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

BY AHU GEMICI AND MATTHEW WISWALL †

New York University, U.S.A; Arizona State University, U.S.A

Running Head: Evolution of College Major Choices

Keywords: human capital, gender differences, college majors

JEL Codes: J16, J24, J71

Abstract

While women in the US now complete more college degrees than men, the distribution of college majors among college graduates remains unequal with women about 2/3 as likely as men to major in business or science. We develop and estimate a dynamic, overlapping generations model of human capital investments and labor supply. We allow for specific college major choices, rather than aggregating these choices to the education level. Results show that changes in skill prices, higher schooling costs, and gender specific changes in home value were each important to the long-term trends.

* Manuscript received June 2011; revised September 2012.

† We thank Jennifer Sallman for excellent research assistance. We thank three anonymous referees and the editor, Holger Sieg, for helpful comments. All errors are our own.

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 four-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.[1]

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 cohort born in 1920, women who graduated from college obtained about 84 percent of their degrees in the humanities, social sciences, or teaching fields, and only 11 percent in science, mathematics, or engineering and 5 percent in business or economics. In contrast, college educated men born in the same year had around 41 percent of their degrees in science, mathematics, or engineering, and 27 percent in business. Forty years later, for the cohort born in 1960, the proportion of women earning degrees in science fields nearly doubled to about 20 percent and the proportion in business increased four-fold to 25 percent.[2]

As Figure 3 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 human capital, making a distinction between college degrees with different majors. We define human

[1]Calculated from Census and CPS data, discussed below. See Figure 3. 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 as men.

[2]See Figures 1 and 2.

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 multiple 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.[3]

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 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

[3]Other 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.

the 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 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 tastes we estimate, we find that higher tuition reduced the proportion of individuals who would have otherwise completed science and business majors from entering college at all, 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 the current literature offers several possible candidate explanations, including changes in the price of home goods (Greenwood, Seshadri and Yorukoglu 2005), an 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).[4] 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.

[4]See 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 laws. In addition, Charles and Luoh (2003) argue for the importance of earnings risk differences.

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).[5] 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.

2 Descriptive Evidence

2.1 Trends in Field of Study Composition

Using the NSCG data, Figures 1 and 2 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 whom graduated from college in the 1940s, 84 percent of all college degrees earned by women were in humanities, social sciences, and

[5]More recent studies use expectations data to study college major choices (Zafar 2011; Arcidiacono, Hotz, and Kang 2011; Wiswall and Zafar 2012). 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).

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 41 percent of degrees earned by men in science, engineering, and mathematics, 27 percent in business and economics, and the remaining 32 percent in humanities, arts, and education fields. College educated men born in 1920 were nearly 4 times more likely to major in a science, mathematics, or engineering field than a college educated women born the same year, and over 5 times more likely to major in a business or economics field.

Examining the Figures 1 and 2 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 78 percent from the 1920 to 1950 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 1963 birth cohort, the proportion of degrees in science, mathematics, or engineering grew from 31 percent to 39 percent. During this period there was a concomitant fall in the proportion of men who graduated with humanities degrees, from 45 percent to 30 percent. The fraction in business also increased from 24 to 31 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. Contrasting the cohorts born in 1950 vs. those born in 1963, the proportion of science graduates increased from 13 to 21 percent, and the number of graduates in business increased from 9 percent to 29 percent, a level approaching that of men.[6]

2.2 Gender Differentials in Human Capital Investments

Figure 3 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, gender differences in field of study composition are larger (more unequal) than gender difference in college attainment. Examining only extensive margin

[6] It is important to note that while the field of study composition among college degree holders changed considerably, at the same time there were many more female and male college graduates, hence the *number* of humanities degrees actually increased over this period.

human capital differences greatly understates the gender differences in field of study composition. Second, the trend for the female-to-male ratio in 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.

2.3 Earnings by Field

Table B-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 considerable 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 across fields, we also estimated a regression of log annual earnings on field of study and a full set of age indicators (full results available on request). From this analysis, we see that 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]

[7]These 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.

2.4 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. 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 4 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 the trend in college major composition is strongly related to the trend in the gender earnings gap.

3 Model

3.1 Overview

The economy consists of a single sector and overlapping generations.[8] Time is discrete and individuals make decisions over a finite horizon. Each period or age for an individual is indexed $a = 16, 17, \ldots, A$, where the initial age of

[8]Implicitly, we allow for generational spill-overs through the equilibrium skill prices (one generation's supply of skills affects equilibrium prices for all other overlapping generations), but we do not model this directly. One can refer to this as an "overlapping generations" model for this reason, despite the fact that generations do not directly interact with each other.

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 expected future labor market returns, their own heterogeneous preferences for working in the labor market, and their tastes for various kinds of 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 $d = 1, \ldots, D$, 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:

$$u_t(a) = \begin{cases} \gamma_d(k) + \gamma_6 \tau_{d,t} + \varepsilon^1_{d,t}(a) \text{ if go to school for degree } d \\ \gamma_7 w_{d,t}(k, a) + \varepsilon^2_t(a) \text{ if work} \\ \gamma_8(k) + \gamma_9(g) + \gamma_{10}(g) c_t(g, a) + \gamma_{11}(g)(t - a) + \varepsilon^3_t(a) \text{ if stay at home} \end{cases} \quad (1)$$

The consumption value of going to school for a particular degree d has a time invariant and type specific component $\gamma_d(k), d = 1, \ldots, D$, a component that depends on a time varying tuition cost, $\tau_{d,t}$, as well as a stochastic component, $\varepsilon^1_{d,t}(a)$. 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, $w_{d,t}(k, a)$, and a stochastic component $\varepsilon_t^2(a)$, which reflects idiosyncratic (dis-) utility of working. 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. Note that skill endowments and therefore wages are type specific, and we allow for and estimate different distributions of types by gender. Hence wages are gender specific, as are all type specific elements of the model.

