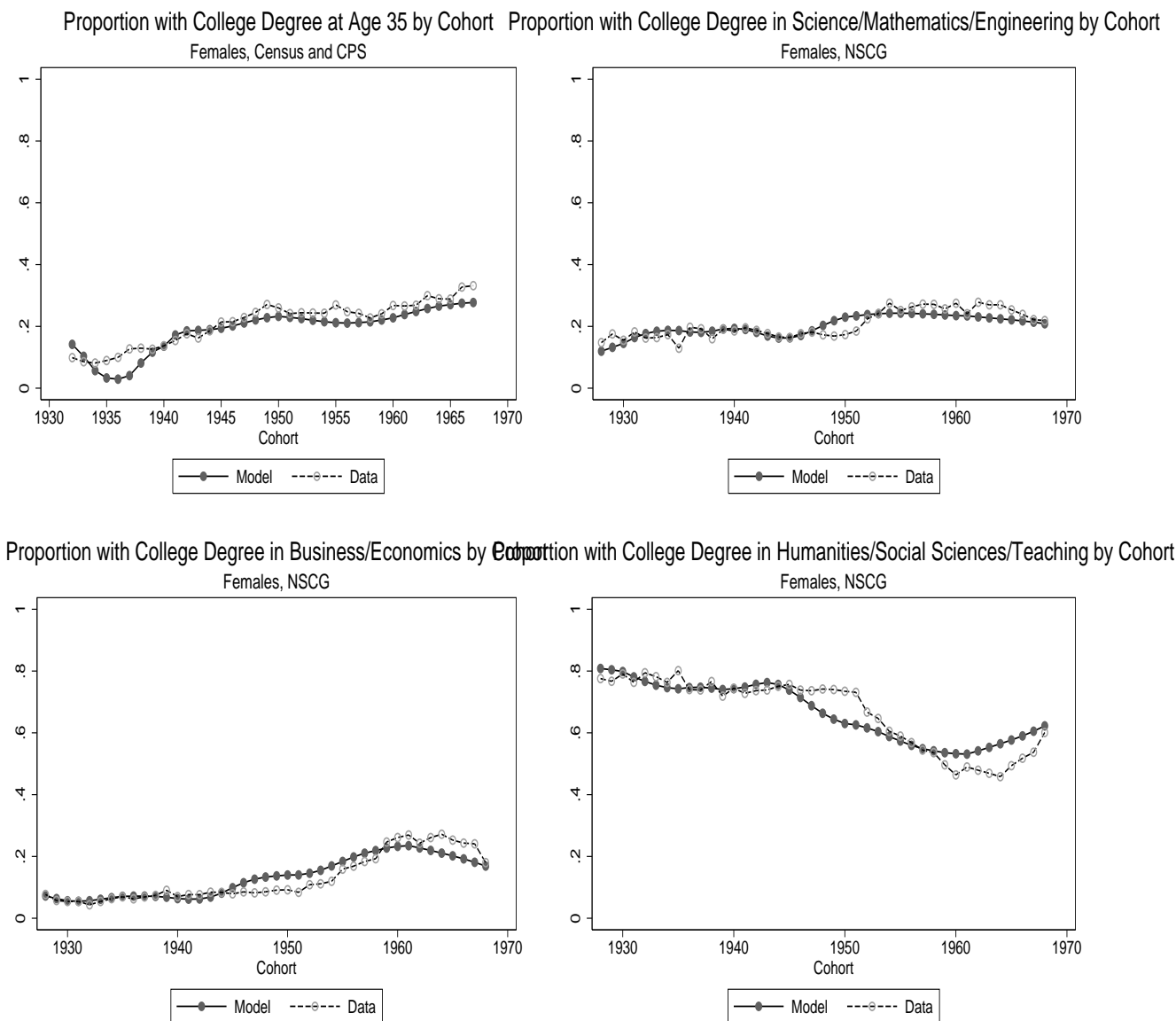


Figure 9: Model Fit

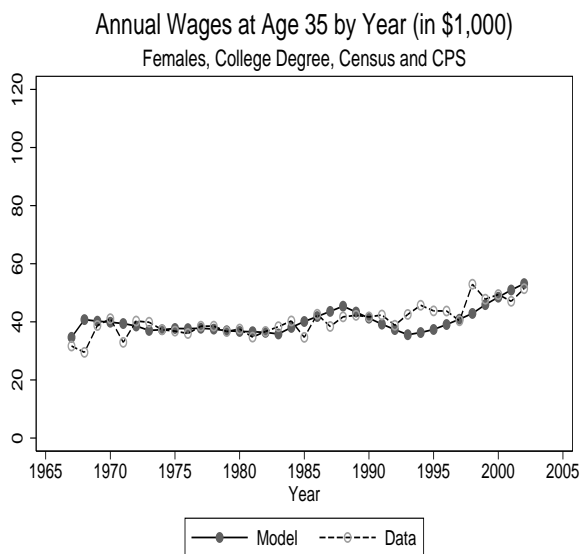
