

Occupational Complexity, Experience, and the Gender Wage Gap

Elisa Keller *

Preliminary

February 2013

ABSTRACT

I explore the role of individuals' skills and work experience in explaining the gender wage gap across occupations. I use the O*NET dataset to build an index of occupational complexity: the ratio of abstract to manual tasks. The ratio of female to male wages is U-shaped across occupations ordered by increasing complexity. The U-shape flattens over the lifecycle and across successive cohorts. I develop an occupational choice model with male and female individuals who are heterogeneous with respect to their level of skill. An individual's skill at a point in time depends on his/her exogenous initial level of skill and his/her work experience. Individuals decide how much time to spend in the labor market. Occupations differ by two features in my model: 1) the skill required to perform, and 2) the marginal product of skill. If occupations involve simple tasks, output and wages vary little across workers of different initial skill levels. Also, acquired work experience influences wages only slightly, since little can be learned by performing simple tasks. I discipline the model with data on occupational complexity, occupational choice, labor supply and male wages. The model reproduces the gender wage gap across occupations for cohorts born between 1915 and 1955. The little work experience of females relative to that of males is a key factor behind the U-shape. It decreases female wages disproportionately across occupations and it influences female occupational selection. I find that 69% of the lifecycle gender wage gap is attributable to work experience. Removing differences in work experience between genders results in a larger fraction of females choosing occupations for which the gender wage differential is smaller.

JEL: J1, J2, J3.

Key words: Occupation. Gender wage gap. Human capital. Earnings inequality.

*University of Iowa, Department of Economics. Email: elisa-keller@uiowa.edu I am indebted to B. Ravikumar and Guillaume Vandenbroucke for their continual guidance and support. I thank Todd Schoellman, and the participants at the 2012 Midwest Macro Meetings for their helpful comments.

1 Introduction

On average, females earn less than males. While there is no dispute on the existence of a gap in wages between males and females, there is one as to what drives these differences. In this paper, I use data on the tasks content of occupations to study the roles of skills and work experience in determining the gender wage gap. The task content of an occupation reveals two types of information about the individual employed in the occupation: the lower bound of his/her level of skill and the value of his/her skill. The intuition is straightforward. If an individual is employed in an occupation involving tasks of a certain complexity, then he/she needs to possess at least enough skills to perform these tasks at the lowest accepted standard. If an occupation involves simple tasks, there is little value to skill, both endowed and accumulated through work experience, since there is little room for improvement when performing simple tasks.

I use the Occupational Information Network (O*NET) dataset to construct an index of occupational complexity: the ratio of abstract to manual tasks content. I then combine these data with Census data on occupations and wages to document patterns of the gender wage gap across occupations of different complexity, for cohorts born between the years 1915 and 1955. I focus on two patterns in particular. First, the ratio of female to male wages for young individuals is U-shaped across occupations ordered by increasing complexity. The U-shape becomes flatter for successive cohorts of young individuals. Second, over the lifecycle, the ratio of female to male wages decreases faster for more complex occupations. This decrease becomes weaker for successive cohorts over the lifecycle.¹ In addition, the data reveals that females, especially those of earlier cohorts, systematically choose occupations with low wages relative to males.

The purpose of this paper is to investigate the determinants of the above mentioned patterns of the gender wage gap by occupational complexity in order to shed lights on the driving factors behind the overall gender wage gap. I argue that gender wage gap and occupational choices need to be analyzed jointly and develop a quantitative theory that delivers both as a result of differences in endowed skills and work experience between males and females.

¹Section 2 discusses these patterns of the gender wage gap by occupational complexity in detail.

A counterfactual experiment reveals that work experience accounts for 69 percent of the lifecycle gender wage gap. Removing differences in work experience between genders results in a larger fraction of females choosing occupations for which the difference in wages between males and females is smaller.

I develop an occupational choice model with overlapping generations of female and male individuals who are heterogeneous in their level of skill. An individual's skill at a point in time depends on his/her exogenous initial level of skill along with his/her work experience. Individuals divide their time between work in the market and work at home. Individuals accumulate skills while working, through learning by doing, and those with higher initial skill accumulate faster. Occupations differ by their level of complexity. Occupations with high complexity have high skill requirements and high marginal products of skill. Occupations produce *occupational goods*, which are aggregated to produce a final market good. Home production uses female time and market goods as inputs. I model technological change in both market and home good production by allowing input shares to change over time. Females are predicted to accumulate less work experience and therefore have flatter skill profiles over the lifecycle than males. This affects the ratio of female to male wages in two ways. First, it implies a flatter wage profile over the lifecycle, which pushes the ratio of female to male wages down. Second, it reduces the benefits of occupations with high complexity and thus increases the ratio of female to male wages. The fraction of females in occupations with high complexity decreases and therefore the average initial skill of females relative to males increases in each occupation.

For each occupation, I measure the minimum skill requirement with the index of the occupational complexity, which I construct from the O*NET dataset. I discipline the technology for learning by doing with the lifecycle profile of male wages and the technology for home production with female labor supply over the lifecycle, for the cohorts born between 1915 and 1925. I infer the distribution of initial skill for males from their relative wages across occupations and that for females from their occupational choices, for the cohorts born between 1915 and 1925. My calibration implies that the distribution of initial skill for females first order stochastically dominates the distribution of initial skill for males. This is consistent with Rendall (2010) and Yamaguchi (2012). Finally, I calibrate technological changes in

the market and home good production to, respectively, the evolution of male occupational choices and female labor supply over successive cohorts born between 1915 and 1955.

I find that initial skill and work experience go a long way in explaining the observed patterns in the gender wage gap. Overall, the model generates the ratio of female to male wages for the cohorts born between 1915 and 1925 (the average ratio of female to male wages over the lifecycle is 0.68 both in the model and in the data), and 60 percent of the increase in the ratio of female to male wages for successive cohorts born between 1915 and 1955. More precisely, first, the model generates the U-shaped pattern of the ratio of female to male wages across occupations for young individuals as well as the successive flattening of that U-shape. Second, the model reproduces the lifecycle pattern of the gender wage gap for the cohorts born between 1915 and 1925. For these cohorts, the ratio of female to male wages decreases between age 25 and age 65 of 27 percent on average across occupations, compared to the 25 percent decrease in the data. The decrease is stronger in occupations with high skill requirements. Also, the model reproduces 50 percent of the attenuation of the decrease in the ratio of female to male wages over the lifecycle that happens for successive cohorts born between 1915 and 1955.

The gender wage gap is decided by two margins. The first margin is the gender difference in average skill within each occupation. The second margin is the gender difference in occupational choice. The second margin matters because there are differences in the price of occupational output and in the minimum skill requirement (hereafter “occupational characteristics”) across occupations. I find that the first margin explains the majority of the gender wage gap. Blau and Kahn (1997) and Card and DiNardo (2000) draw similar conclusions. Over time, technological change reinforces the narrowing of the difference in female to male wages. In the home good production, it leads to a decline in the share of time, accounting for 80 percent of the increase in the work experience of females born between 1915 and 1955. In the market good production, it increases the price of occupational output requiring high levels of skill and improves the occupational composition of female labor supplied: females move toward occupations for which the difference in wages between genders is smaller.

My model is also broadly consistent with the structure of average wages across occupations.

There is a large interest in understanding how much of what an individual earns is due to his/her skills and how much is due to the returns to skills (see for example Bowlus and Robinson, 2010, and Hendricks and Schoellman, 2011). I decompose average wages in each occupation in two components: one that depends on occupational characteristics and one that depends on individuals' skill. I find that occupational characteristics make up for about 60 percent of wages. Also, about 30 percent of the difference in wages between occupations requiring high levels of skill and occupations requiring low levels of skill is due to occupational characteristics.

The papers that are the closest to mine are Erosa, Fuster and Restuccia (2005), Rendall (2010), and Hsieh, Hurst, Jones and Klenow (2012). Erosa, Fuster and Restuccia (2005) consider work experience as a determinant of the lifecycle pattern of the gender wage gap. I extend their analysis by exploring how work experience affects wages across occupations with a different scope for learning by doing. Rendall (2010) investigates the role of brain-biased technological change in explaining the narrowing of the gender gap in wages. She considers a single sector model and therefore cannot speak about the non-linearities of the gender wage gap across occupations. Hsieh, Hurst, Jones and Klenow (2012) explain the evolution of the gender wage gap and gender differences in occupational choice using evolving education and labor market frictions. I explicitly model these frictions by endogenizing work experience through home production and learning by doing across heterogeneous occupations.

The rest of the paper is organized as follows. I next describe the two main patterns of the gender wage gap by occupational complexity that are the focus of my analysis along with major patterns of occupational choice and labor supply. Section 3 outlines the model and section 4 calibrates it. Section 5 discusses the results and section 6 concludes.

2 Facts

Building on the work of Acemoglu and Autor (2010), I construct an index describing the complexity of an occupation. Acemoglu and Autor (2010) utilize the O*NET Work Activities, Work Context, Skills and Abilities files to construct six tasks and measure their

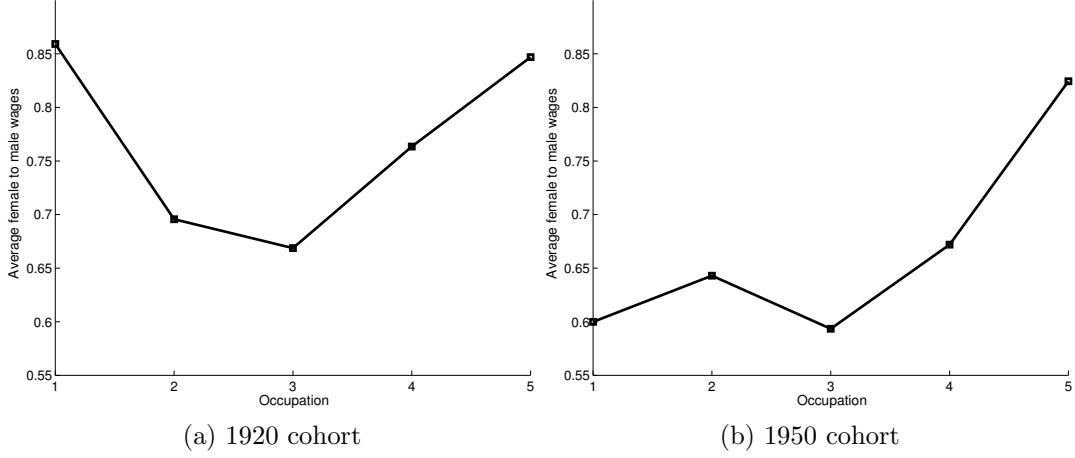


Figure 1: Gender wage gap between ages 25 and 35. Source: IPUMS-USA and O*NET.

intensity for the three digit occupational categories of the 1990 Census. The tasks are: (a) non-routine cognitive analytic, (b) non-routine interpersonal, (c) routine cognitive, (d) non-routine manual physical, (e) non-routine manual interpersonal, and (f) routine manual. I combine the six tasks measures into two aggregate measures: *Brain*, summarizing the intensity of analytical tasks, and *Brawn*, summarizing the intensity of physical tasks. *Brain* is the average of tasks measures (a), (b), and (c). *Brawn* is the average of tasks measures (d), (e), and (f). Both aggregate tasks measures are standardized to zero mean and standard deviation of one. Appendix A shows the distribution of occupations on the *Brain* and *Brawn* dimensions for the 1990 three-digit coding system of the Census. Finally, I define my index of tasks complexity for occupation i as follows:

$$o_i = \frac{e^{Brain_i}}{e^{Brawn_i}}$$

where $Brain_i$ and $Brawn_i$ are the two aggregate tasks measures for occupation i and e is the exponential operator. I order occupations according to their complexity, and group them into five groups that correspond to the five quintiles of the empirical distribution of o . The first group (hereafter “occupation 1”) contains the occupations with complexity in the first quintile of the empirical distribution of o . Occupation 1 is the group with the lowest average complexity: $o_1 = 0.54$. The fifth group (hereafter “occupation 5”) contains the occupations with complexity in the fifth quintile of the empirical distribution for o and has the highest