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 normalize the gender specific intercept, $\gamma_9(g)$, for males to 0. 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, who has obtained degree d supplies $s_{d,t}(k, a)$ 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:

$$s_{d,t}(k, a) = \exp\left(\alpha_d(k) + \beta_1 x_t(a) + \beta_2 x_t(a)^2\right) \tag{2}$$

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 $w_{d,t}(k,a)$ at period t and age a is given by:

$$w_{d,t}(k,a) = r_{d,t}s_{d,t}(k,a) \tag{3}$$

where $r_{d,t}$ is the period t rental rate of skill degree class d, and $s_{d,t}(k,a)$ is the level of accumulated degree d skill.[9]

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 $\varepsilon_t(a) = [\varepsilon^1_{1,t}(a), \ldots, \varepsilon^1_{D,t}(a), \varepsilon^2_t(a), \varepsilon^3_t(a)]$. 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, $R_t, R_{t+1}, \ldots, R_{t+A-a}$, where $R_t = \{r_{1t}, \ldots, r_{Dt}\}$, the present and future cost of schooling T_t, T_{t+1}, T_{t+A-a}, where $T_t = \{\tau_{1,t}, \ldots, \tau_{D,t}\}$, and the present and future value of home, which depends on the age and cohort specific fertility rates $\{c_t(a), c_{t+1}(a+1), \ldots, c_{t+A-a}(A)\}$, and an additional cohort trend. Other state variables include the individual's type k (which determines her degree specific skill endowments, and schooling and leisure tastes), total labor market experience at t $x_t(a)$, highest degree already obtained $d_t(a) \in \{0, 1, \ldots, D\}$, and whether the agent was in school for degree d in previous period, $y_{d,t}(a)$.[10]

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

$$\Omega_t(a) = [k, x_t(a), y_{d,t}(a), R_t, R_{t+1}, \ldots, R_{t+A-a},$$

$$c_t(a), c_{t+1}(a+1), \ldots, c_{t+A-a}(A), T_t, T_{t+1}, \ldots, T_{t+A-a}, \varepsilon_t(a)].$$

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

[9] Gender 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.

[10] Note 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.

problem is then

$$V(\Omega_t(a), a) = \max_{h_t(a), q_t(a)} u_t(a) + \delta E[V(\Omega_{t+1}(a+1), a+1)|h_t(a), q_t(a)], \tag{4}$$

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 $q_t(a)^*(\Omega_t(a)) \in \{0,1,\ldots,D\}$. $h_t(a)^*(\Omega_t(a))$ and $q_t(a)^*(\Omega_t(a))$ 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 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, say, 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. As in these studies, modeling the equilibrium in the market for skills in our setting 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 lay 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. In general, changes in each of the non-stationary features of the economy (skill prices, tuition costs, and home value) alters both the number of individuals who choose to go to college *and* college major composition. 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 have different type distributions and therefore different levels of skill endowments, changes in skill prices can affect the labor market return to human capital differently by gender. In addition, if 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 would 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 the number of college graduates. However, the effect on college major composition is ambiguous given that tuition is field of study neutral, with a single cost for all degrees of a given type (we do allow for different tuition levels for 2 and 4 year college 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.

On the other hand, the overall effect of higher tuition on the composition of college majors also depends on the distribution of skills and tastes in the population, which determines who selects into college. For example, higher college tuition will particularly deter enrollment for individuals with an absolute skill advantage across all degrees or with high tastes (relative to working or staying home) across all degrees. These individuals' skill endowments and tastes are similar across different degrees (e.g. no high school, high school, as well as the various college major categories) and therefore have high potential earnings or psychic utilities from attending school regardless of the education level they choose. Therefore, these individuals have a relatively low marginal benefit of obtaining a college degree and are most easily deterred from college enrollment in response to college tuition cost increases.

In contrast, individuals with a comparative advantage in college related skills or tastes relative to non-college are not as responsive to tuition costs. These individuals have only high skill endowments or tastes in college degree categories (e.g. science, business, humanities) and therefore have a relatively high level of potential earnings or utility from going to college. These individuals have a high marginal benefit of obtaining a college degree and are not easily deterred from college enrollment. The factor that determines the impact of college tuition cost increases on college major composition is then the characteristics of the individuals who comprise the pool of each major prior to tuition cost changes. For example, the majors that attract individuals with an absolute advantage across all degrees would then be those most affected by a tuition change. The effect of a tuition cost increase on the gender composition of college majors then depends on gender differences in the distributions of major specific skills and tastes.

3) Home value: In the model, there are three channels through which a reduction in the value of home can impact the number of college graduates and college major composition.

First, a reduction in the value of home has a short-term effect of lowering the opportunity cost of schooling. This short-term effect increases the number of individuals who choose to go college, but does not directly impact college major composition, as the opportunity cost of schooling is the same regardless of the major chosen.

Second, a reduction in the value of home has the long-term effect of increasing expected labor supply and therefore lifetime gains from human capital investment. With increased expected labor supply, degrees with higher pecuniary returns (due to either higher skill rental prices relative to other degrees or the individual's higher skill endowments in these degrees) become more attractive as these degrees are expected to generate more labor income. If the home value for women declines more than for men, women become more similar to men in terms of the importance they place on the pecuniary returns to each degree, and degrees in higher earning fields such as science and business become more attractive.

Third, the long-term effect of increased expected labor supply changes the nature of selectivity into college enrollment, as tastes become less important relative to skill endowments in determining who chooses to go to college. In this regard, the differential selection into college education after the reduction in the home value can also be an important factor in determining how college major composition changes. For example, the implications of lower home value depends on the characteristics of the marginal individual who switches to obtaining a college degree. A lower home value does not necessarily increase the proportion of college graduates in science and business fields if the marginal individual has higher tastes or skills in humanities fields. In this case, a reduction in the value of home could shift the composition of individuals graduating from college toward those 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 intervaled 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 hours 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.[11] 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,

[11] Several 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.

respondents were instructed to write 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).[12] 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 collected 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

[12]Unlike 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.