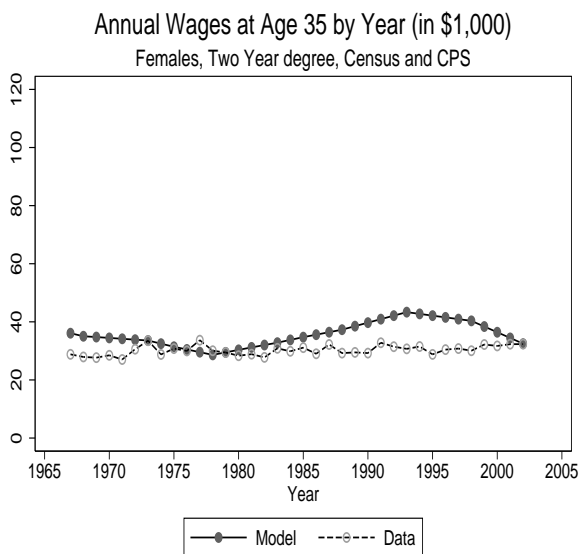
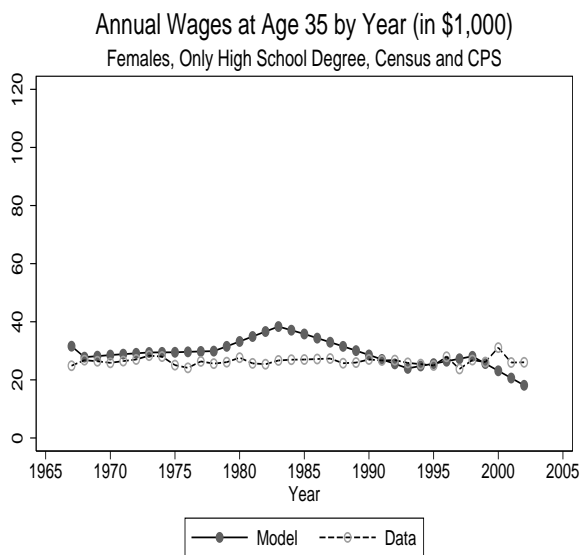
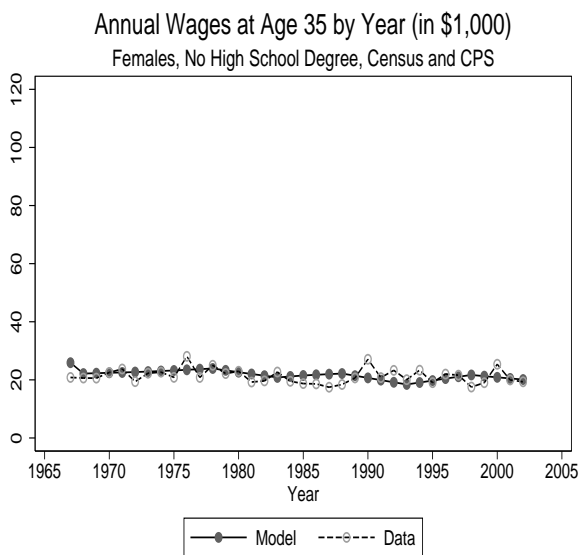


