# GREAT EXPECTATIONS? Women's Work and Fertility in the Face of Career Uncertainty\*

Joyce Cheng Wong<sup>†</sup> Department of Economics, New York University

Job Market Paper

#### Abstract

Women in different occupations exhibit markedly different patterns of fertility and labor supply over the life-cycle. Those in occupations which require relatively short periods of career training (e.g., waitresses, teachers, secretaries) become mothers earlier and have labor force participation rates which are lowest in their 20s and increasing thereafter. "High-powered" professional women, on the other hand, give birth at later ages, have the highest participation rates during their 20s, and decreasing participation thereafter. I examine the hypothesis that these different patterns can be explained by the differences across occupations in: (i) the number of years that it takes for an individual to learn her true productivity and (ii) the wage penalties that accompany work interruptions.

To investigate the quantitative contributions of each of these features, I use an incomplete markets life-cycle model in which married households make fertility and female labor supply decisions. In the model, agents are heterogeneous in the parameters which govern their life-cycle wage profiles. Households do not observe their own parameters and must instead learn about them through occupation specific wage draws and productivity signals. The model calibrated to the National Longitudinal Survey of Youth 1979 (NLSY79) finds that the time required for the resolution of uncertainty about one's wage growth (i.e. learning) is key in explaining the different fertility patterns of women in high-powered professions. On the other hand, the differences in wage penalties due to work interruptions help explain why a significant proportion of high-powered professional women do not return to the labor market after exiting upon childbirth. These features combined generate both the increasing participation for women in professions with low training and the large drop in participation rates of high-powered professional women, across the life-cycle. The welfare losses due to imperfect information are negligible for most of the economy, but for the households with high-powered wives they correspond to almost 5% of consumption.

<sup>\*</sup>November 11, 2011. I thank Raquel Fernández, Gianluca Violante, Thomas Sargent and Christopher Flinn for their invaluable support and advice. I am grateful to David Backus, Laura Veldkamp and Matthew Wiswall as well as the participants in the Stern Macro Lunch Seminar for their helpful comments and suggestions. All errors are my own.

<sup>&</sup>lt;sup>†</sup>19 W 4th St. 6FL, New York, NY 10012. Email: cwong@nyu.edu Website: https://files.nyu.edu/cw526/public

# 1 Introduction

Women in different occupations exhibit very different work and fertility patterns across the life-cycle. If one compares those women in occupations which require relatively low amounts of vocational learning (e.g., waitresses, teachers, secretaries) to their "high-powered" professional counterparts (e.g., the CEOs, lawyers, doctors) significant differences emerge in both work and fertility behaviors. First, the life-cycle labor participation profiles of these two groups of women are strikingly different: while those in occupations which require relatively short periods of training have lower participation rates in their 20s (around 73%) and *increasing* thereafter, their "high-powered" counterparts have, instead, very high participation rates during their 20s (over 95%) but *decreasing* thereafter. By age 48, both groups have participation rates around 85%. Examining fertility behaviors, both the average and the distribution of mother's age at first birth in each of these groups exhibit remarkable differences. Almost half (47%) of the women in short-training occupations who ever become mothers do so before the age of 23, and the distribution of the mother's age at first birth is very concentrated, with a clear peak at age 22. On the other hand, only 34% of the high-powered mothers have given birth by age 23 and the distribution is much more spread out, with the majority of first time mothers evenly distributed throughout all ages from 22 to 26.

In this paper, I examine whether these different patterns can be explained by differences across occupations in: (i) the number of years that it takes for an individual to learn her true productivity in that occupation and (ii) the wage penalties that accompany work interruptions. For example, while a cashier or a waitress can discover her true aptitude for the job in a relatively short amount of time, most lawyers must work for about eight years after law school before they are considered for a partnership, and academic tenure is only considered about seven years after the doctoral degree. Work interruptions also have differential effects depending on the occupation. While skills lost during inactivity can be easily and quickly reacquired for those women in short-training occupations, non-participation periods for those in "high-powered" careers can be very costly both due to technological improvements in the field and because lost skills take longer to be replaced. One year out of the labor market for a cashier has only small consequences for career advancement, but the same interruption for a lawyer can have significant effects on partnership prospects.

I develop a life-cycle model of married households which incorporates both of these features. In my model, agents work in one of two possible occupations: one in which agents can quickly learn their true productivity and one in which agents must work for many years before their productivity is revealed. Agents are also endowed with occupation-specific heterogeneous productivity levels which are unknown in the beginning of their working life. Individuals learn about their own income prospects by observing wage draws every period, and those who choose to work observe further signals about their own productivity. Each period, households choose consumption/savings and the labor participation decision of the wife. Couples also choose the age at which they first become parents. Children provide utility to the household but they also deflate household consumption and impose a psychic and a monetary cost (childcare) to female work. Work interruptions lead to wage loss, the magnitude of which is occupation specific.

The model calibrated to a few aggregate statistics from the National Longitudinal Survey of Youth 1979 (NLSY79) does a very good job capturing the differences in work and fertility decisions of women in different occupations. Using counterfactual experiments, I quantify the contributions of learning and wage penalties in generating the different participation profiles and fertility timing. With regards to work behaviors, I find that a combination of both of these features is necessary in order to explain the declining participation rates for women in the high-learning occupation. Both imperfect information and wage penalties increase the value of working when agents are young: the former through informational gains and the latter through costly non-participation. For women in high-learning occupations, the combination of these features generates strong incentives to work while young and delay fertility. However, in the model as in the data, most women become mothers at some point in their lives. Some of these mothers choose to exit the labor market after childbirth due to the high costs of participation when there are young children in the household. After children grow up, the significant wage loss associated with a career interruption can explain why a many women in high-learning occupations do not return to the labor market.