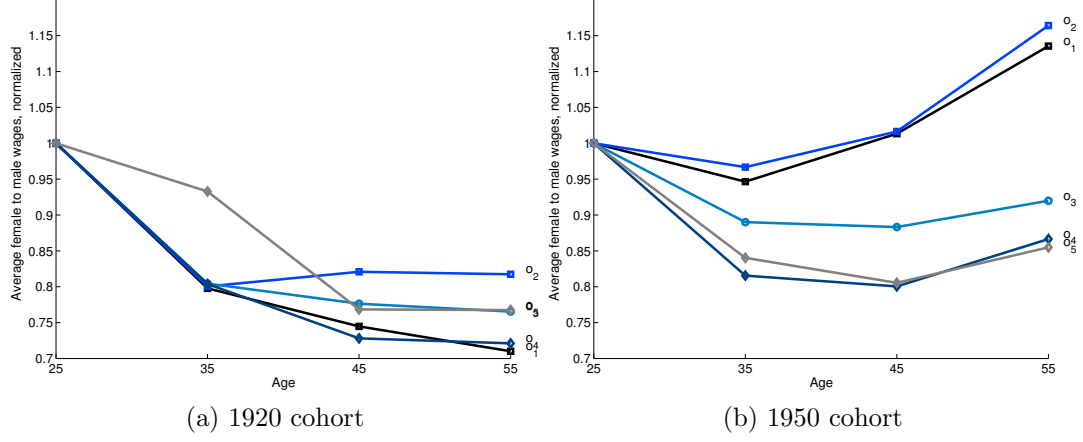


Figure 2: Gender wage gap over the lifecycle. Data for each occupational group are normalized to the value between ages 25 and 35. Source: IPUMS-USA and O*NET.

average complexity, $o_5 = 2.17$. Occupations 2 to 4 have complexity in between o_1 and o_5 : $o_2 = 0.69$, $o_3 = 1.18$, and $o_4 = 1.49$.

I use IPUMS-USA data from years 1950 to 2010 to document patterns of the gender wage gap for occupations 1 to 5 for successive cohorts of married individuals. Details on sample selection are in Appendix A.² I focus on two sets of cohorts: the cohorts born between 1915 and 1925 (hereafter “1920 cohort”) and the cohorts born between 1945 and 1955 (hereafter “1950 cohort”). Patterns of the cohorts born between 1925 and 1945 are smooth transitions from those of the 1920 cohort to those of the 1950 cohort. Figure 1 shows the gender wage gap, measured as the ratio of female to male wages, between ages 25 and 35 for the 1920 and the 1950 cohorts. The gender wage gap for young individuals is U-shaped across occupations in the 1920 cohort. The ratio of female to male wages is the highest in occupations 1 and 5, with value 0.85. The U-shape flattens across successive cohorts of young individuals: in the 1950 cohort, the ratio of female to male wages is the highest in occupation 5, with value 0.82, and the lowest in occupation 1, with value 0.6. Figure 2 documents how the gender wage

²The focus on married individuals is motivated by the fact that the model I later develop naturally maps to a household composed of at least two individuals. However, the patterns here presented are not qualitatively dependent on sample selection criteria and the index of occupational complexity chosen. In particular, the same qualitative patterns result under the three following alternative scenarios: 1) when both married and un-married individuals are considered, 2) when the index of occupational complexity is only a function of *Brain*, i.e., $o_i = e^{Brain_i}$, and 3) when the occupational complexity index is computed from the Dictionary of Occupational Titles dataset for equivalent tasks.

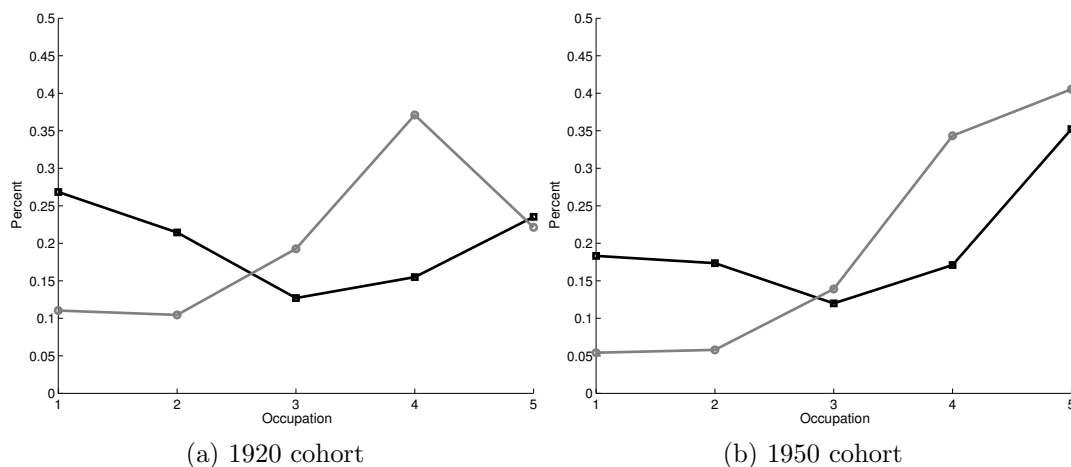


Figure 3: Occupational composition by gender. Females (light-gray line) vs. Males (black line). Source: IPUMS-USA and O*NET.

gap evolves over the lifecycle for occupations of different complexity. The ratio of female to male wages decreases over the life-cycle in both the 1920 cohort and the 1950 cohort and the decrease is on average stronger for occupations with higher complexity. In the 1920 cohort, the ratio of female to male wages decreases 24 percentage points between ages 25 and 65 in occupation 2, and 28 percentage points in occupation 4. In the 1950 cohort, the ratio of average female to male wages decreases 8 percentage points between ages 25 and 65 in occupation 2 and 15 percentage points in occupation 4 .

In addition, the data reveals three further relevant facts. First, females of the 1950 cohort supply significantly more hours to market work than those of the 1920 cohort. Among the married females of the 1920 cohort, 47 percent supply positive hours to market work between ages 45 and 55, and they supply on average 64 percent of the amount of hours supplied by males. Among the married females of the 1950 cohort, 85 percent supply positive hours to market work between ages 45 and 55, and they supply on average 75 percent of the amount of hours supplied by males. Second, males and females choose on average different occupations. As showed in Figure 3, males choose occupations with low complexity more frequently than females. Among the 1920 and the 1950 cohorts, on average 20 percent of male workers are in occupation 1, while only 8 percent of female workers are in occupation 1. Females tend to select occupations with complexity in the upper-half of the the distribution

of o , with a substantial difference across successive cohorts. Female workers of the 1920 cohort concentrate for the most in occupation 3, while female workers of the 1950 cohort concentrate for the most in occupation 4.³ Finally, females, especially those of less recent cohorts, systematically choose occupations with low wages relative to males. Define the *weighted gender wage gap* as the weighted average of the gender wage gaps across the five occupations, with weights the frequency of females in each occupation. For the 1920 cohort, the weighted gender wage gap is 15 percentage points lower than the gender wage gap; while for the 1950 cohort the weighted gender wage gap is 10 percentage lower higher than the gender wage gap.

3 Model

3.1 Setup

Time is discrete and runs from $t = 1, 2, \dots, T$. Each model period corresponds to 10 years of calendar time. The economy is populated by overlapping cohorts of individuals who live for 4 periods: individuals enter the model at age 25 and exit at age 65. I use τ to denote a cohort: cohort τ is composed of individuals of age 1 at time t . For ease of notation, I denote age by j , that is $j = t - \tau + 1$.

Within a cohort, individuals are heterogeneous with respect to their gender $g \in \{m, f\}$ and their initial skill level $s_0 \in \mathfrak{R}^+$. There is an equal mass of females and males. Skill summarizes an agent's ability to perform tasks that are valuable in the labor market. Initial skill is distributed across individuals in line with a gender-specific cumulative distribution function Γ_g . The CDF for females is a transformation of the CDF for males. Given a CDF for males Γ_m , the CDF for females is:

$$\begin{aligned} \Gamma_f(s_0^k) &= \xi^k \Gamma_m(s_0^k) \\ \text{for } s_0^k &\in \mathfrak{R}^+, \Gamma_f(s_0) \leq 1 \forall s_0. \end{aligned}$$

³Similar improvements in the occupational condition of females have been reported by others in the literature. See for example Goldin (2006).

Individual types are pairs $p = (s_0, g)$ on the set $\mathcal{P} = \mathbb{R}^+ \times \{m, f\}$. I assume that individuals observe their type before any decision is made. I assume that credit markets are complete and there is no uncertainty.

Within a generation, individuals of different gender, a husband and a wife, match to form a household. Household preferences are defined over joint consumption of market goods and home goods:

$$\sum_{j=1}^4 \beta^{j-1} [U(c_j) + U(x_j - \bar{x}_{j\tau})], \quad (1)$$

where $\beta \in (0, 1)$ is the discount factor, c_j is the family consumption of the market good at age j , $c_j = c_j(\cdot, f) + c_j(\cdot, m)$, and x_j is the family consumption of the home good at age j . There is subsistence consumption for the home good, $\bar{x}$. It represents the amount of housework and child care that allows for the satisfaction of the (physical and mental) basic needs of family life. I do not model fertility. For this reason, I allow $\bar{x}$ to be indexed by age so to accommodate for changes in the amount of housework and child care that come about as family structure changes. The utility function abstracts from leisure. I interpret an individual's time endowment as his/her *working* time endowment. Knowles (2012) finds that the fraction of time in a week allocated to leisure activities (non-working activities) remained basically unchanged for married men and women over the 1965-2003 period. In 1965, married men dedicated on average 57 hours per week to non-working activities, while in 2003 they dedicated on average 60 hours per week. In 1965, married females dedicated on average 61 hours per week to non-working activities, while in 2003 they dedicated on average 64 hours per week.

Individuals are endowed with one unit of time each period, which can be used for market production and for home production. Let ℓ denote hours supplied to market work. I assume that males supply their entire time to market work, that is $\ell(\cdot, m) = 1$.⁴ At the beginning of

⁴Age-market-hour profiles of married men are essentially identical for cohorts born between 1915 and 1955. On the contrary, the age-market-hour profiles for married females change systematically for successive cohorts during this period. In Appendix B, I decompose the variance of the logarithm of market hours and show that most part of the variance for the 1950-2000 period is due to the dispersion of market hours within the female group. Moreover, most of the change in the variance over time is due to the change in market hours dispersion within the female group. In addition, Knowles (2012) finds that females carry out most part of household chores within a married couple in 1975 and the picture looks very similar 30 years later. In 1975, males supply on average 7 hours to household chores and child care, while females supply on average

each period females must decide whether or not to join the labor market during that period. If they decide to join the labor market, they choose the amount of hours they supply to market work $\ell(\cdot, f)$. There is a minimum amount of hours $\underline{\ell}$ that a person needs to supply to market work once she joins the labor market, that is $\ell(\cdot, f) \in \{0\} \cup [\underline{\ell}, 1]$.