Figure 10: Model Fit

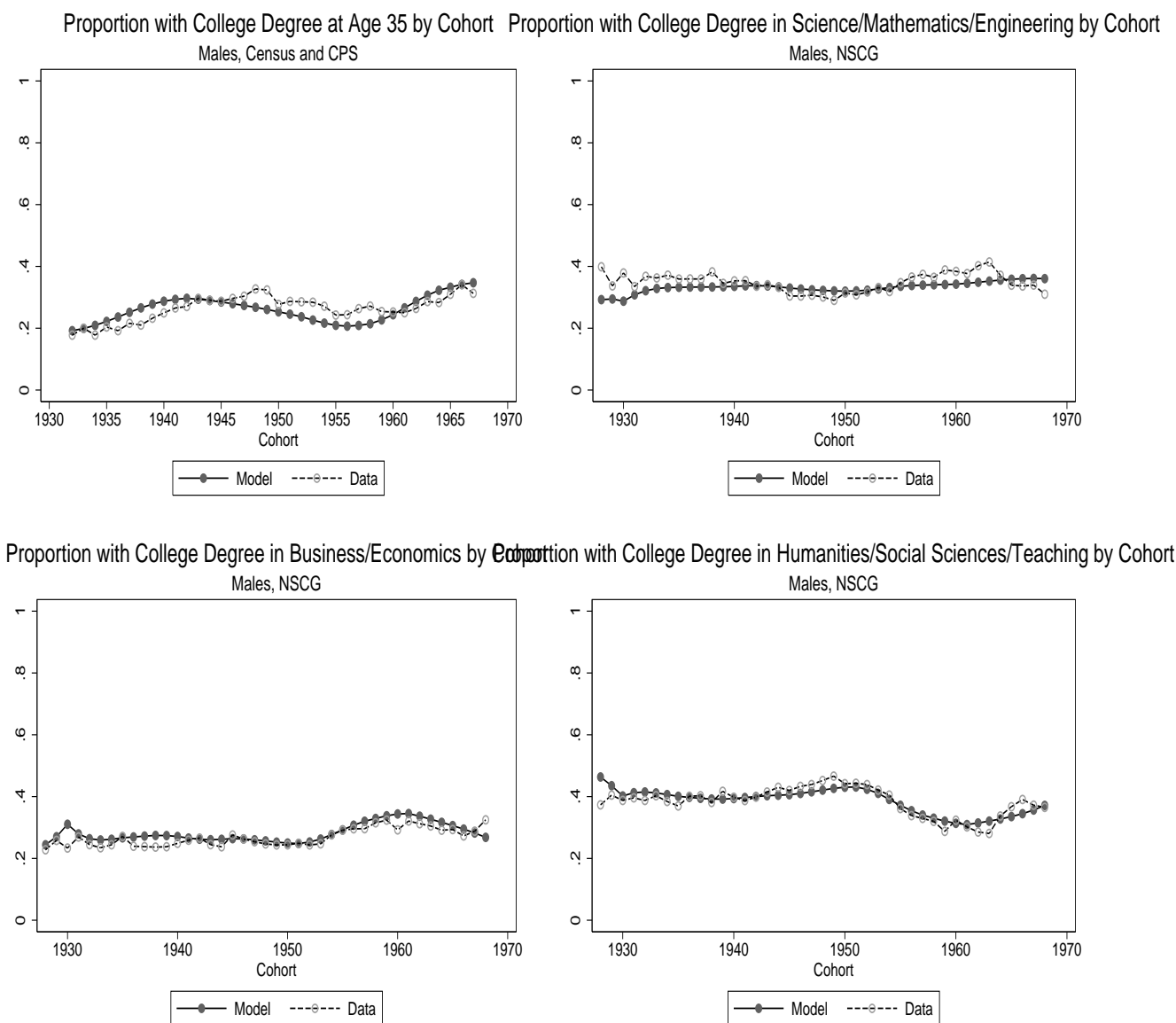


Figure 11: Model Fit

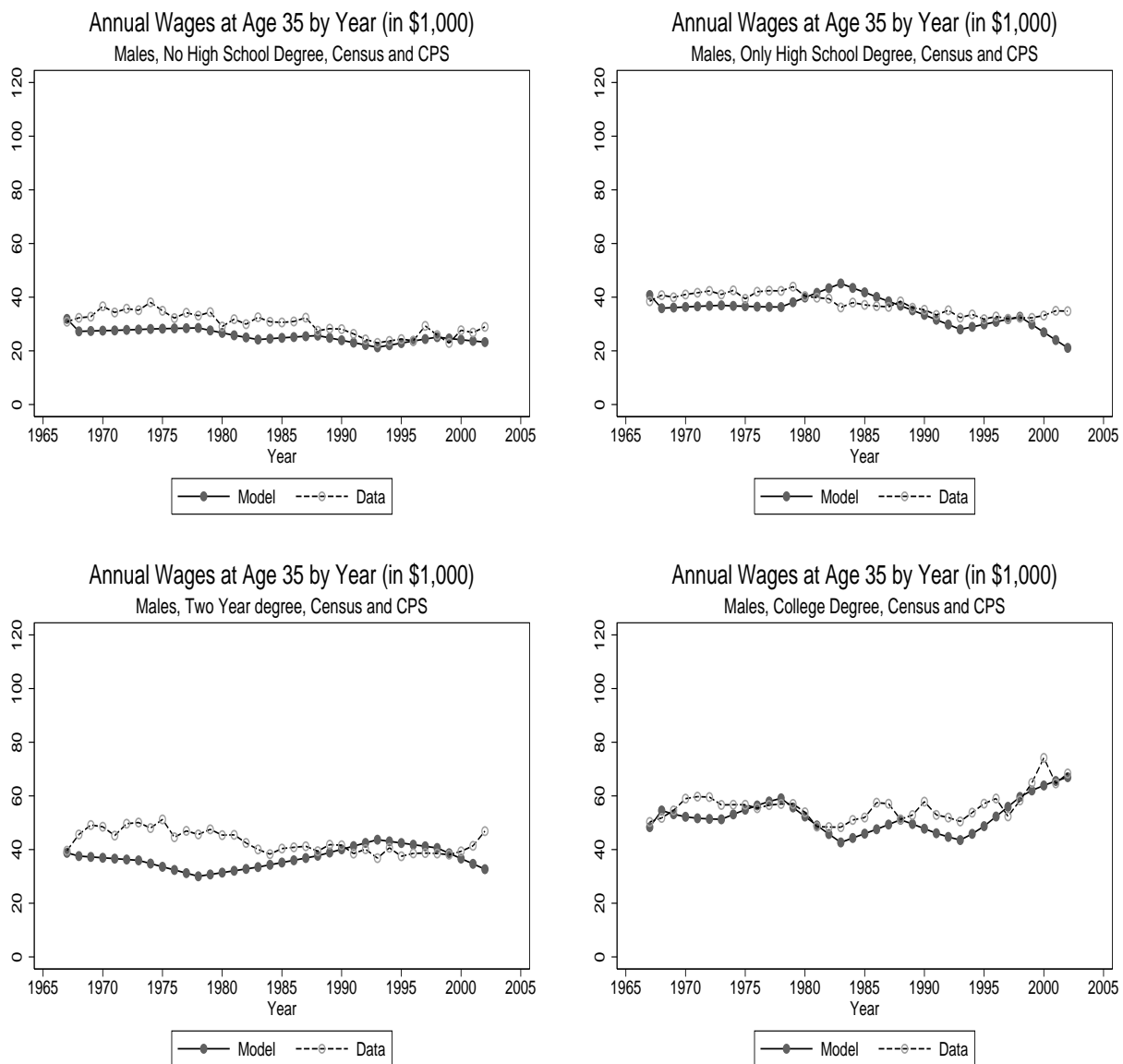


Table 5: Counterfactual Experiments: Determinants of Educational Attainment (1960-1940)

Panel A: Women	(1)	(2)	(3)	(4)	(5)
	Actual Chg. (1960 - 1940)	Baseline No Change	Add Chg. in Skill Prices	Add Chg. in Tuition Rate	Add Chg. in Home Value
% in College	+12	0	+6	-10	+16
% of Col. Grads in ...					
Sci./Math/Eng.	+8	0	+6	-3	+5
Bus./Econ.	+20	0	+31	-25	+14
Hum./Soc.Sci./Teach.	-28	0	-36	+28	-20
Panel B: Men	(1)	(2)	(3)	(4)	(5)
	Actual Chg. (1960 - 1940)	Baseline No Change	Add Chg. in Skill Prices	Add Chg. in Tuition Rate	Add Chg. in Home Value
% in College	-3	0	+16	-21	+2
% of Col. Grads in ...					
Sci./Math/Eng.	+4	0	+7	-7	+4
Bus./Econ.	+2	0	+13	-6	-5
Hum./Soc. Sci./Teach.	-7	0	-20	+12	+1

Notes: This table reports results from a counterfactual experiment in which we progressively add in elements of the model to explain the change in educational attainment from 1940 to 1960 (age 35 men and women in both years). The baseline change is the predicted model change, e.g. +12 is the increase in the percentage of women graduating college from 14 percent to 23 percent. The marginal changes in columns (3)-(5) sum to the baseline change, e.g. percentage women in college $+12 = +6 -10 +16$.

Columns (3)-(5) progressively add in model elements. As an example, consider the first row of Panel A. The first column allows the skill prices to change as estimated from 1940 to 1960, increasing the percentage of women graduating from college by 6 percentage points (from 14 to 20 percent). Next, we add-in changes in tuition rates along with changes in skill prices. This change reduces the percentage of women in college by 10 percentage points. Finally, we add-in all model elements: skill prices, tuition, and home value. The marginal change from the lower home value increases the fraction in college by 16 percentage points. The sum of these marginal changes is the actual predicted change of +12 percentage points.

APPENDICES

A NSCG Data Appendix

A.1 Administration