While a combination of features is necessary to explain participation profile differences, I find that imperfect information alone can account for differences in fertility timing. The key for the differential effect of learning on fertility versus work decisions lies in the dynamics links of each of these choices. While the decision to become a parent has significant consequences for future periods (e.g., higher disutility from labor for the wife, consumption deflation), the participation decision has relatively weak dynamic links. Thus uncertainty and beliefs about future wages prospects plays a much larger role in the once-in-a-lifetime decision to become parents than in the period-by-period participation choice. This effect is so strong that in a counterfactual where households have perfect information about their wage parameters, the proportion of women in the high-learning occupation who become mothers at age 23 is 27.3% compared to 8.2% in the benchmark.

Given that I've constructed a model in which learning plays a central role, I also examine the effects of different beliefs on household decisions. While they have some effect on the decisions of households with wives in the low-learning occupation, they are crucial in determining outcomes in households with high-learning wives. Examining two high-learning women who are identical except for their initial priors, I find that initial over-optimism about one's productivity is related to both delayed fertility and surprisingly, lower participation across the life-cycle. In the model, this is because household assets at the time of childbirth have a negative effect on the wife's participation. Thus the woman who delayed fertility in order to work also has high household wealth and lower incentives to work while she has young children at home.

Welfare losses due to imperfect information reflect the importance of learning on the fertility and labor supply behavior of women in the high-learning occupation. While they are negligible for the majority of the households in the model (less than 1%), they are around 5% for households whose wives work in occupations with high-learning. These losses stem from two different channels: first, and as explained before, learning causes these women to delay fertility beyond what they would like under perfect information. Second, and perhaps more interestingly, some welfare loss comes from the interaction between imperfect information and career interruptions. Since the probability of conception declines with age, some women in the high-learning occupation choose to become mothers and exit the labor market before they have fully learned their productivity. In the model, these are also the women who are extremely productive and although some of them return to work after a few years, their wages are much lower than they would have been in the absence of a career interruption. As a result, the top observed female wage at ages 45-50 is only 34.9% of the top male wage, at a time when the *average* female wage is 67.6% of the average male wage. Thus the model finds that imperfect information may play a role in generating what appears to be a glass-ceiling on wages.

This paper contributes to the literature on women's work and fertility choices. The large increase in female labor force participation across the 20th century has been analyzed in countless papers.<sup>1</sup> As women

<sup>&</sup>lt;sup>1</sup>Some explanations for the rise in the labor market participation of married women include technological change in the household (e.g., Greenwood, Seshadri, and Yorukoglu (2005)), changes in the gender wage gap (e.g., Jones, Manuelli, and McGrattan (2003)) and returns to experience (e.g., Olivetti (2006)), the introduction of the infant formula (e.g. Albanesi and Olivetti (2009)) and the drop in child care costs (e.g., Attanasio, Low, and Sanchez-Marcos (2008)), both of which lightened

became ubiquitous in the labor market, they have also began entering the ranks of the "high-powered" professionals. The reasons provided in the literature for this shift can be separated into two groups: factors which changed labor demand and advances in technology which influenced fertility. Papers in the first group argue that increased within gender inequality (e.g., Mulligan and Rubinstein (2008)) and skill-biased technological change (e.g., Rendall (2010), Black and Juhn (2000)) lead women to invest in human capital and enter the highly skilled careers. Papers which explore the role of fertility include Goldin and Katz (2002), Bailey (2006) and Knowles (2007) who argue for the importance of the dissemination of the use of the oral contraceptive and the legalization of abortion in changing women's career choices.

Although women have entered high-powered careers, their work and wage histories are still not commensurate to those of men. For example, women in top business careers (e.g., Bertrand, Goldin, and Katz (2010)), lawyers (e.g., Wood, Corcoran, and Courant (1993)), and those women who have elite higher education (e.g., Goldin and Katz (2000) and Herr and Wolfram (2010)) by age 50, have, on average lower wages and experience than their male counterparts. A large portion of these differences can be explained by shorter hours and career interruptions associated with maternity.

Given its importance in a woman's life, fertility has long been studied as a key process to explain differences in the work behaviors of men versus women. Starting with the seminal work of Mincer and Polachek (1974) who argued that women who expect future career interruptions associated with fertility underinvest in human capital, a large body of literature has examined the interactions between women's work and fertility decisions. A few papers such as, for example Erosa, Fuster, and Restuccia (2002) and Caucutt, Guner, and Knowles (2002), study these decisions in a general equilibrium model and find that fertility delay plays a large role the determination of women's earnings later in life. In a different vein, a mostly empirical literature has focused on the estimation of wage differences between mothers and women without children - the so called "family gap" (e.g., Waldfogel (1997, 1998) and Budig and England (2001)).

To my knowledge, however, there are no papers which explicitly examine the effects of learning on the life-cycle work and fertility choices of women. Although Fernández (2011) and Fogli and Veldkamp (2011) analyze the effects of learning on female labor force participation throughout the 20th century, both papers focus on the aggregate dynamics of women's participation and abstain from the life-cycle features of women's work. Moreover, in those models women learn about either the long-run disutility

the burden of working mothers.

from labor or the effects of maternal work on children. Closer to my setup is Guvenen (2007) who used a life-cycle model in which agents learned about their own productivity in order to study the effects of imperfect information on life-cycle consumption. However, households in that model were composed of a single male agent who was assumed to always work, and thus the paper was silent on the implications of imperfect information on women's work and household fertility choices.