Home goods are produced using a combination of market goods and the wife's time. That is, $x_j = f_{j\tau}(\ell_j(\cdot, f), y_j)$ where y_j is the input of market goods in the production of the home good at age j . The production function $f_{j\tau}$ is allowed to change over the lifecycle and across cohorts. The change over the lifecycle is meant to capture changes in the relative intensity of inputs that come about as family composition changes. An important fraction of home production is related to children. As children grow older, child care services involve more market goods and less time.⁵ The change across cohorts is meant to capture changes in the technology of home production as in Greenwood, Seshadri and Yorukoglu (2005) and changes in cultural norms with regard to working females as in Fernández, Fogli and Olivetti (2004). I assume the following functional form for $f_{j\tau}$:

$$f_{j\tau}(\ell_j(s_0, f), y_j) = \begin{cases} [\varphi_{j\tau}(1 - \ell_j(\cdot, f))^\alpha + (1 - \varphi_{j\tau})(y_j)^\alpha]^{\frac{1}{\alpha}} & \text{if } x_j \leq \bar{x}_{j\tau} \\ y_j & \text{otherwise.} \end{cases} \quad (2)$$

Technological change in home production is modeled as a change in the share of the wife's time across cohorts: $\varphi_{j\tau+1} = \varphi_{j1}g_\varphi^\tau$. The production of home goods takes the wife's time and market goods as inputs up to the subsistence level. After that, any additional unit of home good is transformed one to one from the market good. I assume that the subsistence level of home production in each period can be satisfied by the wife devoting all her time to home production. That is, $\bar{x}_{j\tau} = \varphi_{j\tau}^\alpha$.

Individuals decide the occupation they want to work in. They do so at the beginning of life. Once an occupation is chosen, it cannot be changed.⁶ There is a finite number I

32 hours. In 2003, males supply on average 10 hours to household chores and child care, while females supply on average 26 hours.

⁵Olivetti (2006) makes similar assumptions for the childcare production function.

⁶Kambourov and Manovskii (2008) find evidence of substantial mobility across occupations classified according to the three-digit coding system of the Census. However, they find little mobility when classifying the occupations according to the two-digit coding system of the Census. In the quantitative analysis, I consider only 4 occupational groups.

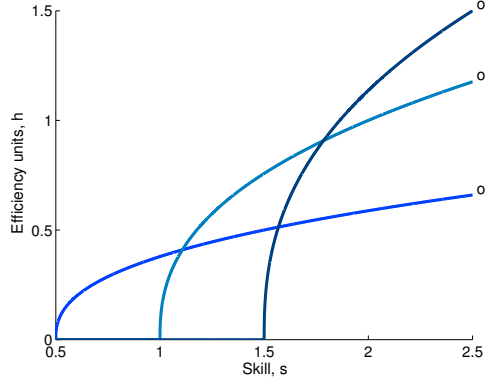


Figure 4: Skill-efficiency unit schedules.

of occupations indexed by i . Occupations differ by two features: 1) the skill required to perform, and 2) the marginal product of skill. An individual can be employed in occupation i if and only if his/her initial skill is at least o_i . o_i indicates the complexity of occupation i , takes non-negative values and increases with i . The marginal product of skill is assumed to increase with the complexity of an occupation. This reflects the idea that the value of skill in an occupation depends on the occupational complexity. If an occupation involves simple tasks, there is little room for improvement and so workers of different skill turn out to be similarly productive. For example, Newton and a monkey would be equally productive in an occupation that involves the only task of peeling a banana. However, this would not be true in an occupation consisting of assembling a computer. When occupation i is filled by an individual of skill s he/she produces $h(g, s, i)$ units of *occupational output* (“occupational output” and “efficiency unit” are used exchangeably hereafter):

$$h(g, s, i) = \begin{cases} 0 & \text{if } s < o_i \\ o_i^{1-\rho}(s - o_i)^\rho \omega_g & \text{otherwise.} \end{cases} \quad (3)$$

Figure 4 shows the skill-efficiency units schedules for three occupations. Efficiency units increase with an individual’s skill. Across occupations, output is more sensitive to individuals’ skills when occupations are more complex. I assume that $\omega_m = 1$; ω_f represents female TFP in occupational output production.

Individuals accumulate skill through learning by doing while working. The amount of skill

an individual accumulates in a period depends on the amount of skill accumulated up to the current period and the amount of hours supplied to market work. In particular,

$$s_{j+1} = s_j(1 - \delta) + \eta_j s_j^\beta \ell_j^\psi, \quad j = 0, \dots, 3, \quad (4)$$

given initial skill s_0 . Individuals with high initial skill are more efficient in learning new skills. Notice that learning on the job is not occupation-specific. However, because the marginal product of skill is occupation-specific, so is the scope of learning by doing and the value of acquired skill. In occupations with low complexity, acquired experience influences output only slightly, while the opposite is true in occupations with high complexity.

Individuals are paid wages, E , according to the marginal product:

$$E(g, s, i) = \underbrace{w_i}_{\text{price}} h(g, s, i),$$

where w_i is the price of occupational output i . Total earnings at age j are: $TE_j(g, s, i) = E(g, s, i)\ell_j(s_o, g)$. Notice that individuals who supply hours to an occupation with a requirement higher than their current skill but lower than their initial skill have zero earnings. I think of this case as that arising in the context of training programs. Individuals accept zero earnings for a period in exchange for higher future earnings that will follow from the newly acquired skills.

3.2 Individual Problem

Households maximize lifetime utility in eq. 1. They do so by choosing the occupation of the husband, that of the wife, the wife's hours of market work and family consumption of both the market good and the home good.

The setup of the model implies that occupational and time allocation decisions are independent from consumption decisions. Present value income maximization implies utility maximization. Husband and wife choose their occupation and hours of market work to maximize their discounted lifetime earnings net of the cost of market goods used in the

production of the subsistence requirement of the home good. Then, given the discounted lifetime earnings of the family, they decide how to split their resources between market good consumption and discretionary consumption of the home good. Occupational and time allocation decisions are the only ones that matter for the determination of wages and therefore for the study of the gender wage gap. In the rest of the paper I will disregard consumption decisions. The occupational and time allocation decisions of the husband are independent from those of the wife, and vice-versa. This allows for the husband's and the wife's problem to be analyzed separately.⁷

Males. Males decide their occupation. They do so to maximize the present value of earnings over their lifetime. Let $V_j(s, m; i, \tau)$ denote the present value of future earnings at age j of a male individual of cohort τ and skill s who is employed in occupation i . $V_j(s, m; i, \tau)$ is:

$$V_j(s, m; i, \tau) = \sum_{u=j}^4 \left(\frac{1}{R} \right)^{u-1} w_{it} h_u(m, s_u, i), \quad (5)$$

for $t = \tau + u - 1$.

given eqs. 3 and 4. R is the gross interest rate, which is the reciprocal of the discount factor. The value of being employed in occupation i for a male individual of type $(m, s_0) \in \mathcal{P}$ in cohort τ is $V_1(s_0, m; i, \tau)$.

Females. Females decide their occupation and each period's amount of market hours. They do so to maximize the present value of earnings over their lifetime net of the cost of market goods used to satisfy the subsistence requirement. Let $V_j(s, f; i, \tau)$ denote the present value of future earnings at age j of a female individual of cohort τ and initial skill s that is employed

⁷With respect to occupational and time allocation decisions, my setup is equivalent to constraining the household to produce a level of the home good equal to the subsistence level and take the consumption of home good out of the utility function. Erosa, Fuster and Restuccia (2005) find that once conditioned by the number of children, marriage is not crucial for understanding female labor supply. The independence between the decisions of husband and wife implies that the conclusions drawn in this paper for the gender wage gap do not depend on how mating in the marriage market is modeled.

in occupation i . $V_j(s, f; i, \tau)$ is:

$$\begin{aligned}
V_j(s, f; i, \tau) &= \max_{\{\ell_u^i, y_u^i\}_{u=j}^4} \sum_{u=j}^4 \left(\frac{1}{R} \right)^{u-1} [w_{it} h_u(f, s_u, i) \ell_u(s, f; i, \tau) - y_u(s, f; i, \tau)], \\
\text{for } y_u(s, f; i, \tau) &= \left(\frac{\bar{x}_{u\tau} - \varphi_{u\tau} (1 - \ell_u(s, f; i, \tau))^\alpha}{1 - \varphi_{u\tau}} \right)^\alpha, \\
y_u(s, f; i, \tau) &\geq 0, \\
\ell_u(s, f; i, \tau) &\in \{0\} \cup [l, 1], \\
t &= \tau + u - 1.
\end{aligned} \tag{6}$$

subject to eqs. 3 and 4. $l_j(s, f; i, \tau)$ and $y_j(s, f; i, \tau)$ are the policy functions for market hours and for market goods used in the production of the subsistence requirement of the home good for a female individual employed in occupation i , at age j . The benefits of supplying hours to home production are saved expenses on market goods. Substituting market hours with home hours entails two costs. First, there is a direct cost due to the foregone earnings of the period ($wh\ell$). Second, learning-by-doing implies the additional cost of the lost future earnings that would have resulted from the additional skills accumulated through learning by doing while working in the current period. The benefits of increasing home hours do not depend on an individual's skill, while the costs of increasing home hours increase with an individual's skill. Therefore, females with high skill supply more time to market work than females with low skill. There is a threshold level of initial skill below which females do not join the labor market, i.e., $\ell(\cdot, f) = 0$. The value of being employed in occupation i for a female individual of type $(f, s_0) \in \mathcal{P}$ in cohort τ is $V_1(s_0, f; i, \tau)$.

Occupation Choice. Individuals of cohort τ choose their occupation based on their initial skill s_0 and their gender g . The occupational problem is:

$$\max_i V_1(s_0, g; i, \tau),$$

for $(s_0, g) \in \mathcal{P}$ and $V_1(s_0, g; i, \tau)$ defined as in 6 and 5. The decision rule stemming from this

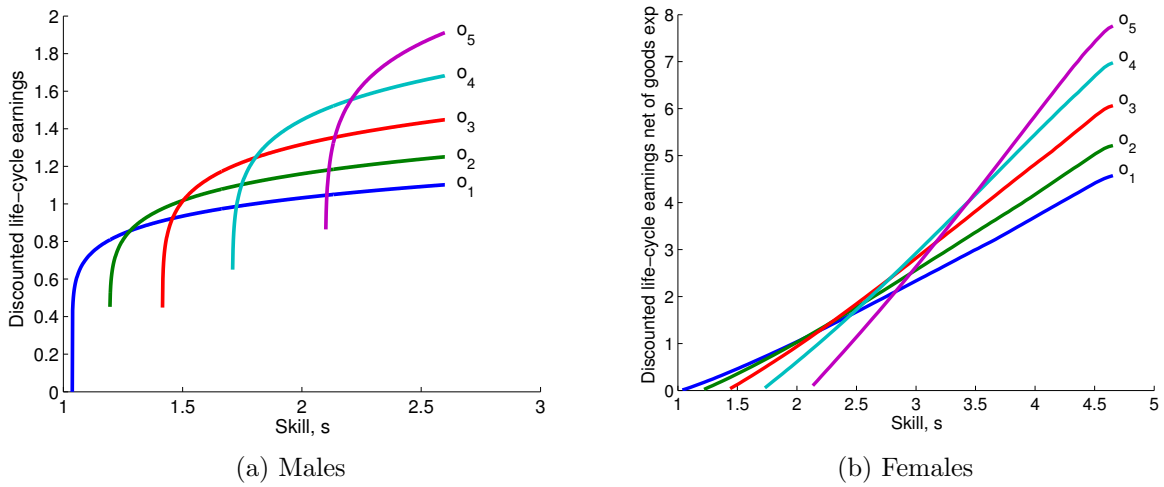


Figure 5: Occupational choice in the model.

problem can be characterized by a simple threshold rule:⁸

$$\mathbf{1}(s_0, g; i, \tau) = \begin{cases} 1, & \text{if } s_0 \in [\underline{s}_0^i(g; \tau), \bar{s}_0^i(g; \tau)] \\ 0 & \text{otherwise.} \end{cases}$$

Figure 5 characterizes the occupational problem for males and for females. As in Roy (1951) individuals sort across occupations based on their comparative advantage, which is determined by a combination of the individual's skill and occupational complexity. The model produces perfect positive sorting of individuals by initial skill across occupations of increasing complexity. Individuals with high initial skill choose occupations with higher complexity than those chosen by individuals with low initial skill. The advantage of occupations with high complexity over those with low complexity is that an increase in skill translates into a bigger increase in efficiency units and so a bigger wage increase. Individuals with high initial skill have more use of this advantage since they are more productive learners and therefore have steeper lifecycle skill profiles than individuals with little initial skill. The discounted value of lifecycle earnings for females are more convex on the skill domain than those for

⁸It can be shown that the value functions for the different occupational options have the single crossing property for the case of males. For the case of females, the single crossing property holds for a region s_0 close to the equilibrium in which there are females working in each occupation. This is the case in the data. The single crossing property is violated for some values of s_0 to the left of $s_0^1(f)$ at which it is not optimal for females to join the labor market and the value functions for all occupational options equal zero.

males, for all occupations. Moreover, those for females intersect at higher values of skill respect to those for males skill, i.e., $\underline{s}_0^i(f; \tau) \geq \underline{s}_0^i(m; \tau)$ and $\bar{s}_0^i(f; \tau) \geq \bar{s}_0^i(m; \tau)$. Females supply less time to market work than males and therefore have flatter lifecycle skill profiles for the same initial skill. Notice that, as female experience decreases, so does the fraction of females in occupations with high complexity and, for the same distribution of initial skill, the average initial skill of females increases relative to that of males in each occupation.