The 1993 and 2003 National Survey of College Graduates (NSCG) are part of the NSF's SESTAT data collections. See *sestat.nsf.gov* for more information. The 1993 NSCG sample selected individuals from the 1990 Census who reported a baccalaureate or higher education attainment as of April 1, 1990 and were age 72 or younger. In 1993, 216,643 individuals meeting these requirements were mailed the 1993 NSCG survey instrument. The reference week for all questions is April 1, 1993. Non-respondents were later contacted by phone and through in-person interviews. The final response rate was 78 percent. Upon receipt of the completed surveys, an additional 19,224 completed cases were discovered to be ineligible for interview (e.g. deceased, no longer living in the United States, misreported age (now report age over 75), and misreported education level (now report no baccalaureate degree as of April 1, 1990)). Dropping these observations leaves a an initial sample for the 1993 NSCG of 148,905 respondents.¹⁷

The 2003 NSCG was based on a sample of 170,797 individuals from the 2003 Census who reported a baccalaureate degree or higher as of April 1, 2000 and were age 72 or younger. As with the 1993 NSCG, the 2003 NSCG was administered using mailings, phone interviews, and personal visits. The 2003 NSCG was administered between October 2003 and August 2004. The reference week for the survey is October 1, 2003. With a response rate of 63 percent and the exclusion of observations which did not meet the sampling frame (as with the 1993 NSCG), the initial sample for the 2003 NSCG is 100,402.

A.2 Sample Selection

We exclude from the sample the following observations with incomplete or nonsensical information:

- 1) We exclude the 0.07 percent of the sample which listed either no degree information (e.g. did not indicate a specific field of study) or listed "other" as their first degree type.
- 2) We exclude the additional 0.77 percent of the sample which reported that they were 18 or younger at the time of earning any of their degrees. This criteria also excluded respondents which reported earning a degree before their reported year of birth.
- 3) We exclude the additional 0.17 percent of the sample which did not report the year of their high school graduation.