The paper is organized as follows. Section 2 presents facts which motivate the main question of the paper. Section 3 develops the dynamic life cycle model and section 4 discusses its parameterization. Section 5 explores the model's implications concerning the roles of learning and beliefs versus wage penalties. Section 6 analyzes the welfare losses due to imperfect information and section 7 concludes.

### 2 Some Facts

This section describes some facts which motivate the main question of this paper. Statistics on wages, work (hours and occupation), and fertility presented in this section are computed from the National Longitudinal Survey of Youth 1979 (NLSY79) while occupational characteristics are drawn from the Dictionary of Occupational Titles (DOT). The NLSY79 is a panel dataset composed of a representative sample of 12,686 young men and women in the US who were between the ages of 14-22 when the survey began in 1979. The DOT provides detailed characteristics for over 12,000 occupations in the United States. Each occupation is given a measure for characteristics such as the formal and informal education needed in order to enter the occupation, aptitudes which are required of the worker (math, coordination, reasoning skills, physical strength, among others), the environmental conditions surrounding it, among others.

The main variable of interest for this paper is the Specialized Vocational Preparation (SVP) variable. It measures the amount of time (in months and years) for a typical individual to learn the necessary occupation specific skills in order to preform at an average level. The SVP variable is coded as a categorical variable ranging from "a short demonstration only" to "more than 10 years".<sup>2,3</sup> In order to investigate

<sup>&</sup>lt;sup>2</sup>The exact categories are (1) a short demonstration only, (2) more than a short demonstration, up to a month, (3) more than one month, up to three months, (4) more than three months, up to six months, (5) more than six months, up to twelve months, (6) 1-2 years, (7) 2-4 years, (8) 4-10 years and (9) more than 10 years. Note that these time requirements may include some vocational education, e.g., medical school, law school, ...

<sup>&</sup>lt;sup>3</sup>Although the DOT has now become the O\*Net, the SVP variable was discontinued in the O\*Net. Although O\*Net reports SVP values, those are simply the ones from the DOT.

the time trends in men and women's occupational choices, I use the crosswalk developed by Meyers and Osborne (2005) in order to map DOT task characteristics to the Census occupation codes in the NLSY79. Census codes which map into more than one DOT occupation are assigned the SVP averaged over all relevant DOT occupations.<sup>4</sup> Lastly, I transform their SVP measure for each Census occupation into a new continuous variable denominated in *months* of training.<sup>5</sup> These continuous SVP values then are merged with the occupation data in the NLSY79.

Occs. with Low Learning $(SVP < 7 \text{ yrs})$	Occs. with High Learning $(SVP \ge 7 \text{ yrs})$
Policemen, Administators	Accountants, Actuaries
Construction workers, Cooks	Lawyers, Physicians
Therapists (Physical, Speech)	Engineers, Designers
Truck drivers, Clerks	Managers, CEOs
Mechanics, Plumbers, Carpenters	Architects, Professors
Cashiers, Waitresses, Secretaries	Inspectors (Insurance, Safety)
Teachers, Nurses, Paralegals	Head Chefs, Farm Owners
Teachers' aids, Teachers	Electricians, Masons

Table 1: Examples of occupations with low and high learning

For ease of exposition, I divide occupations into two groups: those which require under 7 years of SVP (the occupations with "low learning") and those which require more than 7 years of SVP (the occupations with "high learning").<sup>6</sup> Table 1 reports some examples of occupations in each of the learning categories. Note that this split does not directly correspond to a division by education. For example, while several occupations in the low learning category require a college education (teachers, nurses), some occupations in the high-learning category require mostly many years of training and apprenticeship as opposed to higher education (chefs, masons).

Throughout the paper, I will assign individuals into each of the occupation groups in the following way.

<sup>&</sup>lt;sup>4</sup>Their exact mapping system is outlined in Appendix C of Meyers and Osborne (2005).

<sup>&</sup>lt;sup>5</sup>Given that SVP is a categorical variable and the exact categories are given in footnote 2, I proceed in the following manner. First, I assign each category it's median length, e.g. if the occupation has been assigned a category 6 in SVP (1-2 years), it will correspond to a continuous length of 18 months; 36 months if it's category 7. For each occupation code in the CPS, the mapping system of Meyers and Osborne (2005) will give it an SVP value which will correspond to the average SVP across all DOT occupations which mapped to that Census occupation code. I respect these intermediate values by giving them a weight average corresponding to the lengths in the adjacent categories. Thus if an occupation has SVP according to Meyers and Osborne (2005) of 6.23, it will be mapped to  $18 + [0.23 \times (36 - 18)]$  months

 $<sup>^{6}</sup>$ An earlier draft of this paper used a finer division into three different learning groups (low = under 2 years, medium = between 2 and 7 years and high = more than 7 years) which yielded very similar patterns between those in the low and medium learning groups. For this reason, I choose to only aggregate occupations into two main learning categories. Data and some results for the three group specification is reported in the Appendix.