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.[13]

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 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.[14]

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

[13] Other 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.

[14] In 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.

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 $c = 1, \ldots, C$ and $d = 1, \ldots, D$, 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), $C \geq D$, at least one sequence of rental prices $r_{1t}, \ldots, r_{Dt}$ is consistent with the observed data.*

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:

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

$$\vdots$$

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

QED

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. For $C > D$, there may be more than one solution.

The Lemma establishes identification using contemporaneous wage and college major composition data. In our more general setup, there are actually two sources of identification of skill prices: data on contemporaneous wages (which are directly a function of skill prices) and work and educational choices (which are a function of present and future skill prices). Future prices (prices outside our sample period) cannot be identified using unrealized, future, wage observations, but can in general be identified from the work and education choices of current agents in the sample period since their actions are a function of expected future prices. Hence, younger agents will be reacting to a different set of future skill prices than older agents. The difference in their work and education choices identifies future skill prices. For example, if a large proportion of a particular birth cohort are observed choosing some particular degree, all else being equal, we would infer that these agents anticipate a high skill price for this degree in the future. Given the perfect foresight assumption and the other modeling assumptions, this observed behavior then identifies the future, unrealized, skill price.

Although identification in this fashion is possible, without future wage observations, admittedly the identification of future skill prices is somewhat tenuous. After considerable trial and error, we reluctantly imposed a restriction that skill prices after 2002 are constant and equal to the 2002 level. There are several reasons we chose to make this particular assumption about the skill prices after 2002. First, for early cohorts, skill prices after 2002 do not matter for decisions. Since the parameters characterizing stationary distributions of skills and tastes can be identified using these early cohorts alone, we believe that the assumptions we make about post-2002 skill prices have little effect on our parameter estimates. Second, for later born cohorts, given a reasonable level of discounting, the relatively more distant future prices (i.e. 10 or 20 years after 2002) are not particularly important for contemporaneous decisions, which we observe in our data. Given these factors, we felt that this assumption, while seemingly strong, is not particularly crucial for the identification of the main model parameters, as well as for the counterfactual experiments we conduct using the estimated model.

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 $a = 16, \ldots, A$. 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 $\ln w^*_{dt}(k, a) = \ln w_{d,t}(k, a) + \omega_d$, 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 (2), 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

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

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 B-2 displays the parameter estimates for the experience component of the skill production technology (2). 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 full estimates of our 5 point distribution for the $\alpha_d(k)$ type specific skill function intercept terms are presented in an on-line appendix. 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 $s_{dt}(k, a)$ (for type k 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 $s_{dt}(1, a) = 1$ for all d, t. The wage for type 1 agents with no experience is then $w_{d,t}(1, a) = r_{d,t}$ for all d, t, where the $r_{d,t}$ prices are freely varying parameters we directly estimate, as discussed above.[15]

In order to interpret the heterogeneity in skill levels we estimate, Table B-3 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. 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

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

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.[16]

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

[16]Note that a key identifying assumption is that the initial (age 16) skill distribution is stationary. The implicit assumption we are making is that the investment in human capital (skills) comes only from education choices, including of course degree level and major field. Thus men and women can, and do, differ substantially in their skill levels after age 16 given their endogenous human capital choices. Contradicting this stationarity assumption is the evidence that high school course-taking in science and mathematics increased for girls over this period (see Goldin, Katz, and Kuziemko 2006). We do not model high school course-taking as we do for the post-secondary period. Incorporating field of study selection in high school is an important area for future work.

when examining gender differences in cognitive test scores.[17]

Turning next to the skill rental rate component of earnings, Figure 5 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-specific taste for each degree $\gamma_d(k)$ and ii) a homogeneous non-stationary tuition cost given by $\gamma_6 \tau_{d,t}$. The estimated type specific degree tastes are available in an on-line appendix. 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 B-3 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

[17] Previous research is somewhat mixed but generally supports the conclusion of small innate differences. 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 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.

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 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 B-2 shows that the coefficient on the time varying tuition level variable is $\gamma_6 = -0.0005$. With the marginal utility of income set at $\gamma_7 = 0.00003$,[18], 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. Our model assumes utility is linear so the utility consequences of upfront payment of tuition is not as severe as it would be in a model with non-linear (concave) utility. Hence, we interpret our estimate of the "utility cost" of tuition parameter to reflect that very low consumption in a given period is not penalized in utility terms as much as it would be with a non-linear utility function (and no consumption smoothing because of a possible credit constraint). This utility cost is instead being captured by this high free parameter on the utility cost of tuition. 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 $\gamma_8(k)$ which varies by type k and gender intercept $\gamma_9(g)$ and ii) a non-stationary component consisting of a term that depends on cohort,

[18] Our choice of this particular scaling factor was for computational reasons because the choice probabilities involve expressions including exponential functions (see Appendix for model solution). Evaluating the exponential functions at these large values leads to numerical overflow problems. For this reason, we fix the level of marginal utility of income to a relatively small value 0.00003.

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 B-2 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 B-3 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.

We find that this large gender difference in the value of home is a very important factor in accounting for gender differences in education choices and their responses to various changes in the environment, such as skill rental rates and tuition costs. Due to their lower value of home, men have a higher expected labor supply over their lifetime, so that the pecuniary advantage of degrees like science or business translate into higher lifetime utility gains for them. On the other hand, such pecuniary advantages do not constitute an important factor in decision making for women, as they do not expect to work as much as men over their lifetime. This gives rise to a differential selectivity into educational categories between men and women. While men select into different degrees based on their pecuniary returns determined by skill rental prices or individual-specific skill endowments, women's selection is stronger on taste-related factors. We find that this gender difference in selectivity is an important component of men and women's responses to changes in skill rental prices as well as tuition costs. For example, due to the fact that women's schooling decisions are motivated by their taste for degrees rather than pecuniary returns, we find that they are much less responsive to changes in skill rental prices compared to men. This can be seen in Section 7, where we provide a decomposition of the different channels that give rise to such gender differences.