¹⁷The original 1993 NSCG documentation indicates a final sample of 148,932. The data we downloaded from the SESTAT website has only 148,905 observations.

4) We exclude the additional 0.58 percent of the sample which reported that they earned their first college degree less than 1 year after their reported year of high school graduation.

5) We exclude the additional 2.81 percent of the sample which did not report at least one bachelor degree.

This last sample exclusion restriction is the most restrictive and deserves some comment. The observations excluded because of this criteria did not follow the survey instructions to record their first bachelor degree as one of their three reported degrees. Instead, most of these observations recorded three graduate degrees. Some of these respondents did record a bachelor degree as one of their degrees, but indicated that they earned the bachelor degree after they earned the graduate degree. This error may be due to the confusing wording of the survey. Although the third degree category is supposed to include only bachelor degrees, as indicated at the top of the survey instrument, the 1993 and 2003 NSCG survey instruments still allow respondents to check boxes for masters, doctorate, and professional degrees.

There is a strong reason to believe that these observations are not due to random mis-reporting. The individuals excluded are likely to have three or more graduate degrees. Consistent with this hypothesis, in a comparison of these observations with the rest, the sample excluded because of this error has a higher proportion of men and are older on average than the rest of the sample. However, without a first bachelor degree reported, we cannot construct a degree sequence for these respondents.

With these 5 sample restrictions, the initial sample includes 238,344 total individuals observations.

A.3 Field Aggregations

We aggregate the 150 different bachelor degrees into 3 categories:

1) *Science, Mathematics, and Engineering*

a) Mathematical Sciences (mathematics, statistics, computer science, computer programming, operations research)

b) Physical Sciences (physics, chemistry, astronomy, geology, and earth sciences)

c) Biological Sciences (biology, botany, zoology, animal sciences, genetics, environmental sciences)

d) Engineering (all of the engineering sub-fields)

e) Medical Sciences (clinical and counseling psychology, audiology and speech pathology, pharmacy, hospital administration, physical therapy, public health, medical technologies) Medical Sciences does not include nursing.

2) *Business and Economics* (accounting, business administration, actuarial sciences, finance, economics, and marketing)

3) *Humanities, Social Sciences, and Teaching*

a) Social Sciences (psychology (non-clinical and non-counseling), sociology, political science, geography, linguistics, public affairs, international relations, criminology)

b) Humanities and Arts (English, history, fine arts, architecture and design, non-English languages, philosophy) This category also includes the 0.7 percent of degrees which respondents reported the major field as “other fields (not listed)”.

c) Teaching (elementary, secondary, and kindergarten, and pre-school teaching majors, educational counselors, and educational administration)

d) Traditional Female includes nursing, home economics, and social work.

A.4 Multiple Majors and Multiple Bachelor Degrees

About 41 percent of the sample reported a second degree field for at least one of their reported degrees. In addition, about 3.7 percent of the sample reported earning more than one bachelor degree. There are two complications in interpreting this information. First, the second major listed may be a minor field or a true second major. The wording of the survey instrument does not allow the research to distinguish between these possibilities. Second, respondents with one bachelor degree and two majors may choose to record this as two separate bachelor degrees rather than as one degree with two majors.

To deal with multiple majors and multiple bachelor degrees, we make the following changes:

Multiple Bachelor Degrees: We combine second and third bachelor degrees into one bachelor degree if these later bachelor degrees were earned within 2 years of the first bachelor degree. The fields for the second and third bachelor degrees are treated as additional (second, third, etc.) majors or fields for the first bachelor degree. The date of this combined bachelor degree is the date of the last earned bachelor degree.

Combining bachelor degrees in this way eliminated about 56 percent of all second and third bachelor degrees, leaving 2 percent of the sample with an additional bachelor degree. 29 percent of these second and third bachelor degrees have a major within the same aggregate major group based on the 10 aggregate bachelor categories defined above. For the analysis here, we ignore these additional bachelor degrees and statistics calculated are based only on first bachelor degrees.

Multiple Majors: We treat second majors the same as first majors. The analysis uses dummy variables which indicate whether the individual reports earning a degree in a field for either the first or second reported degree field. Therefore, if the respondent reports a second degree field in the same aggregate category as the first degree field, this field is counted only once. For example, an individual reporting a bachelor degree with a first major in philosophy and a second major in English is counted as having earned one humanities bachelor degree. However, if the first and second majors differ, the individual is counted as having earned a degree in both fields. For example, an individual reporting a bachelor degree with a first major in chemistry and a second major in English is counted as having earned a degree in both the Physical Sciences and Humanities. This way of treating first and second majors is the most inclusive, as

it indicates when individuals have completed enough coursework in a field to warrant this field being listed as a first or second major. Without further information, it is not possible to accurately weight the coursework each individual has completed in the different fields.