3.3 Equilibrium

Market good production. There is one homogeneous final market good Y . This is produced by a representative firm that combines the aggregate occupational outputs:

$$Y_t = A_t \left(\sum_j a_{it} H_{it}^\theta \right)^{\frac{1}{\theta}},$$

where H_{it} is the aggregate output of occupation i at time t . A_t is general TFP at time t and a_{it} is the productivity of output i at time t . I assume that both general technological progress and occupation-specific technological progress occur at constant rates g and g_i , respectively. The representative firm maximizes profits as follows:

$$\max_{\{H_{it}\}_{i=1}^I} Y_t - \sum_{i=1}^I w_{it} H_{it}.$$

The solution of the firm problem implies that prices for each occupational output equal their marginal products at each point in time, $w_{it} = \frac{\partial Y_t}{\partial H_{it}}$

Aggregation. The aggregate output of occupation i at time t is:

$$H_{it}^S = \sum_{j=1}^4 \sum_{g \in \{m, f\}} \int_{s_0} \mathbf{1}(s_0, g; i, \tau) \ell_j(s_j(s_0, g; i, \tau), g; i, \tau) h(g, s_j(s_0, g; i, \tau), i) d\Gamma_g(s_0).$$

for $\tau = t - j + 1$. Starting from the right, this equation sums total occupational output i produced by males and females of the four age groups considered in the model, for the

cohorts alive at time t . Total consumption of market goods at time t is:

$$C_t = \sum_{i=1}^I \sum_{j=1}^4 \sum_{g \in \{m,f\}} \int_{s_0} c_j(s_j(s_0, g; i, t-j+1), g; i, t-j+1) d\Gamma_g(s_0),$$

where $c_j(s_j, g; i, \tau)$ is the policy function for market good consumption for an individual of cohort τ , gender g , age i and skill s_j . Starting from the right, this equation sums total market good consumption of males and females of the four age groups considered in the model, employed in each of the I occupations, for the cohorts alive at time t . Lastly, total expenditures on market goods used in the production of the subsistence level of home goods and total expenditure on home goods on top of the subsistence requirement at time t , aggregate into:

$$Y_t^D = \sum_{i=1}^I \sum_{j=1}^4 \sum_{g \in \{m,f\}} \int_{s_0} (x_j(s_j(s_0, g; i, \tau), g; i, \tau) - \bar{x}_{j\tau}) + y_j(s_j(s_0, g; i, \tau), g; i, \tau) d\Gamma_g(s_0).$$

for $\tau = t - j + 1$. $x_j(s_j, g; i, \tau)$ is the policy function for home good consumption for an individual of cohort τ , gender g , age i and with skill s_j . The summation follows the same intuition as the previous one.

Equilibrium. Given R , a competitive equilibrium consists of (1) allocations for individuals of each type: $\{\{\ell_j(s_j, g; i, \tau), c_j(s_j, g; i, \tau), x_j(s_j, g; i, \tau), y_j(s_j, g; i, \tau)\}_{j=1}^4, \mathbf{1}(p; i, \tau)\}_{i=1}^I\}_{\tau=1}^\infty$ for $p \in \mathcal{P}$, and allocations for the firm $\{\{H_{it}\}_{i=1}^I\}_{t=1}^\infty$; (2) prices $\{\{w_{it}\}_{i=1}^I\}_{t=1}^\infty$; Such that:

1. The allocations of the individuals solve the optimization problem of each individual $(g, s_0) \in \mathcal{P}$ given prices;
2. The allocations of the firm solve the firm's optimization problem given prices;
3. The price of the output of each occupation, w_{it} , clears the labor market for each occupation at each point in time, that is $H_{it} = H_{it}^S$;
4. The aggregate resource constraint holds: $Y_t = C_t + Y_t^D$.

4 Calibration

The calibration strategy is as follows. (i) Use the O*NET dataset to characterize the skill-efficiency unit profiles across occupations. (ii) Use the 1950-2010 IPUMS-USA data on male wages and female market hours to calibrate the learning on the job and the home production functions. (iii) Use the 1950-2010 IPUMS-USA dataset combined with the O*NET dataset to compute occupational choices of males and females and discipline the productivities of occupational outputs and the distribution of initial skill.

I assign values to some parameters using a-priori information, these are shown in Table 1. A model period corresponds to 10 years of calendar time: individuals enter the model at age 25 and exit at age 65. The number of occupations is set to 4. I set the occupational requirements o_i for each of the four occupations to the complexity index constructed in section 2. I merge occupations 1 and 2 in a single occupation since they have similar complexity indexes. I set the gross interest rate R to 1.4802 (decennial rate) and the lower bound for market hours to 0.2 to be consistent with the criteria used in sample selection. I set the depreciation rate $\delta - 1$ to 0. Rupert and Zanella (2010) show that the decrease in wages toward the end of the life cycle is marginal. The productivity of occupational output 1 at time $t = 1$ is normalized to 1. I let general TFP grow at 0.14 percent per year. There is little information on the elasticity of substitution across occupational outputs, θ . I avoid the perfect substitution case following Firpo, Fortin and Lemieux (2011) and set θ to $2/3$ as in Hsieh, Jones, Hurst and Klenow (2012)'s baseline experiment. I assume that the distribution of initial skill for males, Γ_m is uniform. This class of distributions is characterized by two parameters, $\Gamma_m(\underline{s}_0, \bar{s}_0)$. I normalize the lower bound of the distribution, $\underline{s}_0$, to a number greater than the skill requirement of occupation 1 and calibrate the upper bound, $\bar{s}_0$, within the model.

The list of remaining parameters is:

$$\begin{aligned}\Lambda_1 &= (\rho, \beta, \{\eta_j\}_{j=1}^4, \bar{s}_0, \{a_{i1}, g_i\}_{i=2}^4), \\ \Lambda_2 &= (\psi, \alpha, \{\varphi_{j1}\}_{j=1}^4, \omega_f, \{s_0^k, \xi^k\}_{k=1}^3, A_1, g_\varphi).\end{aligned}$$

Λ_1 is calibrated to data on male wages and male occupational choice. Λ_2 is calibrated to

PARAMETER	SYMBOL	VALUE
Model period		10 years
Number of occupations	ℓ	4
Gross interest rate	R	1.4802
Min. hours	$\underline{\ell}$	0.2
Learning by doing, depreciation	δ	0
Market good prod., TFP growth	g	0.14%
Market good prod., elasticity of subst.	θ	2/3
Occupational requirements	o_i	brain/brawn index
NORMALIZATIONS:		
Market good prod., share occupational output 1 at $t = 1$	a_{11}	1
Lower bound males and females	$\underline{s}_0$	2.3923

Table 1: Calibration, parameters computed without solving the model.

data on female wages, female working hours, and female occupational choice. I target the following moments:

1. Distribution of workers across the four occupations, for the 1920 cohort, measured between ages 45 and 55 – Gender: males and females;
2. Distribution of workers across the four occupations, for the 1950 cohort, measured between ages 45 and 55 – Gender: males;
3. Ratio of average wages in occupation 4 to that in occupation 1, for the 1920 cohort, measured between ages 25 and 35 – Gender: males and females;
4. Lifecycle profile of average wages in occupation 1 normalized to average wages in occupation 1 at ages 25 to 35, for the 1920 cohort, measured between ages 35 and 45, ages 45 and 55, and ages 55 and 65 – Gender: males;
5. Coefficient of variation of wages, for the 1920 cohort, measured between ages 45 and 55 – Gender: males;
6. Growth of the coefficient of variation of wages in occupation 1 from ages 25 to 55, for the 1920 cohort – Gender: males and females;
7. Average female to male wages, for the 1920 cohort, measured between ages 25 and 35;

8. Lifecycle profile of average female to male work hours, for the 1920 cohorts, measured between ages 35 and 45, ages 45 and 55, and ages 55 and 65;
9. Average female to male work hours, for the 1950 cohort, measured between age 25 and age 65;
10. Coefficient of variation of female to male work hours, for the 1920 cohort, measured between age 25 and age 65;
11. Ratio of number of female workers relative to that of male workers, for the 1920 cohort, measured between age 25 and age 65.

Even though the parameter values are chosen simultaneously to match the data targets, each parameter has a first-order effect on some targets. The production function of occupational output is characterized by two parameters: ρ and ω_f . ρ is important in matching cross-sectional earnings inequality (target 5). Within an occupation, the elasticity of wages to skill is:

$$\xi_{E,s} = \rho \frac{s}{s - o}.$$

Given the occupational requirements, as ρ increases, the distribution of skill maps into a more and more disperse distribution of wages. Female TFP in the production of occupational output is important in matching the overall gender wage gap for young workers (target 7). Because ω_f is not occupation-specific, it shifts the average gender wage gap leaving the pattern across occupations unaltered.