From Table B-2 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 6 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 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 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.[19]

6.2 Model Fit

In results available upon request, we show that the estimated model fits well the main empirical patterns of interest, including educational attainment, field of study choices, and trends in average wages.

[19] With 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).

7 Counterfactual Experiments

7.1 Counterfactual Experiments I: 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 B-4.

The counterfactual experiments in Table B-4 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 report the 1940 to 1960 *marginal* change in education attainment from each of the non-stationary elements. As an example, consider the first row of Panel A. Column (3) 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 full predicted change in the percentage of women in college given by +12 = +6 -10 +16.

7.1.1 Extensive Margin: College Graduation

Focusing first on the extensive margin of college graduation, in Column (3) of Table B-4 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 B-4. 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 impact of the reduction in home value, and increases in tuition costs dampened this factor.

For men, the across cohort trend in the number of college graduates 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 B-4 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.

There are three reasons men are more "tuition sensitive" in the estimated model. First, 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 skill specific type distributions (available in an on-line appendix). 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. Moreover, Type 2 individuals have higher taste for high school relative to other education categories. The proportion of men who are these types is 38 percent, compared to 24 percent for women. Consequently, compared to women, a larger proportion of men have a higher taste for high school and face a smaller differential in their potential wages across different degrees. Second, our parameter estimates show that the largest male advantage in skill endowments is at lower education levels. In particular, a relatively larger proportion of males is Type 4, which is the type with an advantage in no high school and high school skills. In short, the gender-specific type distributions show that a larger proportion of men have a higher taste for low education levels, are less specialized across degrees and have relatively higher levels of skill endowments in low education levels compared to women. For these reasons, men are more willing to forgo college in response to tuition cost increases.

The estimated model shows that the reduction in value of home for males is considerably smaller than that for females. Hence, the reduction in value of home was a relatively minor factor in accounting for men's college attainment levels and college major composition. For women, however, the reduction in the value of home led to a large increase in the proportion of women graduating from college.

7.1.2 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 shifted 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 B-4. The proportion of women college graduates who studied science/mathematics/engineering increased by 6 percentage points (from 19 percent to 25 percent), and the proportion who studied business/economics increased by 31 percentage points (from 6 percent to 37 percent). Men experienced a less pronounced, but similar, pattern in response to skill price changes.

We next add the increase in tuition rates over this time period, so that Column (4) of Table B-4 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 complete 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. We find that the impact of a reduction in the value of home on college major composition is mainly due to its implications for expected labor supply. A reduction in the value of home has the effect of increasing expected labor supply, which renders majors with higher pecuniary returns more attractive. Consequently, individuals put more importance on the pecuniary returns to each degree and start choosing majors in higher earning fields such as science and business. The estimated model shows that the extent to which home value decreased over time is much larger for women than men and this is the reason why it is a considerably more important factor in accounting for the shift in women's college major composition compared to men.

7.2 Counterfactual Experiments II: Determinants of Gender Wage Gap and College Premium

We next turn to a series of counterfactual experiments to examine how the non-stationary elements of the model contribute to changes in the ratio of female-to-male earnings and the college premium for men and women. In general, there are three channels through which skill prices, tuition, and home value changes affect earnings: i) directly, skill price changes affect wage levels even with human capital composition remaining fixed; ii) as discussed above, these changes in the economy alter the educational choices, both extensive and intensive college major choices, which directly changes the average human capital level of men and women and hence the average earnings of these groups; and iii) indirectly, each of these forces changes the labor supply decisions of different cohorts, which in turn affects who among the population is represented in average earnings and the endogenously accumulated level of labor market experience.

7.2.1 Gender Wage Gap

Table B-5 reports results from a counterfactual experiment in which we progressively add in elements of the model to explain the change in the gender wage gap and college premium from 1940 to 1960 (age 35 men and women in both cohorts). We calculate the female-to-male ratio as the average wages of college graduate females to that of college graduate males. The first column is the full predicted model change: +8 is the increase in the female-to-male wage ratio for college graduates, from 69 percent to 77 percent. During this period the male wage advantage shrunk by 8 percentage points.

To explain this overall change in the gender wage gap, we first add in skill prices changes as estimated from 1940 to 1960. The skill price changes alone actually *increase* the male advantage, and cause the female-to-wage ratio for college graduates to decrease by 5 percentage points, from 69 to 64 percent. Since the overall direction of skill price changes during this period was an increase in the price of science and business skills relative to humanities, college educated men benefited more from this change than college educated women, who have a large share of their degrees in humanities fields. Note that as we allow the skill prices to change, both men and women endogenously change their educational choices, as shown in Table B-4. Therefore, this result takes into account endogenous human capital choices in response to skill price changes, but holds tuition and home value fixed.

Next, we add-in changes in tuition rates along with changes in skill prices. This change affects the wages indirectly through education choices, including the types of individuals who enter college at all on the extensive margin and the choices of college majors on the intensive margin. The overall effect of tuition price increase is relatively modest. Tuition price changes increase the male advantage by about 2 percentage points.

Column (5) shows that the fall in value of home is the most important factor in accounting for the change in female-to-male wage ratio, increasing the female-to-male wage ratio by 15 percentage points. As discussed above, changes in home value effect labor supply choices and indirectly the composition of individuals who completing college degrees. As we show above the changes in home value are small for men, and the most significant impact is on women, causing more women to complete college degrees and choose majors with higher monetary returns. This effect directly increases the average wage of female college graduates much more so than for men, thereby helping to reduce the gender gap in earnings.