	Dependent Variable (age 48)		
	Log Hourly Wages	Participation	
	(OLS)	(Probit)	
Low Learning	-0.0888**	-0.0697	
	(0.0444)	(0.0658)	
High Learning	-0.1878**	-0.3358**	
	(0.0927)	(0.1505)	

Table 2: Effect of a career interruption during ages 23-32, on log hourly wages (45-50) and participation at age 48

The numbers reported are the coefficients of a dummy variable which takes on a value 1 if the individual was observed not working for at least one year between the ages of 23 and 32, given that she was not a full-time student. Both the OLS and the Probit specifications controlled for whether the woman is married with a spouse present and her race. See the Appendix for sample construction. \*\*\* = significant at 2%, \*\* = significant at 5% and \* = significant at 10%.

I take advantage of the NLSY79's panel structure and assign each individual to one occupation group, based on his/her observed work history. In with the findings of Topel and Ward (1992) and Kambourov and Manovskii (2009), occupational switches in my sample occur mostly for individuals under the age of 30. After the age of 35, only 0.6% switch between occupations which belong to different learning categories (62 individuals). For those who never switch across occupational categories (over 80% of the sample), I simply assign them to their observed occupational group. For consistency with my model which does not include occupational switches, I exclude those who switch occupations after the age of 35 from the sample. For those who are observed switching occupational categories at least once, I assign them to an occupational group based on the first occupation they report after the age of 35. This is mostly a concern for people who are observed in low-learning occupations early in life and switch to high-learning occupations later in life, as there are very few people who switch in the opposite direction. Thus I implicitly assume that the switchers in the sample chose an occupation when they were younger and were working towards it until the age of 35.

The occupations in the high learning category are characterized not only by the requirement of long periods of career training but also three other key characteristics. First, it takes many years for individuals who work in these occupations to find out their true productivity in them, i.e. how "good" they really are at those jobs. For example, it takes on average 8 years after law school in order for lawyers to be considered for partnerships, 7 years after the PhD to be considered for tenure in academia and an average partner in an accounting firm has had around 13 years of experience. Physicians must face at least 3 years

of residency after medical school (7 years for surgeons) before becoming attending physicians and in the business world, the median age of the Fortune 500 CEO is 54 years old. These occupations usually have long career ladders and individuals must work for many years in their fields before the acknowledgment that they are truly successful in their profession.



Figure 1: Average log wages of men, by occupational learning intensity

Occupational learning group for each individual in the sample is assigned according to the first occupation observed for that individual after the age of 35.

Second, work interruptions for people in high-learning occupations are associated with both large wage penalties and lower probability of participation later in life. Table 2 reports the results of two different regressions: one in which the dependent variable is log hourly wages during the ages of 45-50, and a second regression in which the dependent variable is an indicator for whether the woman works or not during those ages.<sup>7</sup> The reported coefficients are those for a dummy which takes on a value of 1 if the individual was observed not working for at least one year between the ages of 23 and 32, given that she was not a full-time student. As one can see, women who did not have career interruptions during the ages of 23-32, have significantly higher log wages than those of their counterparts who are observed not working for at least one year of their counterparts who are observed not working for at least one year for those in high-learning occupations is more than twice the size of the gap for those in low-learning occupations. Examining the results when the dependent variable is participation at 45-50 (second column), one finds a similar difference between those women in

 $<sup>^{7}</sup>$ I consider a woman to be working between the ages of 45-50 if she reports positive hours worked during at least three years in that interval. I do this because the NLSY79 is biennial when most of the individuals in the sample reach these ages.

occupations with low versus high learning. In particular, a career interruption significantly decreases the probability of a woman working at age 48 for those women in high learning occupations, but it has no effect on the participation of women in occupations with short learning.



Figure 2: Percent Women Working, by occupational learning intensity

Occupational learning group for each individual in the sample is assigned according to the first occupation observed for that individual after the age of 35.

Note that both lower probability of participation and lower wages associated with career interruptions can be due to a mechanism akin to the one examined in the seminal work of Mincer and Polachek (1974). In that paper, the authors argue that women who do not expect to work later in life underinvest in human capital and incur lower hours and career interruptions during early career. In this case, however, it is interesting how these effects differ across occupations; a few recent papers have examined similar trends. For example, Anderson, Binder, and Krause (2002) show that the the wage difference between mothers and non-mothers (the "motherhood wage penalty") is nearly non-existent for women without a high-school diploma, but around 29% for college mothers of 2 children. In a similar vein, Wilde, Batchelder, and Ellwood (2010) find that wage trajectories diverge sharply for high-ability mothers versus non-mothers after childbirth, while there is little to no difference for low-ability women.

Finally, the wage profiles of individuals in each of these two occupation groups are also very different. Figure 1 plots the evolution of average log hourly wages by age for men in each of the two types of occupations. One striking feature is the fact that not only is the level of wages higher for those in high learning occupations but the growth rate of wages is also much higher. Although log wages are very similar between the group up until age 25, they diverge from then on and by age 48 the gap between them is almost 80 log points. A second interesting fact is that while wages of those in low learning occupations plateau soon after 35, the wages of those in high learning occupations continue growing well into their late 40s. One reason for the low wages of those in high learning occupations before 25 could be that those individuals are still in school during those ages. However, the increase from age 25 to 48 for those in high learning occupations is still larger than the total increase from the age of 19 to 48 for those in low learning occupations (94 log points versus 76 log points).