The parameters describing learning on the job, namely β, η_j, γ , are important in matching lifecycle wage profiles. In the model, an individual's wage evolves over the lifecycle as his/her skill grows:

$$\frac{\partial E_t}{\partial s_{t-1}} = \underbrace{w_i o^{1-\rho} \rho (s_t - o_i)^{\rho-1} \omega_g}_{\text{return per unit of skill}} \underbrace{\frac{\partial s_t}{\partial s_{t-1}}}_{\text{learning by doing}},$$

which implies

$$g_E = g_{wi} + \rho \frac{s}{s - o_i} g_s.$$

g_E is the growth rate of the wage, g_{wi} is the growth rate of the price of occupational output

PARAMETER	SYMBOL	Moment	TARGET		Value	
			Model	Data		
Market good production, shares at $t = 1$	$a_{i,1}$	Occ. comp males, 1920 chrt				
Market good production, shares growth	g_i	Occ. comp males, 1950 chrt				
Distribution of s_o , males	$\bar{s}_0$	$E(m, \cdot, 4)/E(m, \cdot, 1)$	1.162	1.113	3.642	
Occupational output elasticity, s	ρ	$cv(E(m, \cdot, \cdot), \text{ages 45-55})$	0.260	0.265	0.7	
Learning by doing, TFP	η_j	LC earn. growth				
Learning by doing, skill elasticity	β	growth, $cv(E(m, \cdot, 1), \text{age 25-55})$	1.145	1.135	1.4	
Learning by doing, hours elasticity	ψ	growth, $cv(E(f, \cdot, 1), \text{age 25-55})$	1.565	1.631	0.943	
Occupational output, female TFP	ω_f	Avg. gwg, age 25-35	0.754	0.795	0.65	
Home good prod., elasticity of substit.	α	Hours ineq.	0.113	0.084	0.45	
Home good prod., shares at $\tau = 1$	φ_{j1}	LC avg. hours, 1920 chrt				
Home good prod., shares growth	g_φ	Avg. hours, 1950 chrt	0.746	0.778	0.924	
Distribution of s_o , females	ξ^k	Occ. comp females				
		$E(f, \cdot, 4)/E(f, \cdot, 1)$	1.210	1.270	5.962	
Market good production TFP at $t = 1$	A_1	Female LF part.	0.473	0.473	0.392	

Table 2: Calibration, parameters computed by solving the model. The missing values are shown in Figures 6 and 7. The calibration matches the fraction of males and females in each occupation to the second decimal.

i , and g_s is the growth rate of skill. Higher η_j implies higher wage growth between ages $j - 1$ and j for all males. The η_j 's are calibrated to the lifecycle profile of average male wages in occupation 1 (target 4). The lifecycle pattern of wage inequality for males (target 6, males) disciplines β . Wage inequality among males decreases over the lifecycle if β is lower than one, while it increases if β is greater than one. Lastly, γ is important in matching the lifecycle pattern of wage inequality for females (target 6, females). Because females decide the amount of time they spend on the market, γ influences g_s and hence it influences the lifecycle pattern of female wage inequality.

The parameters describing home production, namely $\varphi_{j1}, g_\varphi, \alpha$, are important in matching female market hours. In the model, the cost of supplying hours to market work is the amount of additional market goods that needs to be bought to satisfy the subsistence requirement of the home good. This cost varies with market hours supply as follows:

$$\frac{\partial y}{\partial \ell} = \frac{\varphi}{1 - \varphi} \left(\frac{1}{1 - \ell} \right)^{1-\alpha} \left(\frac{\bar{x}^\alpha - \varphi(1 - \ell)}{1 - \varphi} \right)^{\frac{1}{\alpha}-1}.$$

When $\alpha < 1$, the marginal cost of working in the market depends on the amount of hours worked in the market and it changes with it depending on the degree of complementarity

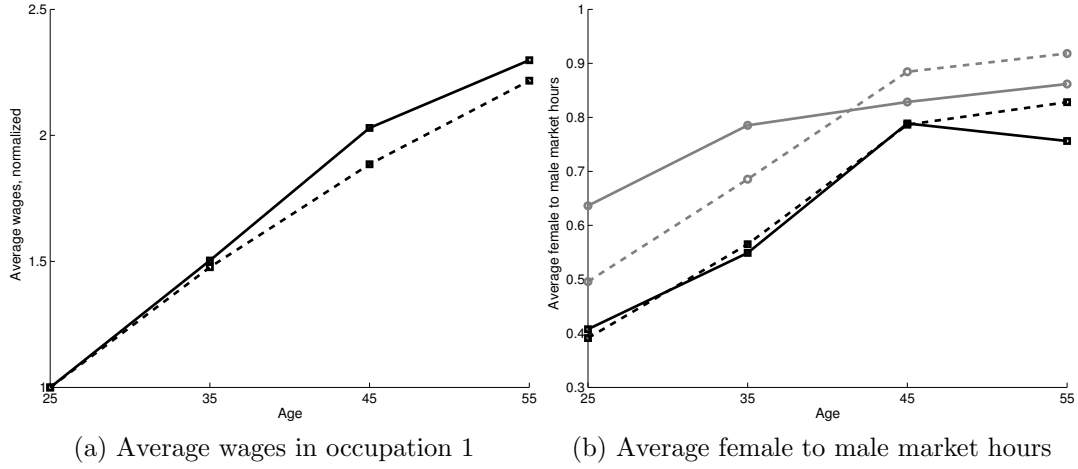


Figure 6: Model fit. Lifecycle moments for the 1920 cohort. Data (dashed lines) vs. Model (solid lines). Source: IPUMS-USA, O*NET and the author.

between hours and market goods. Because the benefits of working in the market increase with the individual's skill and there is positive sorting of individuals across occupations, α determines the dispersion of market hours across occupations. For an example, with the baseline parameterization, increasing α from 0.25 to 0.35 decreases the difference of female market hours between occupation 1 and occupation 4 by 16 percent. I target the dispersion of the *ratio* of female to male work hours, since the time endowment in the model maps to work hours of full time workers (target 10). I set the lifecycle profile of φ , i.e., φ_{j1} and g_φ , to match the lifecycle profile of average hours worked by females relative to those worked by males (targets 8 and 9). As φ increases, the cost of supplying hours to market work increases and therefore female market hours decrease.

The production function for market goods is parameterized by a set of values for the productivity parameters, A, a_{i1}, g_i . General TFP is important in matching female labor force participation (target 11). I discipline the productivities of aggregate occupational outputs with data on male occupational choice (targets 1, for males, and 2). A variation in a_i changes the slope of the profile of lifecycle discounted wages of occupation i on the initial skill domain. When the a_i 's change non-proportionally, the points of intersections on the initial skill domain of the profiles of discounted lifecycle wages for the four occupations change. Hence, the fraction of males in each occupation changes.

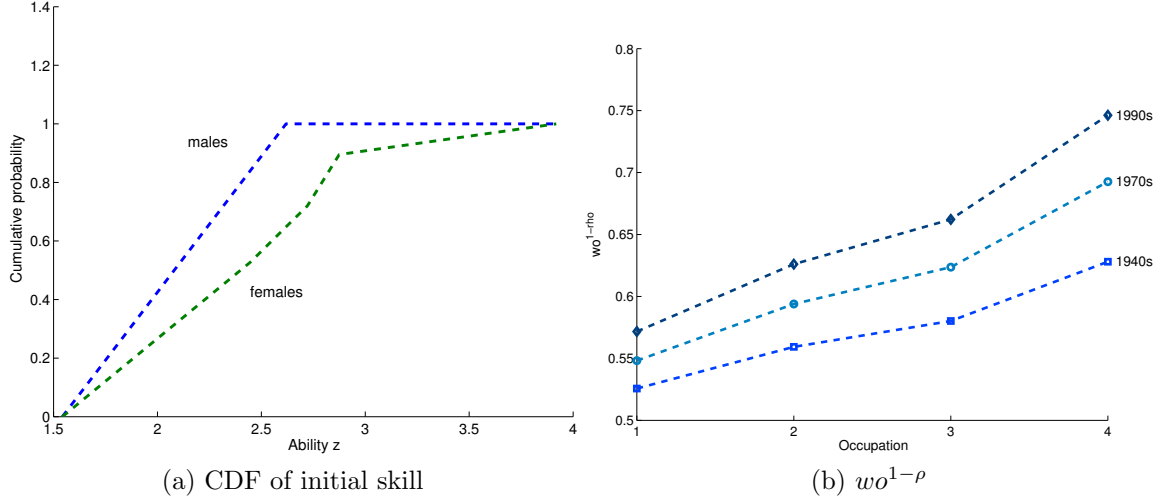


Figure 7: Calibration, implications.

The distribution of initial skill for males is parameterized by a value for $\bar{s}_0$. I discipline $\bar{s}_0$ with average wages of male workers employed in occupation 4 relative to those employed in occupation 1 (target 3 for males). The ratio of average wages of individuals in occupation i relative to those in occupation i' , at age j and time t is:

$$OP_{jt}(i, i', g) = \frac{w_{it}}{w_{i't}} \left(\frac{o_i}{o_{i'}} \right)^{1-\rho} \left(\frac{\int_{s_0} \mathbf{1}(s_0, g; i, t-j+1)(s_j - o_i)^\rho d\Gamma_g(s_0)}{\int_{s_0} \mathbf{1}(s_0, g; i', t-j+1)(s_j - o_{i'})^\rho d\Gamma_g(s_0)} \right).$$

Given the occupational requirements and ρ , $\bar{s}_0$ determines the conditional mean of initial skill in occupation 4 relative to that in other occupations. Hence, it determines average wages in occupation 4 relative to those in other occupations. For a given distribution of initial skill for males, the distribution of initial skill for females is characterized by the vector of parameters s_0^k, ξ^k . I pick four points on the s_0 domain, i.e., $k = 4$. Three points are the points of intersection on the initial skill domain of the occupational profiles of discounted lifecycle wages for females. The fourth point is chosen to match the ratio of average female wages in occupation 4 relative to average female wages in occupation 1. Finally, the ξ^k 's are set to match the fraction of females in each occupation (target 1 for females).⁹

Formally, the calibration strategy consists of solving a system of equations. For a given

⁹Schoellman (2012) uses a similar strategy for estimating immigrants skills compared to natives.

$\Lambda = (\Lambda_1, \Lambda_2)$, I compute the model moments, $X(\Lambda)$, that correspond to the targets described above. I then solve for the zero of the function $F(\Lambda)$ defined by

$$F(\Lambda) = \tilde{X} - X(\Lambda),$$

where $\tilde{X}$ are the targets described above.¹⁰ Table 2 and Figure 6 display the calibration performance. Table 2 and Figure 7 display the calibrated parameters. The distribution of initial skill for females first order stochastically dominates the one for males. From the 1940s to the 1990s, the price of output of occupations with high complexity increases relative to that of occupations with low complexity. Given the calibrated distributions of initial endowments, the evolution of prices over time favors the “working” condition of females relative to that of males. This results is consistent with the findings of Black and Spitz-Oener (2010) for the case of Germany.¹¹

5 Results

The main quantitative implication of the model is in terms of the gender wage gap. I present the model implications on the gender wage gap for the 1920 and the 1950 cohorts. I then analyze the structure of wages across occupations, separately for males and for females.

¹⁰Note that the maximizer in eq. 6 is not always concave. The policy function for female market hours is computed by using the grid search method. I compute aggregate statistics using numerical integration.

¹¹Blau and Kahn (2000) argue that technological change is unfavorable to women in the 1980s since it appears to favor individuals with high education and work experience. In my model, female work experience is endogenous. Technological change favors females because it increases the price of output of occupations that require skill at levels for which the CDF of females is more dense than the one of males.

5.1 Gender Wage Gap

The gender wage gap is the ratio of average female to male wages. Measured for individuals of age j , in occupation i at time t , it reads:

$$GWG_{jt}(i) = \omega_f \left(\frac{\int_{s_0} \mathbf{1}(s_0, g; i, t - j + 1)(s_i - o_i)^\rho d\Gamma_f(s_0)}{\int_{s_0} \mathbf{1}(s_0, g; i, t - j + 1)(s_i - o_{i'})^\rho d\Gamma_m(s_0)} \right).$$

The gender wage gap in an occupation is determined by female TFP in occupational output production, ω_f , and the ratio of average skill of females relative to that of males in that occupation. Because ω_f is constant across occupations, the pattern of the gender wage gap across occupations is determined only by the pattern of the ratio of average skill across occupations, which is in turn determined by individuals' initial skill and experience. The overall gender wage gap is decided by three margins. The first two margins are the female TFP in occupational output production and the gender skill differential. The third margin is gender differences in occupational choice. This third margin matters when occupations differ by the price of occupational output and the minimum skill requirement (hereafter “occupational characteristics”). That is, when $w_i o_i^{1-\rho}$ is not the same occupations.