7.2.2 College Premium

Table B-5 also examines the college premium for men and women. We calculate the college premium as the average wages of college graduates to that of high school graduates, separately for females and males. For the 1940 cohort, the college premium is 1.28 and 1.50 for females and males, respectively. This is partly a result of the fact that the composition of major choices for college graduates is considerably different by gender. For the older cohorts, females chose humanities majors with low monetary returns, and males chose majors with high monetary returns majors in science and business. By the 1960 birth cohort, the college premium increased by 18 percent for women and 13 percent for men to 1.46 and 1.63, respectively.

Decomposing the changes in the college premium due to the non-stationary elements of our model, we see in Column (3) that changes in skill prices increases the college premium of females by 24 percentage points from 1.28 to 1.52. While the changes in skill prices had a relatively modest effect on college completion, the shift toward higher paying majors (as discussed above) had a large effect on the average earnings of college graduates. Thus, not only were the skill prices for the science and business majors increasing but a higher proportion of women completed these degrees. Both effects pushed the college premium for women higher.

For men, the effect of skill price changes was even larger. Skill price changes alone would have increased the college premium for men by 40 percentage points. The effect is larger for men because men are more sensitive to skill price changes (due largely to higher anticipated labor supply) and their college major composition consisted of science and business majors which experienced the largest relative price increases.

In Column (4), adding in tuition changes reduces the college premium by 25 percentage points for women and 27 percentage points for men. This reduction in the college premium for both genders mirrors the effect higher tuition costs have in pushing college graduates toward lower paying humanities degrees, as can be seen in Table B-4.

Finally in Column (5), adding in home value changes increases the college premium of females by 19 percentage points, while it has no impact on the college premium of males. As the value of home falls, women start choosing majors with higher monetary returns due to their increased expected labor supply, as we document in Table B-4. In addition, a fall in home value also increases labor market experience and hence the skill levels and earnings of men and women.

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.

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.[20]

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

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

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.

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: In terms of field aggregations, we treat second majors the same as first majors. However since the model allows individuals to choose only one major, we need to make an assumption about which major to assign to each individual. If a respondent reports a second major, then we assign the major using the following ordering: (1) science, (2) business, (3) humanities. For example, if an individual reports earning both a science and humanities major, we classify the individual as earning a science degree.

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 $a = \bar{A} + 1, \ldots, A - 1$, 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 (4).

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 (4).

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 $j \in J$:

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

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 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. However, we ignore these elements here since the model solution structure is the same for each birth cohort.

Given each 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 the prior sequence of choice probabilities using the following procedure:

Let $P(y, d, x \mid k, g, a)$ denote the probability of state vector (y, d, x) for an individual of type k, gender g and age a, so that $\sum_{y,d,x} P(y, d, x \mid k, g, a) = 1 \quad \forall k, g, a$. At age 16, the initial condition for all individuals is $y = 0$, $d = 0$, and $x = 0$, hence $P(0, 0, 0 \mid k, g, a = 16) = 1$. Note that we also 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 subsequent periods after age 16, probabilities of observing an agent at a certain state space point (y', d', x') given her type, gender and age are updated according to,

$$\begin{aligned} P(y' = 0, d' = d, x' = x \mid k, g, a') \quad &= \quad P(0, d, x - 1 \mid k, g, a) \quad \rho(\text{work} \mid d, x - 1, k, g, a) \\ &\quad + P(0, d, x \mid k, g, a) \qquad \rho(\text{stay home} \mid d, x, k, g, a) \\ &\quad + \sum_{\bar{d} \neq d} P(1, \bar{d}, x \mid k, g, a) \end{aligned} \tag{B-1}$$

where $\rho(.)$ denotes the choice probability given the decision rule of the agent, defined above. The components of expression B-1 represent the three possible states and choice combinations that give rise to $(y' = 0, d' = d, x' = x)$. These are:

1. Given $y = 0$ (out of school), d, $x - 1$, agent chooses to work.
2. Given $y = 0$, d, x, agent chooses to stay home.
3. Given $y = 1$ (in school for a previously chosen degree d), $\bar{d}$ (highest degree obtained as of age a), x, agent remains in school.

Similarly, $P(y' = 1, d' = d, x' = x \mid k, g, a)$ is calculated as,

$$P(y' = 1, d' = d, x' = x \mid k, g, a') = \sum_{\tilde{d} \neq d} P(0, d, x \mid k, g, a)\ \rho(\text{attend school for } \tilde{d} \mid d, x, k, g, a) \tag{B-2}$$

In other words, agent's school attendance status is updated to $y' = 1$ if she chooses to attend school to obtain a degree $\tilde{d}$ in the previous period. The degree variable at age a' remains at d (e.g. high school) and is not changed in subsequent periods until the years required to complete the new degree $\tilde{d}$ (e.g. college) are finished. More generally, degrees in our model last for more than one period and we do not allow individuals to drop out of school before finishing their degree. Therefore, decisions rules are restricted so that an individual must decide to go to school in subsequent periods until graduation. 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 this type of state dependence in our computer algorithm to solve the model.

Using the above probabilities given by B-1 and B-2, we calculate the probability of an agent of gender g and age a having degree d by integrating over type k, experience level x and schooling state y,

$$P(d \mid g, a) = \sum_k \pi(k \mid g) \left[\sum_x P(y = 0, d, x \mid k, g, a) + P(y = 1, d, x \mid k, g, a) \right] \tag{B-3}$$

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 it yet, hence the highest degree remains degree d. $\pi(k \mid g)$ denotes the gender-specific type probabilities.