Figure 3: Distribution of mother's age at first birth, by occupational learning intensity



Next, I proceed to examine work and fertility behaviors of women across these different occupation groups. Figure 2 plots the percentage of women working, by age and type of occupation.<sup>8</sup> Overall, women in high-learning occupations work more than their low-learnign counterparts at all ages. Their participation profiles across their life-cycle, however, are distinctly different. Those women in occupations with low levels of learning have low participation rates during their early to mid 20s (around 73%), increasing thereafter until they reach almost 85% by age 42. Women in high learning occupations,

<sup>&</sup>lt;sup>8</sup>Many of those women in high-learning occupations (e.g., lawyers, doctors, chefs...) are still full-time students in professional or vocational schools during most of their 20s. I will consider these women to be working since they can be perceived as "apprenticing" towards their profession.

however, have participation rates which are over 95% between the ages of 23 and 26, but which decrease thereafter. These changes are significant enough that by age 48 the participation rates of women in these two occupation group are almost the same: 84.5% for low-learning versus 85.8% for high-learning.





women aged 30 and over of getting pregnant within 12 months of trying.

Women in different occupation groups also exhibit markedly different fertility timing behaviors. Figure 3 plots the distribution of ages at first birth for women in each of these groups. As expected, women in low-learning occupations have children earlier while those in high-learning occupations have them later. However, it is interesting to point out that the shapes of the distribution are also very different. The distribution of women's age at first birth for those in low-learning occupations is more concentrated with a clear peak at the age of 22. On the other hand, women in high learning occupations have a distribution of ages at first birth which is much more spread out, with a peak which lasts all throughout ages 22 to 26. A second interesting feature is the fact that although more women in high learning occupations have children later, still less than 6% of them become mothers for the first time after the age of 35. This is simply due to the biological fact that a woman's probability of conceiving a (healthy) child declines very quickly after the age of 30. As shown in Figure 4 (taken from van Noord-Zaadstra, Looman, Alsbach, Habbema, te Velde, and Karbaat (1991)) a woman's relative probability of conceiving a healthy child at age 35 is half the probability faced by her 20-29 year old counterpart.

## 3 Economic Model

In this section, I develop a model whereupon households composed of a married man and a married woman jointly decide on consumption, savings and the wife's labor force participation status each period. The couple also decides when to become parents to two children. I focus on the woman's participation decision since most prime-age males are observed working in the data. I choose to model married households because even though the age at first marriage has been increasing, the vast majority of the population marries at least once during their lives. In my sample by 2008, only 8.8% had never been married (ages of 44-51). A key feature of the model is that agents differ in both the parameters which govern their wage profiles and in their occupations. Furthermore, households are assumed to not know the husband and wife's wage parameters and must learn about them through wage draws and signals. In this section, I describe the model's demographic structure, preferences and agents' life-cycle structure with special focus on the agents' learning process over income parameters.

### 3.1 Demographics, Preferences and Life-cycle

The environment is stationary, time is discrete and indexed by t = 1, 2, ..., T. The economy is populated by a continuum of households indexed i = 1, ..., I. Each household is composed of two agents, one of each gender  $g \in \{m, f\}$ : a male (m) and a female (f) agent who are assumed to be married to each other for life. I will from now on refer to the male agent as the husband and the female agent as the wife. Since this is a stationary model with only one cohort, the time index t corresponds to the age of the household at that time. Each household enters the labor market together at age 23 (t = 1 in the model) and retire at age  $t = t_R$ . The husband and wife are assumed to die together at age t = T; thus there is no uncertainty in longevity or retirement.

Every period, the household decides on consumption, savings, and the wife's labor participation. As it is common in the literature on female labor force participation and for computational simplicity, men are assumed to always work. During ages  $t \leq t_F$ , the household will also make a decision about when to become parents. If they choose to do so at time t, with probability  $p_t$ , their first child arrives at time t + 1 and their second child arrives two periods after the first. With probability  $1 - p_t$ , the couple does not receive children. The arrival of children is modeled as an absorbing state. This means that once the couple receives children, they will not be allowed to make any more fertility decisions thereater. Parents derive utility from children who in turn, deflate household consumption by their adult equivalent and affect the mother's disutility of labor. During retirement, there is no more uncertainty. Households consume and save and they receive a constant pension payment b every period until death.

Households in the model are unitary in the sense that the husband and wife are assumed to have common preferences  $u_t(c_t, P_t; k_t)$ . These are defined over market consumption per adult equivalent  $(c_t \ge 0)$ , the wife's labor market participation  $(P_t \in \{0, 1\})$  and they are also a function of  $k_t$ , the age of the youngest child in the household. In particular, the household's instantaneous utility function is given by:

$$u_t(c_t, P_t; k_t) = \frac{c_t^{1-\sigma}}{1-\sigma} - \left(\psi + \mathbb{I}_{\{k_t \le 6\}}\psi_k\right) P_t + \mathbb{I}_{\{k_t \le 18\}} \ln \chi$$

where  $\mathbb{I}_{\{k_t \leq 6\}}$  and  $\mathbb{I}_{\{k_t \leq 18\}}$  are indicators for the presence of a child under the ages of 6 or 18, respectively, in the household. This specification implies that children under 18 bring utility to his parents bit a child under 6 implies that households with a working mother incur a higher disutility from work ( $\psi + \psi_k$  versus  $\psi$  only).

### 3.2 Endowments

A key component of this model is the individual specific income process which depends on occupation his/her occupation  $\lambda_{ig} \in \{L, H\}$ . Given that the model does not include education or occupational decisions, the underlying assumption is that households in the model have already made all their education and occupation choices and that the permanent heterogeneity across agents is both a result of their innate ability and all their past choices.