B Model Solution Appendix

The model is solved through a backward recursion, starting from the last period. In the last period, the agents only have a decision to make between working and staying at home because we set a maximum age $\bar{A} < A$ beyond which the agents cannot attend school. Hence, for all ages $a = \bar{A} + 1, \dots, A - 1$, the value functions simplify to

$$V(\Omega_t(a), a) = \max_{h_t(a)} u_t(a) + \delta V(\Omega_{t+1}(a), a)$$

Once we get to an age below the maximum schooling age, the agent will also have decisions to make regarding attending school. Therefore, in any period $a < \bar{A}$, the Bellman equation takes the form given above (3).

The extreme value distribution for the preference shocks implies that the expectation of the continuation value takes a closed form:

$$EV(\Omega_t(a), a) = \bar{\gamma} + \lambda \ln \sum_{j \in J} \exp\{\bar{V}_j(\Omega_t(a), a)/\lambda\},$$

where $\bar{\gamma}$ is Euler's constant and $\bar{V}_j(\Omega_t(a), a)$ is the non-stochastic portion of the value function given a particular choice j , e.g. for the choice to attend school for degree d if type k , $\bar{V}_j(\Omega_t(a), a) = \gamma_d(k) + \gamma_6 \tau_{d,t} + \delta EV(\Omega_{t+1}(a+1), a+1)$, and $\Omega_{t+1}(a+1)$ is appropriately updated given this choice. Note that for choices not allowed given the current state vector, we set $\bar{V}_j(\Omega_t(a), a) = -\infty$. For example, working is not an option while an individual is in school. We denote the finite set of choices $j \in J$ for convenience here, but the full choice set and constraints are laid out in (3).

With the expectation of the continuation values in hand, we then move to calculating the probability of each choice, conditional on the feasible points in the state space. We calculate these probabilities for each period, birth cohort, and gender. For a given state vector, we use the properties of the extreme value distribution to provide a closed form for the probability of choices $j \in J$:

$$\rho(\text{choice} = j | \Omega_t(a), a) = \frac{\exp\{\bar{V}_j(\Omega_t(a), a)\}}{\sum_{j \in J} \exp\{\bar{V}_j(\Omega_t(a), a)\}}$$

Using these choice probabilities, we obtain the analytical expression for the probabilities of each point in the state space through an iterative procedure starting from age $a = 16$. To simplify the notation, let (y, d, x) be the state vector, where y denotes whether the agent is required to be in school this period given unfinished years of schooling for a previously chosen degree ($y = 1$ required to be in school, $y = 0$ otherwise), d denotes the highest degree he/she has by age a , and x is the accumulated labor market experience by age a . More formally, as discussed in the main text, the state vector involves other elements, including non-stationary skill rental rates, tuition levels, and home values. But we ignore these elements here since the model solution structure is the same for each birth cohort.

Given a certain type k and gender g , we start from $a = 16$ and update the probability of each state space point (y, d, x) at age a according to all the prior sequence of choice probabilities. Let $P(y, d, x | k, g, a)$ denote the probability of the state vector (y, d, x) for type k of gender g at age a , where $P(y, d, x | k, g, a) \in [0, 1]$ and $\sum_{y,d,x} P(y, d, x | k, g, a) = 1$ for all k, g, a . At age 16, the initial conditions for all individuals are $y = 0, d = 0$, and $x = 0$, hence $P(0, 0, 0 | k, g, 16) = 1$. Note that we suppress the calendar time t indices for simplicity since the model solution structure is the same at each period, although the non-stationary elements of the state vector vary.

For all the subsequent periods after age 16, probabilities of observing an agent at a certain state space point given his type, gender and age are updated according to the following:

$$\begin{aligned}
 P(y = 0, d, x | k, g, a) &= P(0, d, x - 1 | k, g, a - 1) \rho(\text{work} | d, x - 1, k, g, a - 1) \\
 &\quad + P(0, d, x | k, g, a - 1) \rho(\text{not work} | d, x, k, g, a - 1) \\
 &\quad + P(1, d', x | k, g, a - 1) \\
 P(y = 1, d, x | k, g, a) &= P(0, d, x | k, g, a - 1) \rho(\text{go to school for } d' | d, x, k, g, a - 1),
 \end{aligned}$$

where $\rho(\cdot)$ denotes the choice probabilities given the decision rule of the agent, defined above. The first expression follows since there are three possible states and choice combinations that lead to an individual not being in school in at age a : i) not being in school at age $a - 1$ and choosing to work at a , ii) not being in school at age $a - 1$ and choosing to stay home, and iii) being in school in $a - 1$ (with highest degree d'), and in the last year of school required to earn a new degree $d \neq d'$, and hence starting age a not in school ($y = 0$). The second expression follows since an individual who decides to attend school in $a - 1$ (to earn a new degree d') must be in school at age a . Note that since degrees in our model last for more than 1 period, and we do not allow individuals to drop-out of school before finishing the degree, the decisions rules are restricted such that an individual must decide to go to school in the next period if she has not completed the necessary years of schooling for this degree. In the formal exposition of the dynamic program, this kind of state dependence in decision making is reflected in the state vector, but we ignore this complication here for simplicity. We mechanically impose these type of state dependence in our computer algorithm to solve the model.