The probability of working and average wages are calculated as,

$$
\begin{aligned}
P(\text{work} \mid d, g, a) = \quad & \sum_{x,k} P(y, x, k \mid d, g, a) \quad \rho(\text{work} \mid y, d, x, k, g, a) && \text{given } y = 0 \\
E(\text{wage} \mid \text{work}, d, g, a) = \quad & \sum_{x,k} P(y, x, k \mid d, g, a) \quad \rho(\text{work} \mid y, d, x, k, g, a) \ \text{wage}(y, d, x, k, g, a) && \text{given } y = 0
\end{aligned}
$$

where wage(y, d, x, k, g, a) is simplified notation for wages at the given state variables. The probability weights above, which are used to aggregate individual employment probabilities and wages, are calculated using the probabilities of state space points, i.e. $P(y, d, x \mid k, g, a)$ given by B-1 and B-2. For example,

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

where $P(d \mid g, a)$ is given by B-3. The expressions we obtain at the end of this procedure, such as $P(\text{work} \mid d, g, a)$ and $E(\text{wage} \mid \text{work}, d, g, a)$, are computed for each cohort. These model objects correspond directly to observed objects in the data and are then used as the basis of our method of moments estimator.

References

ARCIDIACONO, P., “Ability Sorting and the Returns to College Major,” *Journal of Econometrics* 121 (2004), 343–75.

ARCIDIACONO, P., V. J. HOTZ AND S. KANG, “Modeling College Major Choice using Elicited Measures of Expectations and Counterfactuals,” *Journal of Econometrics* (2011).

BAILEY, M. J., “More Power to the Pill: The Impact of Contraceptive Freedom on Women’s Life Cycle Labor Force Participation,” *Quarterly Journal of Economics* 121 (2006), 289–320.

BANK, B. J., R. L. SLAVINGS AND B. J. BIDDLE, “Effects of Peer, Faculty, and Parental Influences on Students’ Persistence,” *Sociology of Education* 63 (July 1990), 208–225.

BERGER, M. C., “Predicted Future Earnings and Choice of College Major,” *Industrial and Labor Relations Review* 41 (April 1988), 418–29.

BETZ, N., “What Stops Women and Minorities from Choosing and Completing Majors in Science and Engineering,” in D. Johnson, ed., *Minorities and Girls in School: Effects on Achievement and Performance* (Russell Sage Publications, 1997).

BLACK, D., A. HAVILAND, S. SANDERS AND L. TAYLOR, “Gender Wage Disparities among the Highly Educated,” *working paper* (2004).

BLACK, D., S. SANDERS AND L. TAYLOR, “Measurement of Higher Education in the Cenus and CPS,” *Journal of American Statistical Association* 98 (September 2003), 545–54.

BLAKEMORE, A. E. AND S. A. LOW, “Sex Differences in Occupational Selection: The Case of College Majors,” *Review of Economics and Statistics* 66 (February 1984), 157–163.

BROWN, C. AND M. CORCORAN, “Sex Based Differences in School Content and the Male-Female Wage Gap,” *Journal of Labor Economics* 15 (July 1997), 431–65.

CARD, D. AND T. LEMIEUX, “Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis,” *Quarterly Journal of Economics* 116 (May 2001), 705–46.

CHARLES, K. K. AND M.-C. LUOH, “Gender Differences in Completed Schooling,” *Review of Economics and Statistics* 85 (2003), 559–77.

EIDE, E., “College major choice and changes in the gender wage gap,” *Contemporary Economic Policy* 12 (1994), 55–64.

EIDE, E. AND G. WAEHRER, “The Role of the Option Value of College Attendance in College Major Choice,” *Economics of Education Review* 17 (1998), 73–82.

FERNANDEZ, R., A. FOGLI AND C. OLIVETTI, “Mothers and Sons: Preference Formation and Female Labor Force Dynamics,” *Quarterly Journal of Economics* 119 (November 2004), 1249–1299.

FERNANDEZ, R., N. GUNER AND J. KNOWLES, “Love and Money: A Theoretical and Empirical Analysis of Household Sorting and Inequality,” *Quarterly Journal of Economics* 120 (January 2005), 273–344.

FIGLIO, D. N., “Why Barbie Says ’Math is Hard’,” *working paper* (2005).

FREEMAN, R., “Legal Cobwebs: A Recursive Model of the Market for Lawyers,” *Review of Economics and Statistics* (May 1975a).

———, “Supply and Salary Adjustments to the Changing Science Manpower Market: Physics,1948-1975,” *American Economic Review* 65 (March 1975b), 27–39.

———, “A Cobweb Model of the Supply and Starting Salary of New Engineers,” *IndustrialLabor Relations Review* (January 1976).

GAO, *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, 2004).

GOLDIN, C., “From the Valley to the Summit: A Brief History of the Quiet Revolution that Transformed Women’s Work,” *Regional Review* 14 (2005), 5–12.

GOLDIN, C. AND L. F. KATZ, “The Power of the Pill: Oral Contraceptives and Women’s Career and Marriage Decisions,” *Journal of Political Economy* 110 (2002), 730–70.

GOLDIN, C., L. F. KATZ AND I. KUZIEMKO, “The Homecoming of American College Women: The Reversal of the College Gender Gap,” *Journal of Economic Perspectives* 20 (2006a), 133–56.

———, “More Power to the Pill: The Impact of Contraceptive Freedom on Women’s Lifecycle Labor Supply,” *Quarterly Journal of Economics* 121 (2006b), 289–320.

GREENWOOD, J., A. SESHADRI AND M. YORUKOGLU, “Engines of Liberation,” *Review of Economic Studies* 72 (2005), 109–133.

GROGGER, J. AND E. EIDE, “Changes in College Skills and the Rise in the College Wage Premium,” *Journal of Human Resources* 30 (1994), 280–310.