Let the wage draw of an agent of gender g, in household i, at age t be denoted  $y_{igt}$ . The log wage process is assumed to be the sum of a gender specific intercept  $y_g$ , an age profile common to all agents  $p(t; \Theta)$ , an individual specific human capital component  $h_{igt}$  and a stochastic term  $z_{igt}$ :

$$\ln y_{igt} = \ln y_g + p(t; \Theta) + \ln h_{igt} + z_{igt}$$

$$h_{igt} = h_{ig,t-1} \exp[(1 - \delta d_{igt})\beta], \quad h_{ig0} = \exp(\alpha_{ig})$$

$$z_{igt} = \rho z_{ig,t-1} + \eta_{igt}, \quad \eta_{igt} \sim^{iid} N(0, \sigma_{\eta}^2)$$
(1)

An agent's human capital level at time t is given by  $h_{igt}$  where the terms  $\alpha_{ig}$  and  $\beta_{ig}$  capture individual-specific

age-depent human capital dynamics.<sup>9</sup> These parameters are jointly distributed  $(\alpha_{ig}, \beta_{ig}) \sim N(\mu, \Sigma)$  in the population, where  $\mu = (\mu_{\alpha}, \mu_{\beta})$  and

$$\Sigma = \begin{pmatrix} \sigma_{\alpha}^2 & \sigma_{\alpha\beta} \\ \\ \sigma_{\alpha\beta} & \sigma_{\beta}^2 \end{pmatrix}$$

Agents draw latent parameters ( $\alpha_{ig}, \beta_{ig}$ ) once in the beginning of t = 1 but they are never directly revealed. Following the setup in Guvenen (2007), agents must "learn" about both  $\alpha_{ig}, \beta_{ig}$  and  $z_{igt}$  in a Bayesian fashion.<sup>10</sup> The agent's draw of ( $\alpha_{ig}, \beta_{ig}$ ) will also determine his/her occupation. In particular, as shown in Figure 5, those agents who are endowed with  $\beta_{ig} \leq B$  will be assigned to the occupation with low levels of learning ( $\lambda_{ig} = L$ ) while those with  $\beta_{ig} > B$  are assigned to the occupation with high learning ( $\lambda_{ig} = H$ ).

Figure 5: Occupations and the distribution of  $\beta$ 



Given the evidence on the importance of work choices early in life (see, for example, Cox (1984) and Mincer and Polachek (1974)), I define  $d_{igt}$  to be an indicator which takes on the value of one if  $P_{ig\tau} = 0$ for any  $\tau \leq \min\{t, \bar{t}\}$ . Thus,  $\delta$  i the log wage equation captures the wage loss which is due to a career interruption which takes place before the age of  $\bar{t}$ . The assumption that  $\delta$  affects the individual's wages

<sup>&</sup>lt;sup>9</sup>Note that  $\ln h_{igt} = \alpha_{ig} + (1 - \delta d_{igt})\beta_{igt}$ .

<sup>&</sup>lt;sup>10</sup>Although the key learning dynamics in this model are all about  $\beta_{ig}$ , if agents were allowed to separately observe  $\alpha_{ig}$  and  $z_{igt}$ , the learning problem would become trivial. Note that I could have defined the learning process over only  $\beta_{ig}$  and  $z_{igt}$ , but it seems implausible that agents would have perfect information over  $\alpha_{ig}$  but be ignorant of  $\beta_{ig}$  and  $z_{igt}$ .

through the growth rate of human capital can be interpreted as career interruptions in one's past leading to lower human capital investment in the future. As argued by Weiss and Gronau (1981) and Weiss (1986), agents who anticipate career interruptions will undertake fewer activities which raise his earnings potential, for example, by investing less in human capital accumulation. Thus, agents who went through career interruptions in the past and have low attachment to the labor market, may opt to under-invest in human capital expecting future interruptions.

The stochastic portion of the income process,  $z_{it}$ , is modeled as an AR(1) process; this specification is standard in the literature. Since this is a model of a married household, a key assumption concerns the correlation structure between the husband and wife's error terms. Although many papers assume a positive correlation between the innovations to the AR(1) process in the spouses' incomes, it is not clear that such an assumption would be appropriate in this model (see, for example, Attanasio, Low, and Sanchez-Marcos (2008) and Heathcote, Storesletten, and Violante (2010)). Specifically, if  $corr(z_{imt}, z_{ift}) \neq 0$  then each of the agents in the household could, in theory, learn about their own  $z_{igt}$  by observing the wage draws and signals of their spouses. Since this does not appear to be a realistic assumption, I assume that the husband and wife's income shocks are ex-ante uncorrelated.<sup>11</sup>

#### The Learning Problem

In this section, I outline the learning process about the individual hidden Markov State  $\mathbf{S}_{t}^{ig} \equiv (\alpha_{ig}, \beta_{ig}, z_{igt})$ . First of all, within each occupation  $\lambda = L, H$ , there is a proportion of agents  $\Upsilon_{\lambda}$  who know their true  $\mathbf{S}_{t}^{ig}$  for all  $t \geq 1$ . For the remaining agents in each occupation, I follow Guvenen (2007), and assume that they do not know  $\mathbf{S}_{t}^{ig}$  but must instead learn about their individuals hidden state as they work. Since women's participation is endogenous in my model, I must also make further assumptions about the amount of information a woman gets before and after her work decision is made. Consistent with the timing used in most of the labor literature, I assume that husband and wife observe their respective wage draws in the beginning of each period. The household then updates its beliefs using those observations, and chooses the wife's participation.