Given the probabilities of each state space point, we calculate the probability of an agent of gender g and age a having degree d by integrating over the probabilities of experience, type, and in schooling state:

$$P(d | g, a) = \sum_k \pi(k, g) \left[\sum_x P(y = 0, d, x | k, a) + P(y = 1, d, x | k, a) \right],$$

This expression follows since there are two possible states in which degree d is the highest degree: i) the individual is not in school at age a ($y = 0$), and ii) the individual is in school for another degree ($y = 1$), but has not completed this degree yet, hence the highest degree remains degree d .

The probability of working and average wages conditional on highest degree is obtained by

$$P(\text{work} \mid d, g, a) = \sum_{y,x,k} \text{weight}^e(y, x, k, g, a) \rho(\text{work} \mid y, d, x, k, a)$$

$$E(\text{wage} \mid \text{work}, d, g, a) = \sum_{x,k} \text{weight}^w(y = 0, x, k, g, a) \text{wage}(d, x, k, a),$$

where $\text{wage}(d, x, k, a)$ is simplified notation for the wage at the given state variables. The weights are given as

$$\text{weight}^e(y, x, k, g, a) = \frac{\pi(k, g)P(y, x \mid d, k, g, a)}{\sum_{y,x,k} \pi(k, g)P(y, x \mid d, k, g, a)}$$

$$\text{weight}^w(y, x, k, g, a) = \frac{\pi(k, g)P(y, x \mid d, k, g, a)\rho(\text{work} \mid y, d, x, k, a)}{\sum_{y,x,k} \pi(k, g)P(y, x \mid d, k, g, a)\rho(\text{work} \mid y, d, x, k, a)},$$

where $P(y, x \mid d, k, g, a)$ is the probability of y and x conditional on having obtained degree d so far for a type k agent at age a :

$$P(y, x \mid d, k, g, a) = \frac{P(y, x, d \mid k, g, a)}{P(d \mid k, g, a)}$$

where $P(d \mid k, g, a) = \sum_{y,x} P(y, d, x \mid k, g, a)$.

The expressions we obtain at the end of this procedure, i.e. $P(\text{work} \mid d, g, a)$, $P(d \mid g, a)$, and $E(\text{wage} \mid \text{work}, d, g, a)$, are computed for each cohort. These model objects correspond to directly observed objects in the data and are then used as the basis of our method of moments estimator.

C Type Parameter Estimates

Table C-1: **Parameter Estimates: Skill and Taste Distribution**

		Parameter	S.E.
$\alpha_d(k), d = 0, k = 1$	Type 1 No High School Skill	0.00	-
$\alpha_d(k), d = 1, k = 1$	Type 1 High School Skill	0.00	-
$\alpha_d(k), d = 2, k = 1$	Type 1 Two Year Skill	0.00	-
$\alpha_d(k), d = 3, k = 1$	Type 1 Science Skill	0.00	-
$\alpha_d(k), d = 4, k = 1$	Type 1 Business Skill	0.00	-
$\alpha_d(k), d = 5, k = 1$	Type 1 Humanities Skill	0.00	-
$\gamma_d(k), d = 1, k = 1$	Type 1 High School Taste	13.43	(0.044)
$\gamma_d(k), d = 2, k = 1$	Type 1 Two Year Taste	14.47	(0.161)
$\gamma_d(k), d = 3, k = 1$	Type 1 Science Taste	13.54	(3.254)
$\gamma_d(k), d = 4, k = 1$	Type 1 Business Taste	20.98	(0.192)
$\gamma_d(k), d = 5, k = 1$	Type 1 Humanities Taste	13.13	(6.079)
$\alpha_d(k), d = 0, k = 2$	Type 2 No High School Skill	-0.11	(0.015)
$\alpha_d(k), d = 1, k = 2$	Type 2 High School Skill	-0.09	(0.055)
$\alpha_d(k), d = 2, k = 2$	Type 2 Two Year Skill	-0.01	(0.028)
$\alpha_d(k), d = 3, k = 2$	Type 2 Science Skill	-0.17	(0.030)
$\alpha_d(k), d = 4, k = 2$	Type 2 Business Skill	-0.13	(0.036)
$\alpha_d(k), d = 5, k = 2$	Type 2 Humanities Skill	-0.13	(0.065)
$\gamma_d(k), d = 1, k = 2$	Type 2 High School Taste	5.77	(0.169)
$\gamma_d(k), d = 2, k = 2$	Type 2 Two Year Taste	-0.02	(0.118)
$\gamma_d(k), d = 3, k = 2$	Type 2 Science Taste	-0.01	(0.394)
$\gamma_d(k), d = 4, k = 2$	Type 2 Business Taste	0.99	(0.355)
$\gamma_d(k), d = 5, k = 2$	Type 2 Humanities Taste	0.03	(8.387)

Notes: “-” is for normalized parameters. Skill prices for each degree are normalized with respect to the type 1 skill levels. See text.