HEATHCOTE, J., K. STORESLETTEN AND G. VIOLANTE, “The Macroeconomic Implications of Rising Wage Inequality in the United States,” *Journal of Political Economy* 118 (August 2010), 681–722.

HEDGES, L. V. AND A. NOWELL, “Sex Differences in Mental Test Scores, Variability, and Numbers of High-Scoring Individuals,” *Science* 269 (July 1995), 41–45.

HUNT, J., “Why Do Women Leave Science and Engineering?,” *NBER Working Paper 15853* (2010).

HYDE, J. S., “The Gender Similarities Hypothesis,” *American Psychologist* 60 (September 2005), 581–92.

JACOB, B. A., “Where the boys arent: non-cognitive skills, returns to school and the gender gap in higher education,” *Economics of Education Review* 21 (2002), 589–98.

KATZ, L. F. AND K. M. MURPHY, “Changes in Relative Wages, 1963-87: Supply and Demand Factors,” *Quarterly Journal of Economics* 107 (February 1992), 35–78.

KRUSELL, P., L. E. OHANIAN, J.-V. RIOS-RULL AND G. L. VIOLANTE, “Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis,” *Econometrica* 68 (September 2000), 1029–1053.

KRUSELL, P. AND A. SMITH, “Income and Wealth Heterogeneity in the Macroeconomy,” *Journal of Political Economy* 106 (1998), 867896.

LEE, D., “An Estimable Dynamic General Equilibrium Model of Work, Schooling and Occupational Choice,” *International Economic Review* 46 (2005), 1–34.

LEE, D. AND K. WOLPIN, “Intersectoral Labor Mobility and the Growth of Service Sector,” *Econometrica* 74 (2006), 146.

———, “Accounting for Wage and Employment Changes in the U.S. from 1968-2000: A Dynamic Model of Labor Market Equilibrium,” *Journal of Econometrics* 156 (2009), 68–85.

LESLIE, L. L., G. T. MCCLURE AND R. L. OAXACA, "Women and Minorities in Science and Engineering: A Life Sequence Analysis," *Journal of Higher Education* 69 (May-June 1998), 239–76.

MACHIN, S. AND P. PUHANI, "Subject of degree and the gender wage differential: evidence from the UK and Germany," *Economics Letters* 79 (June 2003), 393–400.

MONTMARQUETTE, C., K. CANNINGS AND S. MAHSEREDJIAN, "How do young people choose college majors?," *Economics of Education Review* 21 (2005), 543–556.

NSF, *New Formulas for America's Workforce: Girls in Science and Engineering* (National Science Foundation, 2003).

PAGLIN, M. AND A. M. RUFOLO, "Heterogeneous Human Capital, Occupational Choice, and Male-Female Earnings Differences," *Journal of Labor Economics* 8 (January 1990), 123–44.

RUMBERGER, R. W. AND S. THOMAS, "The Economic Returns to College Major, Quality, and Performance: A Multilevel Analysis of Recent Graduates," *Economics of Education Review* 12 (March 1993), 1–19.

SHORE, S. AND R. E. SAKS, "Risk and Career Choice," *Advances in Economic Analysis & Policy* 5 (2005).

SIOW, A., "Occupational Choice under Uncertainty," *Quarterly Journal of Economics* 52 (May 1984), 631–45.

WEINBERGER, C. J., "Race and Gender Wage Gaps in the Market for Recent College Graduates," *Industrial Relations* 37 (January 1998), 67–84.

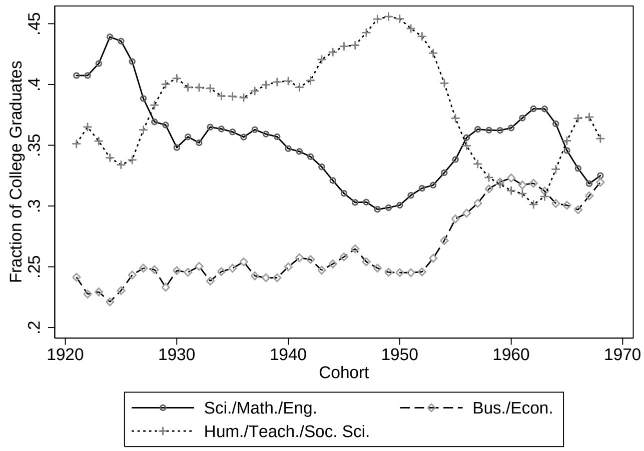
WISWALL, M. AND B. ZAFAR, "Determinants of College Major Choices: Identification using an Information Experiment," *working paper* (2011).

XIE, Y. AND K. SHAUMAN, *Women in Science: Career Processes and Outcomes* (Harvard University Press, 2003).

ZAFAR, B., "How do College Students Form Expectations?," *Journal of Labor Economics* 29 (2011), 301–48.

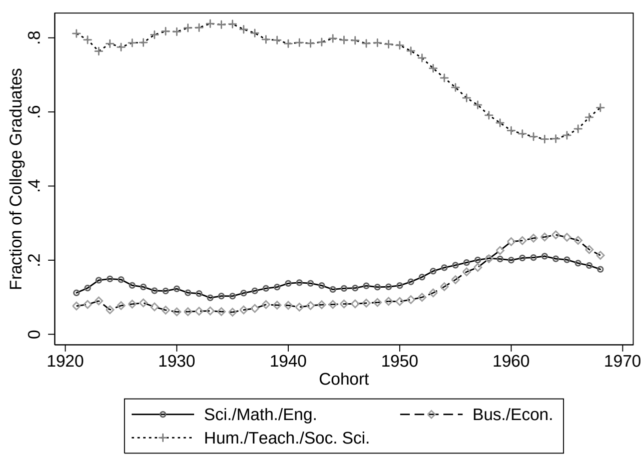
ZARKIN, G. A., "Occupational Choice: An Application to the Market for Public School Teachers," *Quarterly Journal of Economics* 100 (May 1985), 409–446.