 $<sup>^{11}</sup>$ I also solve the model under the assumption that the error terms of the husband and the wife are positively correlated but that agents do not learn about their own parameters from observing their spouse's wage draws and signals. The results of the paper are unchanged.

In any period t where the agent chooses to work, he/she observes  $s = 1, ..., S(\lambda_{ig})$  signals

$$x_{igt}(s) = \alpha_{ig} + \beta_{igt} + z_{igt} + \nu_{igs}, \quad \nu_{igs} \sim^{iid} \mathcal{N}(0, \sigma_{\nu}^2)$$

which can be interpreted as information about his/her true wage parameters which is only discovered upon working. The number of signals the agent receives depends on his/her occupation  $\lambda_{ig}$ . In particular, those agents in the high learning occupation ( $\lambda_{ig} = H$ ) receives  $S_H$  signals, while those agents in occupation  $\lambda_{ig} = L$  receive  $S_L$  signals, where I assume that  $S_L > S_H$ . One can interpret this assumption as capturing the fact that occupations with high levels of learning/training require the worker to master many distinct tasks while occupations with low learning involve only a few tasks which are performed repeatedly.<sup>12</sup> Thus, after the wife's participation decision is made, the household observes these signals and beliefs are updated again.

I assume that all agents know the set of parameters  $\Theta$  of the common  $p(t; \Theta)$  component, the gender specific intercept terms  $(y_m, y_f)$  and the parameters which characterize the distributions of  $(\alpha_{ig}, \beta_{ig})$ ,  $\eta_{igt}$ , and  $\nu_{igt}$ . In order to express the updating/learning problem as a hidden markov model with a continuous state, let  $\tilde{y}_{igt}$  denote wages which are free of the common age components and the gender specific intercept, i.e.  $\ln \tilde{y}_{igt} \equiv \ln y_{igt} - p(\Theta, t) - \ln y_g = \ln h_{igt} + z_{igt} = \alpha_{ig} + \beta_{ig}t + z_{igt}$ .

The equation for the evolution of the hidden Markov state is given by:

$$\mathbf{S}_{t+1}^{ig} \equiv \begin{bmatrix} \alpha_{ig} \\ \beta_{ig} \\ z_{ig,t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \rho \end{bmatrix} \begin{bmatrix} \alpha_{ig} \\ \beta_{ig} \\ z_{ig,t} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \eta_{igt} \end{bmatrix} = \mathbf{F}\mathbf{S}_{t}^{ig} + \mathbf{V}_{t}^{ig}$$
(2)

and also define  $\mathbf{Q}$  as the covariance matrix of  $\mathbf{V}_t^{ig}$ . Note that  $z_{igt}$  is the only dynamic target that the agents are learning about and although  $\alpha_{ig}$  and  $\beta_{ig}$  are fixed for each agent, they are still part of the underlying unknown state. Given that individuals observe both income draws  $y_{igt}$  and signals  $x_{igs}(t)$  which reveal information about their state  $\mathbf{S}_t^{ig}$ , I next define the observation equations for each one of

<sup>&</sup>lt;sup>12</sup>A simple example with 2 agents makes this interpretation clear. There are two working agents and each receives one signal a day. Agent H who works in the high-learning occupation must master K tasks and his overall productivity  $\beta_H$  is given by max $\{\beta_1, \beta_2, ..., \beta_K\}$  where each  $\beta_k$  denotes his productivity in that specific task. His daily signal concerns only his productivity in one specific task. On the other hand, agent L who works in the low-learning occupation only has to master 1 task and his productivity  $\beta_L$  is given by his productivity at that task. From this simple example one can see that if agent L learns at rate l, then agent H will learn at rate l/K.

them. Let the observation equation for the log of the wage draw  $\tilde{y}_{igt}$  be given by:

$$\ln \tilde{y}_{igt} = \begin{bmatrix} 1 & t & 1 \end{bmatrix} \begin{bmatrix} \alpha_{ig} \\ \beta_{ig} \\ z_{igt} \end{bmatrix} = \mathbf{H}'_t \mathbf{S}^{ig}_t$$
(3)

and the one for each signal  $x_{igs}(t)$ ,  $s = 1, ..., S(\lambda_{ig})$  be given by:

$$x_{igt}(s) = \begin{bmatrix} 1 & t & 1 \end{bmatrix} \begin{bmatrix} \alpha_{ig} \\ \beta_{ig} \\ z_{igt} \end{bmatrix} + \nu_{igs} = \mathbf{H}'_t \mathbf{S}^{ig}_t + \nu_{igs}$$
(4)

An individual's initial uncertainty about his/her parameters is modelled as an initial prior over  $(\alpha_{ig}, \beta_{ig}, z_{ig1})$  given by a multivariate normal distribution with mean  $\widehat{\mathbf{S}}_{1|0}^{ig}$  and variance  $\mathbf{P}_{1|0}$ , defined as:

$$\widehat{\mathbf{S}}_{1|0}^{ig} = (\begin{array}{cc} \widehat{\alpha}_{1|0}^{ig} & \widehat{\beta}_{1|0}^{ig} & \widehat{z}_{1|0}^{ig} \end{array})'$$

$$\mathbf{P}_{1|0} = \begin{bmatrix} \sigma_{\alpha(1|0)}^2 & \sigma_{\alpha\beta(1|0)} & 0 \\ \sigma_{\alpha\beta(1|0)} & \sigma_{\beta(1|0)}^2 & 0 \\ 0 & 0 & \sigma_{z(1|0)}^2 \end{bmatrix}$$

where the standard notation  $\widehat{\mathbf{S}}_{t'|t}^{ig}$  denotes the beliefs about  $\mathbf{S}_{t}^{ig}$  of an individual in household *i*, of gender *g*, at time *t'*, given all the available information at time *t*, where  $t' \geq t$ . Thus, at time *t*, individual *i* has beliefs about the unobserved state  $\mathbf{S}_{t}^{ig}$  which have a multivariate normal posterior distribution with mean  $\widehat{\mathbf{S}}_{t|t}^{ig}$  and covariance matrix  $\mathbf{P}_{t|t}$ .