The patterns of the gender wage gap for the 1920 cohort are summarized in Figure 8. Figure 8a shows the lifecycle pattern of the overall gender wage gap. The model generates the main features of the lifecycle gender wage gap in the data. The ratio of female to male wages at age 25 to 65 decreases of 14 percent (11 points) in the model and of 22 percent (17 points) in the data. The decrease in the model is due to the lower work experience of females compared to that of males. The pattern of the gender wage gap across occupations is shown in Figure 8b for ages 25 to 35. The model replicates the U-shape in the data: the ratio of female to male wages is the highest in occupations 1 and 4. The U-shape is determined primarily by gender differences in the distribution of initial skill combined with gender differences in occupational choices. At a young age, skill is mainly a result of its initial level, s_0 . The points of intersection on the initial skill domain of the occupational profiles of discounted lifecycle wages for females are right shifts of those for males. These shifts are of similar magnitude, which causes the gender wage gaps in occupations 2 and 3

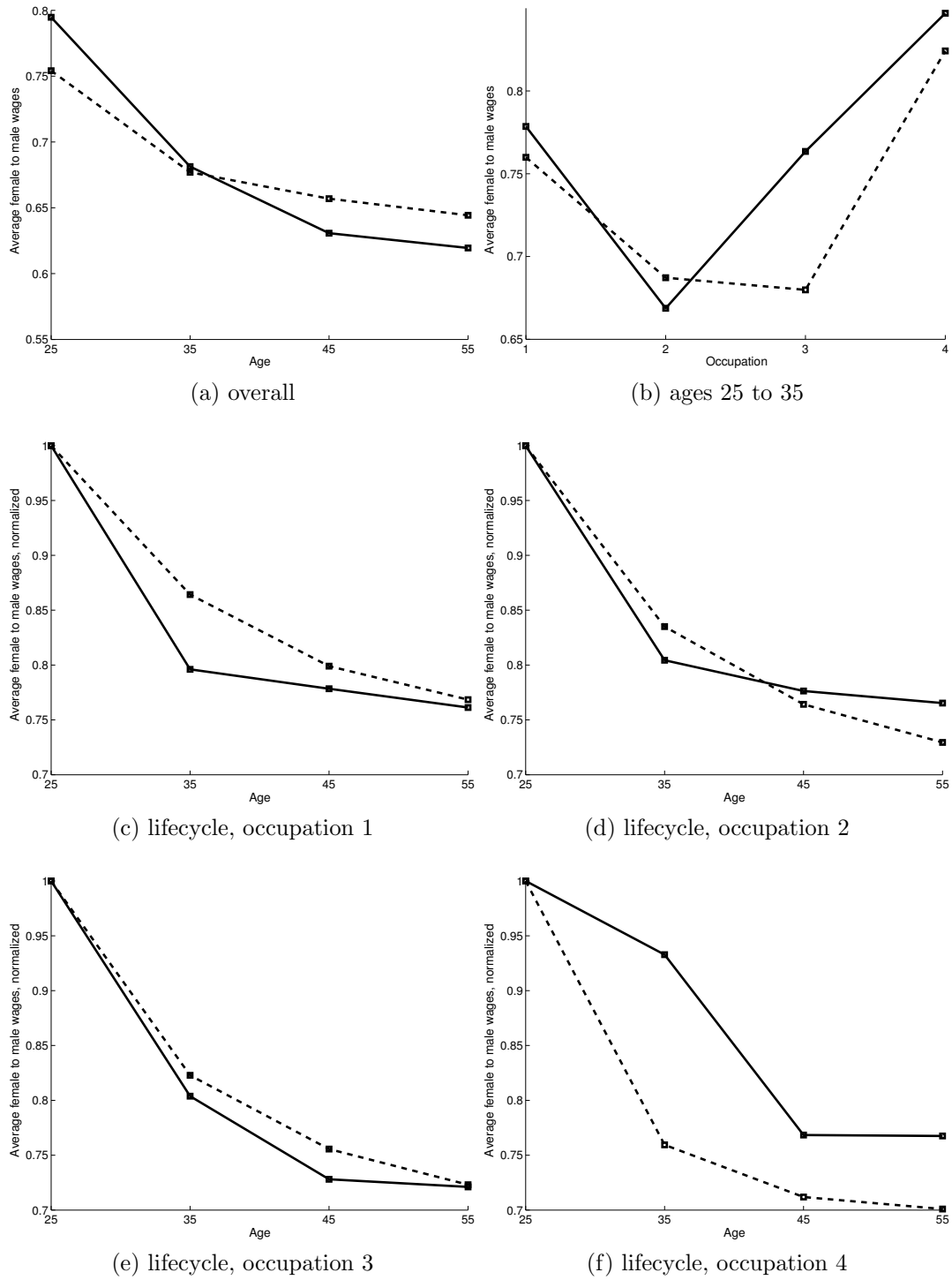


Figure 8: Results. Gender wage gap, 1920 cohort. Data (dashed lines) vs. Model (solid lines). Source: IPUMS-USA, O*NET and the author.

OCCUPATION	MALES		FEMALES	
	$\Delta_{\log(o)}$	$\Delta_{\log(s)}$	$\Delta_{\log(o)}$	$\Delta_{\log(s)}$
1	62.19	38.85	65.01	37.07
2	62.15	39.14	65.50	36.85
3	62.18	39.18	65.54	36.87
4	62.46	38.98	64.76	37.65

Table 3: Decomposition of wages, 1920 cohort. Values in the table do not sum up to one across the rows because they reported as averages over the life cycle.

to be of similar magnitude. The ratio of female to male wages in occupation 1 is pushed up relative to that in other occupations because female labor force participation is low in the 1920 cohort. Finally, the flattening out of the slope of Γ_f for high values of s_0 , increases the ratio of female to male wages in occupation 4 relative to that in other occupations.

Rows 2 and 3 of Figure 8, show the lifecycle properties of the gender wage gap by occupation. On average, the model matches the magnitude of the lifecycle decrease. The model generates a 26.9 percent decrease in average female to male wages at ages 25 to 65, compared to a 24.6 percent decrease in the data. This follows from the compound of two forces that work in opposite directions but have a common denominator: females supply less hours to market work than males. First, for a given initial skill, skill increases less over the lifecycle for females than it does for males since learning on the job increases with market hours. This effect pushes the ratio of female to male wages to decrease over the lifecycle. Second, the initial-skill thresholds for joining each occupation are higher for female than they are for males and so is average initial skill in each occupation. This composition effect pushes average female to male wages to increase over the lifecycle since learning on the job increases with initial skill. Quantitatively, the first effect is the strongest. The lifecycle decrease in the ratio of female to male wages in the model is stronger for more complex occupations. This is because the skill-efficiency profiles of more complex occupations are steeper and because individuals' with high skill are more productive in learning on the job and there is positive sorting of individuals by initial skill across occupations of increasing complexity.

As a first step in the analysis of the determinants of the gender wage gap, I study the determinants of wages. Table 3 decomposes log wages across occupations in a component

GAP/AGE	25-35	35-45	45-55	55-65	AVERAGE
E	0.7541	0.6770	0.6569	0.6443	0.6831
Δ_o	1.0230	1.0253	1.0275	1.0298	1.0264
Δ_s	0.7357	0.6603	0.6395	0.6259	0.6654

Table 4: Decomposition of the gender wage gap, 1920 cohort.

that depends on occupational characteristics ($\Delta_{\log(o)}$) and in a component that depends on individual characteristics ($\Delta_{\log(s)}$), as follows:

$$\log(E(\cdot, i)) = \underbrace{\log(w_i) + (1 - \rho) \log(o_i)}_{\Delta_{\log(o)}} + \underbrace{\log(\omega_g) + \rho \left(\log \left(\int_{s_0} \mathbf{1}(\cdot, i)(s - o_i)^\rho d\Gamma_g(s_0) \right) \right)}_{\Delta_{\log(s)}}.$$

The component that depends on occupational characteristics makes up for about 2/3 of average wages of both males and females. How much of the overall gender wage gap is due to occupational characteristics? Table 4 shows the ratio of female to male averages for three variables: wages (E), occupational characteristics (Δ_o), and individual characteristics (Δ_s). The aggregate gender wage gap is primarily determined by the Δ_s component, i.e., gender skill differential. The ratio of averages for Δ_o is greater than 1 because there are relatively more females in complex occupations than there are males, and complex occupations have higher values for the occupational characteristics (see Figure 7). On the other hand, the ratio of averages for Δ_s is lower than one because females supply hours to home production and therefore have lower experience than males have.

An individual's skill at a point in time depends on his/her initial level of skill along with his/her work experience. How much of the lifecycle gender wage gap is due to work experience? To answer this question, I run a counterfactual experiment in which I decrease the share of female time in home production, φ , to equalize female market hours to male market hour. Table 5 reports the gender wage gap over the lifecycle and across occupations for the baseline experiment (first row of each entry) and for the counterfactual experiment (second row of each entry).¹² Work experience influences the gender wage gap in two ways. First,

¹²I run an alternative counterfactual experiment in which I change the level of TFP in market good production, A , to equalize female market hours to male market hour. The results of this counterfactual experiment are very similar to the results for the counterfactual experiment in which I change φ .

OCCUPATION/AGE	25-35	35-45	45-55	55-65	AVERAGE
1	0.7599	0.6719	0.6400	0.6197	0.6729
	0.6696	0.6618	0.6535	0.6481	0.6582
2	0.6872	0.5969	0.5640	0.5417	0.5974
	0.6784	0.6710	0.6641	0.6594	0.6682
3	0.6798	0.5906	0.5600	0.5388	0.5923
	0.6785	0.6702	0.6629	0.6579	0.6674
4	0.8242	0.7169	0.7000	0.6931	0.7336
	0.7905	0.7664	0.7686	0.7769	0.7756
AVERAGE	0.7541	0.6770	0.6569	0.6443	0.6830
	0.7474	0.7744	0.7916	0.8076	0.7802
Gap Δ_o	1.0230	1.0253	1.0275	1.0298	1.0264
	1.0379	1.0421	1.0463	1.0504	1.0442

Table 5: Counterfactual exercise: the importance of experience.

as female market hours increase, females also choose complex occupations more often. This causes the average initial skill of females in each occupation to decrease and pushes down the ratio of female to male wages in each occupation. This composition effect decreases the ratio of average female to male wages between ages 25 and 35 from 0.82 in the baseline to 0.79 in the counterfactual exercise. Across occupations, the decrease is the strongest in occupation 1. The average initial skill of females in occupation 1 decreases not only because females with high skill move to occupation 2, but also because females with low skill join the labor market. Second, as female market hours increase, the lifecycle decrease of the ratio of female to male wages disappears. Females employed in occupations with high complexity gain the most from this second effect. The compound of these two effects causes 1) the ratio of female to male wages to decrease in occupations with low complexity and to increase in occupations with high complexity, 2) the ratio of female to male wages to decrease for young individuals and to increase for old individuals, and 3) a larger fraction of females to choose occupations for which the difference in wages between genders is smaller. Overall, the ratio of average female to male lifecycle wages increases from 0.68 in the baseline to 0.78 in the counterfactual experiment.