Table C-2: **Parameter Estimates: Skill and Taste Distribution (Continued)**

		Parameter	S.E.
$\alpha_d(k), d = 0, k = 3$	Type 3 No High School Skill	0.31	(0.008)
$\alpha_d(k), d = 1, k = 3$	Type 3 High School Skill	0.10	(0.057)
$\alpha_d(k), d = 2, k = 3$	Type 3 Two Year Skill	-0.03	(0.036)
$\alpha_d(k), d = 3, k = 3$	Type 3 Science Skill	-0.00	(0.042)
$\alpha_d(k), d = 4, k = 3$	Type 3 Business Skill	-0.04	(0.058)
$\alpha_d(k), d = 5, k = 3$	Type 3 Humanities Skill	-0.04	(0.073)
$\gamma_d(k), d = 1, k = 3$	Type 3 High School Taste	0.71	(0.176)
$\gamma_d(k), d = 2, k = 3$	Type 3 Two Year Taste	2.62	(0.186)
$\gamma_d(k), d = 3, k = 3$	Type 3 Science Taste	5.24	(1.757)
$\gamma_d(k), d = 4, k = 3$	Type 3 Business Taste	7.34	(2.076)
$\gamma_d(k), d = 5, k = 3$	Type 3 Humanities Taste	19.17	(0.199)
$\alpha_d(k), d = 0, k = 4$	Type 4 No High School Skill	0.35	(0.014)
$\alpha_d(k), d = 1, k = 4$	Type 4 High School Skill	0.53	(0.046)
$\alpha_d(k), d = 2, k = 4$	Type 4 Two Year Skill	0.00	(0.030)
$\alpha_d(k), d = 3, k = 4$	Type 4 Science Skill	0.14	(0.029)
$\alpha_d(k), d = 4, k = 4$	Type 4 Business Skill	0.07	(0.044)
$\alpha_d(k), d = 5, k = 4$	Type 4 Humanities Skill	0.16	(0.062)
$\gamma_d(k), d = 1, k = 4$	Type 4 High School Taste	0.01	(0.137)
$\gamma_d(k), d = 2, k = 4$	Type 4 Two Year Taste	0.48	(0.340)
$\gamma_d(k), d = 3, k = 4$	Type 4 Science Taste	0.84	(0.430)
$\gamma_d(k), d = 4, k = 4$	Type 4 Business Taste	-0.39	(2.359)
$\gamma_d(k), d = 5, k = 4$	Type 4 Humanities Taste	0.35	(10.473)
$\alpha_d(k), d = 0, k = 5$	Type 5 No High School Skill	-0.02	(0.040)
$\alpha_d(k), d = 1, k = 5$	Type 5 High School Skill	-0.01	(0.065)
$\alpha_d(k), d = 2, k = 5$	Type 5 Two Year Skill	-0.01	(0.041)
$\alpha_d(k), d = 3, k = 5$	Type 5 Science Skill	-0.09	(0.047)
$\alpha_d(k), d = 4, k = 5$	Type 5 Business Skill	-0.19	(0.114)
$\alpha_d(k), d = 5, k = 5$	Type 5 Humanities Skill	-0.18	(0.106)
$\gamma_d(k), d = 1, k = 5$	Type 5 High School Taste	0.56	(0.191)
$\gamma_d(k), d = 2, k = 5$	Type 5 Two Year Taste	-0.03	(0.396)
$\gamma_d(k), d = 3, k = 5$	Type 5 Science Taste	3.47	(0.194)
$\gamma_d(k), d = 4, k = 5$	Type 5 Business Taste	-1.69	(14.588)
$\gamma_d(k), d = 5, k = 5$	Type 5 Humanities Taste	0.68	(44.325)

Table C-3: **Parameter Estimates: Type Probabilities**

		Actual Value ($\pi(k, g)$)	Transformed Value ($\tilde{\pi}(k, g)$)	S.E. of $\tilde{\pi}(l, g)$
$\pi(1, 1)$	Type 1 Probability - Female	0.55	7.79	(0.629)
$\pi(2, 1)$	Type 2 Probability - Female	0.17	2.48	(0.218)
$\pi(3, 1)$	Type 3 Probability - Female	0.17	2.46	(0.165)
$\pi(4, 1)$	Type 4 Probability - Female	0.03	0.49	(0.086)
$\pi(5, 1)$	Type 5 Probability - Female	0.07	0.00	-
$\pi(1, 2)$	Type 1 Probability - Male	0.31	3.52	(0.302)
$\pi(2, 2)$	Type 2 Probability - Male	0.29	3.22	(0.344)
$\pi(3, 2)$	Type 3 Probability - Male	0.13	1.45	(0.130)
$\pi(4, 2)$	Type 4 Probability - Male	0.18	2.08	(0.224)
$\pi(5, 2)$	Type 5 Probability - Male	0.09	0.00	-

Notes: The parameter values for the actual type probabilities ($\pi(k, g)$) are calculated from the transformed values ($\tilde{\pi}(k, g)$) as: $\pi(k, g) = \frac{\tilde{\pi}(k, g)}{\sum_k \tilde{\pi}(k, g)}$.