Figure 1: Field of Study Composition of Bachelor Degrees by Year of Birth (Men)



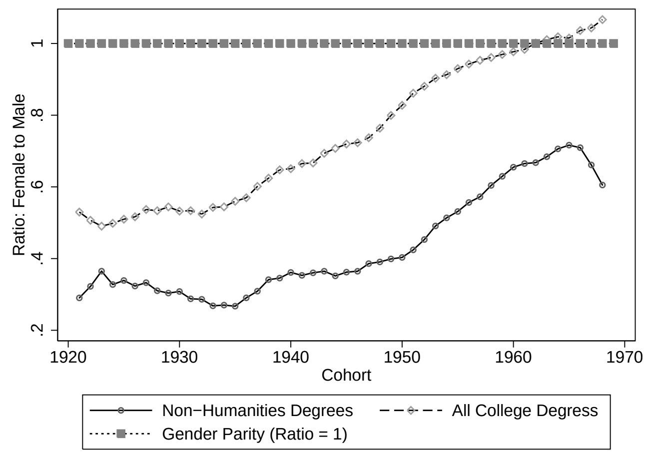
Source: 1993 and 2003 NSCG data.

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



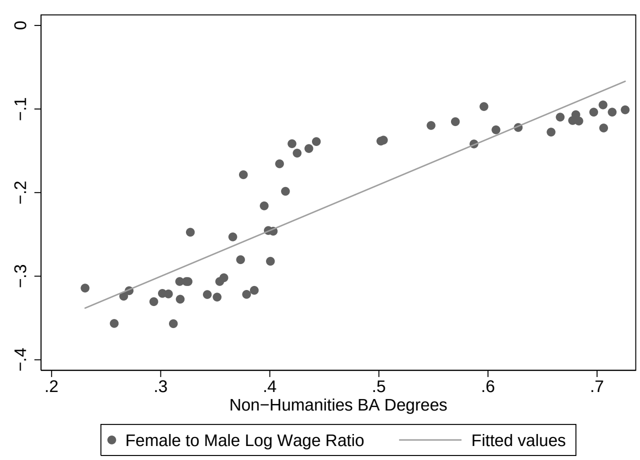
Source: 1993 and 2003 NSCG data.

Figure 3: Ratio of Female-Male College Completion and Non-Humanities Fields



Source: 1993 and 2003 NSCG data.

Figure 4: Cohort Effect Earnings vs. Cohort Fraction in Non-Humanities Degrees



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

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.

Figure 5: Skill Rental Rates

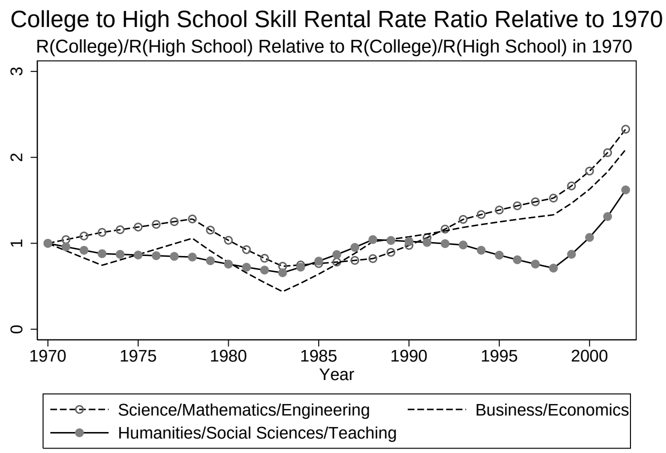


Figure 6: Average Home Value for Women

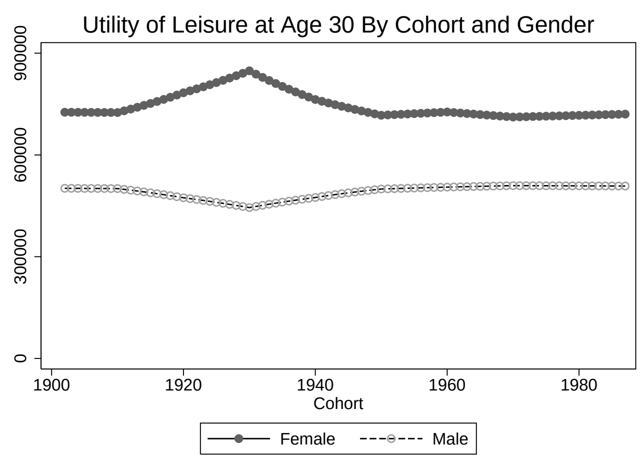


Table B-1: Descriptive Statistics for NSCG Sample of College Graduates

	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			

Source: 1993 and 2003 NSCG data.

Table B-2: 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.

Table B-3: 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 $d = 0,\ldots,5$ 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 $d = 1,\ldots,5$ 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.

Table B-4: Counterfactual Experiments: Determinants of Educational Attainment (1960-1940)

Panel A: Women	(1) Full Chg. (1960 - 1940)	(2) No Chg.	(3) Add Chg. in Skill Prices	(4) Add Chg. in Tuition Rate	(5) 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) Full Chg. (1960 - 1940)**	**(2) No Chg.**	**(3) Add Chg. in Skill Prices**	**(4) Add Chg. in Tuition Rate**	**(5) 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

Table B-5: Counterfactual Experiments: Determinants of Gender Wage Gap and College Premium (1960-1940)

	(1) Full Chg. (1960-1940)	(2) No Chg.	(3) Add Chg. in Skill Prices	(4) Add Chg. in Tuition Rate	(5) Add Chg. in Home Value
Female-to-Male College Wage Ratio	+8	0	-5	-2	+15
College Premium - Females	+18	0	+24	-25	+19
College Premium - Males	+13	0	+40	-27	0