In the data, female labor supply increases across successive cohorts. How well does the model reproduce the gender wage gap for the 1950 cohort? Figure 9 shows the lifecycle profile of the gender wage gap for the 1950 cohort along with that for the 1920 cohort. The model

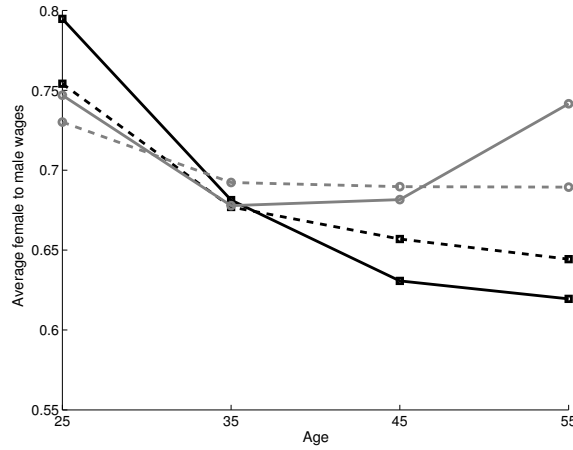


Figure 9: Results. The gender wage gap, 1920 cohort and 1950 cohort. Data (dashed lines) vs. Model (solid lines). Source: IPUMS-USA, O*NET and the author.

generates 50 percent of the decrease in the ratio of female to male wages for young workers that happens in the 1920 to 1950 cohorts: the ratio decreases from 0.75 to 0.73 in the model, and from 0.79 to 0.75 in the data. The model is also consistent with the change in the shape of the profile of the gender wage gap across occupations for young workers. In the model, this profile goes from a U-shape for the 1920 cohort to a non-decreasing shape for the 1950 cohort as it happens in the data.¹³ Over the lifecycle, the decrease in the ratio of female to male wages over the lifecycle in the data is milder for the 1950 cohort than it is for the 1920 cohort. The model reproduces 50 percent of this attenuation. Overall, the model generates 60 percent of the increase in the ratio of female to male lifecycle wages for successive cohorts born between 1915 and 1955.

The cross-cohort patterns are reinforced by the increase in female labor force participation and the improved occupational composition of female labor supply. Among the females of the 1950 cohort, 78.8 percent join the labor market in the model compared to 85 percent in the data. Consistent with Greenwood, Seshadri and Yorokoglu (2005), the increase in the fraction of females joining the labor market in the model is primarily caused by technological change in the home production technology, i.e., a decrease in φ . When I hold the home production technology constant across cohorts, the fraction of females that joins the

¹³I do not report figures for the gender wage gap across occupations for the 1950 cohort. These figures are available upon request.

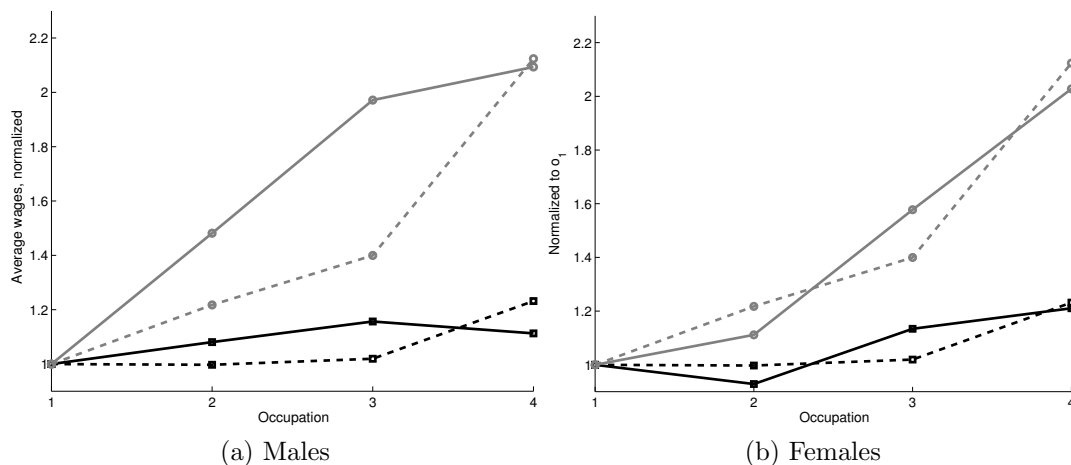


Figure 10: Results. The occupation premium, 1920 cohort. Black lines are for age 25 to age 35, and gray lines are for age 55 to age 65. Data (dashed lines) vs. Model (solid lines). Source: IPUMS-USA, O*NET and the author.

labor market increases only 4 percentage points. The evolution of occupational prices, instigated by technological change in the market good production, reinforces the improvement of the occupational composition of female labor supply. Prices of more complex occupational outputs increase relative to those of less complex occupational outputs. Because the distribution of initial skill for females has higher density toward the right end of the skill domain compared to that of males, females migrate toward complex occupations faster than males. For the 1950 cohort, complex occupations are those where the difference in wages between genders is the smallest.¹⁴

5.2 Occupation Premium

One of the most remarkable features of the twentieth century United States is the rise in the wages of skilled workers relative to those of unskilled workers (hereafter “skill premium”) starting in the 1980s. The literature has tied the concept of skill to various observables, such as the amount of schooling an individual has or the type of occupation an individual

¹⁴Others in the literature found evidence of a role of changes in the wage structure on the increase in female labor supply. See for example, Jones, Manuelli, and McGrattan (2003), Olivetti (2006), Rendall (2010) and Contessi and De Nicola (2012).

OCCUPATION AGE	MALES				FEMALES			
	25-35	35-45	45-55	55-65	25-35	35-45	45-55	55-65
2	0.0620	0.0681	0.0741	0.0801	0.0620	0.0681	0.0741	0.0801
	0.0357	0.1087	0.1455	0.1687	-0.0649	-0.0097	0.0189	0.0342
3	0.0987	0.1089	0.1190	0.1290	0.0987	0.1089	0.1190	0.1290
	0.0317	0.1470	0.2018	0.2359	-0.0796	0.0180	0.0683	0.0960
4	0.1779	0.1969	0.2156	0.2339	0.1779	0.1969	0.2156	0.2339
	-0.0506	0.1673	0.2546	0.3052	0.0306	0.2321	0.3443	0.4171

Table 6: Log decomposition of the skill premium, 1920 cohort. For each entry, the first row shows the component that depends on occupational characteristics, while the second row shows the component that depends on individual characteristics.

is employed in (see Goldin and Katz, 2008, and Acemoglu, 2002). In my model, the natural mapping is to occupations and the notion of the skill premium that follows is the ratio of average wages across occupations. In this section, I define the skill premium in occupation i to be the ratio of average wages in occupation i relative to that in occupation 1.

Figure 10 presents the model implications for the skill premium between ages 25 and 35, and between ages 55 and 65 for the 1920 cohort. The model replicates the structure of the skill premiums across occupations quite nicely for the case of females (Figure 11b). However, for the case of males, the skill premium is reproduced well only between ages 25 and 35. The skill premium for males in occupation 4 between ages 55 and 65 is 1.9 in the model, while it is 2.1 in the data.

What lies behind wage differentials across occupations? To answer this question, I decompose the logarithm of the skill premium in occupation i in a component that depends on occupational characteristics ($\Delta_{\log(o_i)} - \Delta_{\log(o_1)}$) and in a component that depends on individual characteristics ($\Delta_{\log(s_i)} - \Delta_{\log(s_1)}$) as follows:

$$\begin{aligned} \log(OP(i, 1, g)) &= \underbrace{\log(w_i o_i^{1-\rho}) - \log(w_{i'} o_1^{1-\rho})}_{\Delta_{\log(o_i)} - \Delta_{\log(o_1)}} \\ &\quad + \underbrace{\rho \left(\log\left(\int_{s_0} \mathbf{1}(\cdot, i)(s - o_i)^\rho d\Gamma_g(s_0)\right) - \log\left(\int_{s_0} \mathbf{1}(\cdot, 1)(s - o_1)^\rho d\Gamma_g(s_0)\right) \right)}_{\Delta_{\log(s_i)} - \Delta_{\log(s_1)}}, \end{aligned}$$

Table 6 shows the results of the decomposition, separately for males and females. On average,

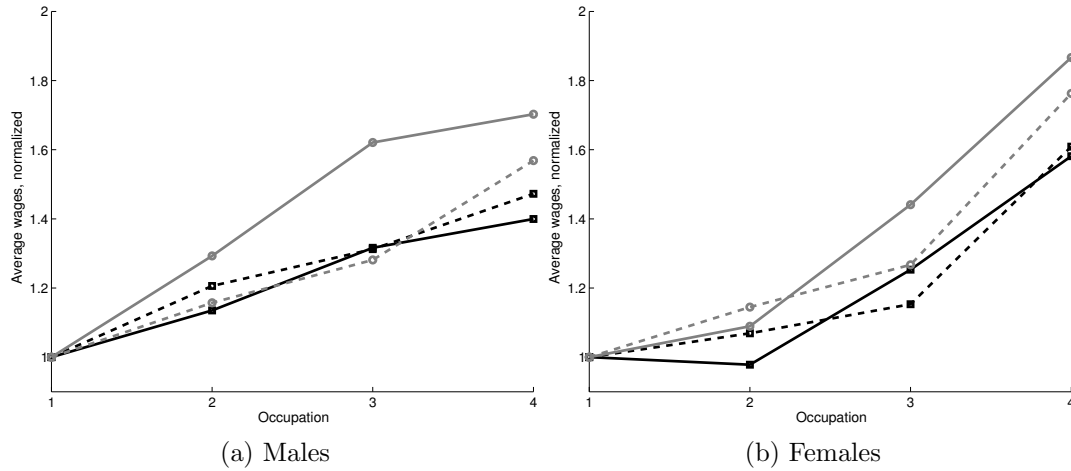


Figure 11: Results. Occupation premium, 1920 cohort (dark-colored lines) and 1950 cohort (light-colored lines). Data (dashed lines) vs. Model (solid lines). Source: IPUMS-USA, O*NET and the author.

the component that depends on occupational characteristics accounts for 30 to 40 percent of the skill premium for males. The importance of this component increases with the complexity of the occupation. For the case of females, the component that depends on occupational characteristics accounts for 65 percent of the skill premium in occupations 2 and 3, and for 40 percent of the skill premium in occupation 4.

Figure 11 presents the model implications for the skill premium for the 1950 cohort. The skill premium in occupation 4 for male workers increases from 1.5 to 1.6 in the 1920 to 1950 cohorts in the model; while it increases from 1.4 to 1.7 in the data. The skill premium in occupation 4 for female workers increases from 1.6 to 1.8 in the 1920 to 1950 cohorts in the model; while it increases from 1.6 to 1.9 in the data. The importance of the component that depends on occupational characteristics and that of the component that depends on individual characteristics for the skill premium across occupations change little across successive cohorts of male individuals. However, the importance of the occupational component in occupations 2 and 3 decreases across successive cohorts of female individuals. This is because as more females join the labor market, the average initial skill in occupation 1 decreases faster than the average initial skill in other occupations.

6 Conclusions

In this paper, I document two patterns of the gender wage gap across occupations of different complexity, for cohorts born between 1915 and 1955. First, the ratio of female to male wages for young individuals is U-shaped across occupations ordered by increasing complexity. The U-shape becomes flatter for successive cohorts of young individuals. Second, over the lifecycle, the ratio of female to male wages decreases faster for more complex occupations. The decrease becomes weaker for successive cohorts over the lifecycle. I argue that understanding these patterns is central to the understanding of the overall gender wage gap, since the value of an individual's skill and the scope of learning by doing in an occupation depend on the complexity of the tasks the occupation entails.

I write a model of occupational choice and learning by doing where individuals differ by their skill and occupations differ by the skill required to perform and the marginal product of skill. I calibrate the model to the task content of occupations, occupational choices, and major patterns of labor supply and male wages, for the cohorts born between 1915 and 1955. The model quantitatively reproduces the two patterns of the gender wage gap across occupations for the cohorts born between 1915 and 1955, the overall gender wage gap for the cohorts born between 1915 and 1925, and 60 percent of the decrease in the ratio of females to male wages for successive cohorts born between 1915 and 1955.

The overall gender wage gap is for the most determined by gender differences in skill, endowed and acquired through work experience, within each occupation. Through counterfactuals, I find that work experience alone accounts for 69 percent of the lifecycle gender wage gap. Part of this percentage is due to occupational choice. As differences in work experience between genders disappears, females migrate toward occupations for which the difference in wages between genders is small. Over time, technological change reinforces the narrowing of the gender gap in wages. In the home good production, it increases female experience and in the market good production, it improves the occupational composition of female labor supplied.

References

- [1] Acemoglu, D. (2002), “*Changes in the Characteristics of American Youth: Implications for Adult Outcomes*,” *Journal of Economic Literature*, 40(1), pp. 7-72.
- [2] Acemoglu, D., Autor, D. (2010), “*Skills, Tasks and Technologies: Implications for Employment and Earnings*,” Chapter in Orley Ashenfelter and David E. Card (eds.) *Handbook of Labor Economics*, vol. 4, Amsterdam: Elsevier.
- [3] Black, S., Spitz-Oener, A. (2010), “*Explaining Women’s Success: Technological Change and the Skill Content of Women’s Work*,” *Review of Economics and Statistics*, 92(1), pp. 187-194.
- [4] Blau, F., Kahn, L. (2000), “*Gender Differences in Pay*,” *Journal of Economic Perspectives*, 14(4), pp. 75-99.
- [5] Bowlus, A., Robinson, C. (2010), “*Human Capital Prices, Productivity and Growth*,” *American Economic Review*, forthcoming.
- [6] Card, D., DiNardo, J. (2002), “*Skill Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles*,” NBER Working Paper 8769.
- [7] Contessi, S., De Nicola, F. (2012), “*Trade, Female Labor, and Female Entrepreneurship in MENA countries*,” Federal Reserve Bank of St. Louis Working Paper 2012-053.
- [8] Erosa, A., Fuster, F., Restuccia, D. (2010), “*A Quantitative Theory of the Gender Gap in Wages*,” IMDEA Ciencias Sociales Working Paper 2010-04.
- [9] Fernández, R., Fogli, A., Olivetti, C. (2004), “*Mothers and Sons: Preference Development and Female Labor Force Dynamics*,” *Quarterly Journal of Economics*, 119(4), pp. 1249-1299.

- [10] Firpo, S., Fortin, N., Lemieux, T. (2011), “*Occupational Tasks and Changes in the Wage Structure*,” IZA Working Paper, Institute for the Study of Labor.
- [11] Goldin, C. (2006), “*The Quiet Revolution That Transformed Women’s Employment, Education, And Family*,” American Economic Review, 96(2), pp. 1-21.
- [12] Goldin, C., Katz, L. (Eds.) (2008), “*The Race between Education and Technology*,” Harvard University Press.
- [13] Greenwood, J., Seshadri, A., Yorokoglu, M. (2005), “*Engines of Liberation*,” Review of Economic Studies, 72(1), pp. 109-133.
- [14] Hendricks, L., Schoellman, T. (2011), “*Student Abilities During the Expansion of U.S. Education, 1950-2000*,” Mimeo, University of North Carolina.
- [15] Hsieh, C.-T., Hurst, E., Jones, C., Klenow, P. (2012), “*The Allocation of Talent and U.S. Economic Growth*,” Working paper.
- [16] Jones, L., Manuelli, R., McGrattan, E. (2003), “*Why are married women working so much?*,” Federal Reserve Bank of MinneapolisMStaff Report 317.
- [17] Kambourov, G., Manovskii, I. (2008), “*Rising Occupational and Industry Mobility in the United States: 1968-1997*,” International Economic Review, 49(1), pp. 41-79.
- [18] Knowles, J. (2012), “*Why Are Married Men Working So Much?*,” Review of Economic Studies, forthcoming.
- [19] Kuruscu, B. (2006), “*Training and Lifetime Income*,” American Economic Review, 96(3), pp. 832-846.
- [20] Mulligan, C., Rubinstein, Y. (2008), “*Selection, Investment, and Women’s Relative Wages Over Time*,” The Quarterly Journal of Economics, 123(3), pp. 1061-1110.

- [21] Olivetti, C. (2006), “*Changes in Women’s Hours of Market Work: The Role of Returns to Experience*,” *Review of Economic Dynamics*, 9(4), pp. 557-587.
- [22] Rendall, M. (2010), “*Brain versus Brawn: The Realization of Women Comparative Advantage*,” Institute for Empirical Research in Economics Working Paper iewwp491.
- [23] Roy, D. (1951), “*Some Thoughts on the Distribution of Earnings*,” *Oxford Economic Papers*, 3 (2), pp. 135-146.
- [24] Ruggles, S., Alexander, T., Genadek, K., Goeken, R., Schroede, M., Sobek, M. (2010), “*Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]*,” Minneapolis: University of Minnesota.
- [25] Rupert, P., Zanella, T. (2012), “*Revisiting Wage, Earnings, and Hours Profiles*,” UCSB Economics Working Paper qt61f2f1hv.
- [26] Schoellman, T. (2010), “*The Occupations and Human Capital of U.S. Immigrants*,” *Journal of Human Capital*, 4(1), pp. 1-34.
- [27] Yamaguchi, S. (2012), “*Changes in Returns to Task-Specific Skills and Gender Wage Gap*,” Available at SSRN: <http://ssrn.com/abstract=2035833> or <http://dx.doi.org/10.2139/ssrn.2035833>.

A Data sources

IPUMSUSA I use 1 percent samples for the period 1950-1970, and 5 percent samples for the period 1980-2008. Observations are weighted. I restrict the sample to employed, married, white individuals who reported their occupation and a total amount of annual hours worked of at least 400. I construct a panel of occupations that includes all occupations that have complete crosswalks across the different Census Bureau occupational classification systems for the period 1950-2010. The panel contains a total of 272 occupations.¹⁵

My measure of earnings is the IPUMS variable INCWAGE. It reports total pre-tax wage and salary income, i.e. money received as an employee for the previous calendar year, as midpoints of intervals (instead of exact dollar amounts). I compute real earnings by applying the CPI weights.

My definition of an individual's occupation is the IPUMS variable OCC1950. This variable reports the occupation of an individual according to the three digits 1950 Census Bureau occupational classification system.

My measure of weekly working hours is the IPUMS variable HRSWORK2 for the period 1950-1970, and the IPUMS variable UHRSWORK for the period 1980-2010. HRSWORK2 reports the total number of hours the respondent was at work during the previous week. UHRSWORK reports the total number of hours that the respondent usually worked during the previous year. My measure of the annual weeks worked is the IPUMS variable WKSWORK2. WKSWORK2 reports the number of weeks that the respondent worked for profit, pay, or as an unpaid family worker during the previous year. I compute annual hours as the product of weekly hours worked and the number of weeks worked in a year.

For each cohort, the number of females joining the labor market is computed as the total number of working females between age 45 and age 55 for that cohort. Similarly, the fraction of individuals of a cohort in an occupation is computed as the fraction of the individuals of that cohort that chooses that occupation between age 45 to age 55. For each cohort, average hours worked at age j are computed as the sum of total hours worked at age j for that cohort divided by the number of individuals in that cohort who are in the labor market between age 45 and age 55. For each cohort, average earning at age j are the sum of annual earnings at age j for that cohort divided by the number of individuals of that cohort who are in the labor market between age 45 and age 55. Finally, an individual's wage is the ratio of his/her annual earnings and his/her annual working hours.

O*NET. To build my index of tasks complexity of an occupation, I use the tasks measures constructed by Acemolgu and Autor (2010) using the Occupational Information Network (O*NET) Database. For these tasks measure, I compute two aggregate indexes, *Brain* and *Brawn*. Figure 12 shows the distribution of occupations on the brain and brawn dimensions. Acemolgu and Autor (2010)'s tasks measures are available for occupations classified according to the three-digit 1990 Census Bureau occupational classification system. I use the crosswalks provided by IPUMS-USA to attribute the tasks measures to occupations classified according to the 1950 Census Bureau occupational classification system. When a three-digit

¹⁵The list of occupation is available upon request.

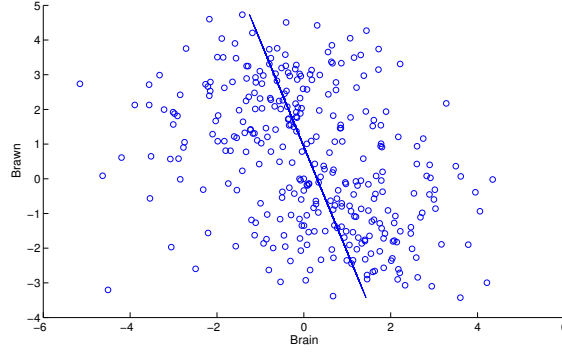


Figure 12: Occupations on the brain/brawn space. Source: O*NET.

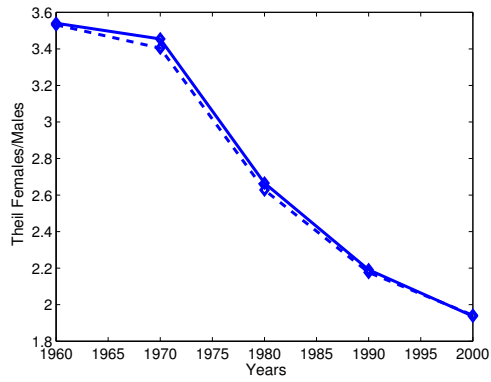
1990 occupational entry is mapped to multiple three-digit 1950 occupational entries, I split the 1990 Census weight equally across the corresponding occupations.

B Decomposition of the Dispersion of Market Hours

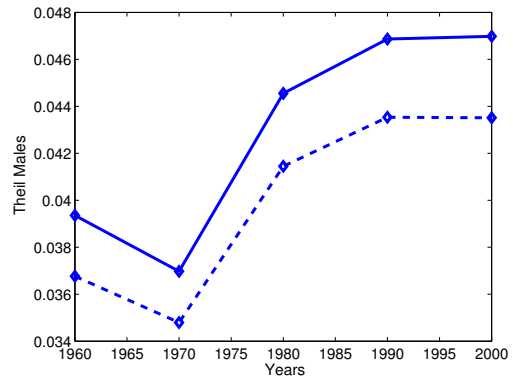
I use the Theil index to decompose the dispersion of market hours in two components: a component that captures within-group dispersion, and a component that captures between-group dispersion. I consider two grouping variables: gender and occupation. Take n to be the number of observations for variable z and define $T \in [0, \log(n)]$ to be the Theil index for variable z under the grouping variable j . The decomposition is as follows: Decomposition

$$\begin{aligned}
 T &= \sum_{j=1}^J y_j T_j + \sum_{j=1}^J y_j \log(J y_j) \\
 &= \underbrace{\sum_{j=1}^J y_j \sum_{i=1}^{n_j} y_{i,j} \log(n_j y_{i,j})}_{\text{within group}} + \underbrace{\sum_{j=1}^J y_j \log(J y_j)}_{\text{between group}},
 \end{aligned}$$

where J is number of groups and $y_i = \frac{z_i}{\sum_{i=1}^n z_i}$. Figure 13 plots the dispersion of working hours within gender groups. The left panel of Figure 13 shows that the dispersion of working hours within females is three times the dispersion of working hours within males in 1960 and it decreases substantially over time. The left panel of Figure 13, shows that the dispersion of working hours within males increases over time starting from 1970. However, the level it reaches in 2000 is still lower than the dispersion of working hours within females in 1960.



(a) Females to males dispersion



(b) Males dispersion

Figure 13: Within group dispersion of market hours. The dash line considers only gender as a grouping variable. The solid line consider both gender and occupation as grouping variables. Source: IPUMS-USA.