An agent who comes into a given period with beliefs  $\widehat{\mathbf{S}}_{t|t-1}^{ig}$ , upon observing the individual wage draw

 $y_{igt}$ , updates his beliefs about his wage parameters using the formulas:<sup>13</sup>

$$\widehat{\mathbf{S}}_{t|t^*}^{ig} = \widehat{\mathbf{S}}_{t|t-1}^{ig} + \mathbf{P}_{t|t-1}\mathbf{H}_t[\mathbf{H}_t'\mathbf{P}_{t|t-1}\mathbf{H}_t + \sigma_{\varepsilon}^2]^{-1}(\ln \widetilde{y}_{igt} - \mathbf{H}_t'\widehat{\mathbf{S}}_{t|t-1}^{ig})$$
(5)

$$\mathbf{P}_{t|t^*} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{H}_t [\mathbf{H}_t' \mathbf{P}_{t|t-1} \mathbf{H}_t + \sigma_{\varepsilon}^2]^{-1} \mathbf{H}_t' \mathbf{P}_{t|t-1}$$
(6)

where  $t^*$  denotes the information set at time t after the agent observes the wage draw  $y_{igt}$ . If the agent works, then he/she also observes  $S(\lambda_{iq})$  signals  $x_{iqs}(t)$  and the updating formulas for each is given by:

$$\widehat{\mathbf{S}}_{t|t_s}^{ig} = \widehat{\mathbf{S}}_{t|t_{s-1}}^{ig} + \mathbf{P}_{t|t_{s-1}}\mathbf{H}_t[\mathbf{H}_t'\mathbf{P}_{t|t_{s-1}}\mathbf{H}_t + \sigma_\nu^2]^{-1}(x_{igt}(s) - \mathbf{H}_t'\widehat{\mathbf{S}}_{t|t_{s-1}}^{ig})$$
(7)

$$\mathbf{P}_{t|t_{s}} = \mathbf{P}_{t|t_{s-1}} - \mathbf{P}_{t|t_{s-1}} \mathbf{H}_{t} [\mathbf{H}_{t}' \mathbf{P}_{t|t_{s-1}} \mathbf{H}_{t} + \sigma_{\nu}^{2}]^{-1} \mathbf{H}_{t}' \mathbf{P}_{t|t_{s-1}}$$
(8)

where  $t_s$  denotes the agent's information set at time t after observing  $y_{igt}$  and s signals  $x_{igt}(s)$  and  $\widehat{\mathbf{S}}_{t|t}^{igg}$ denotes the beliefs of the agent at the end of time t, given all the available information at time t. Thus, for the agent who choses not to work,  $\widehat{\mathbf{S}}_{t|t}^{ig} = \widehat{\mathbf{S}}_{t|t^*}^{ig}$  and  $\mathbf{P}_{t|t} = \mathbf{P}_{t|t^*}$ . Note that the covariance matrix is not individual specific, it does not depend on  $y_{igt}$  or  $x_{igt}(s)$ , and it evolves deterministically.<sup>14</sup> Furthermore, the posterior variances of  $(\alpha_{ig}, \beta_{ig})$  are monotonically decreasing over time, and thus beliefs concentrate more around their true values after every update.<sup>15</sup>

In order to solve the household's problem, one must also define next period's value function which depends on the distribution of income draws tomorrow, conditional on beliefs today. The one-period-ahead forecasts of the hidden state are given by:

$$\begin{split} \widehat{\mathbf{S}}_{t+1|t}^{ig} &= \mathbf{F} \widehat{\mathbf{S}}_{t|t}^{ig} \\ \mathbf{P}_{t+1|t} &= \mathbf{F} \mathbf{P}_{t|t} \mathbf{F}' + \mathbf{Q} \end{split}$$

and thus conditional on  $\widehat{\mathbf{S}}_{t|t}^{ig}$  and  $\mathbf{P}_{t|t}$ ,  $\ln \widetilde{y}_{ig,t+1}$  is distributed:

$$\ln \tilde{y}_{ig,t+1} | \widehat{\mathbf{S}}_{t|t}^{ig} \sim \mathrm{N}(\mathbf{H}_{t+1}' \widehat{\mathbf{S}}_{t+1|t}^{ig}, \ \mathbf{H}_{t+1}' \mathbf{P}_{t+1|t} \mathbf{H}_{t+1} + \sigma_{\varepsilon}^2)$$
(9)

 $<sup>^{13}</sup>$ Since this is a hidden Markov model with a continuous state space and a gaussian assumption on the error terms, all the results and updating steps are simply given by the equations of a Kalman Filter. See Hamilton (1994) for a complete derivation of these formulas.

<sup>&</sup>lt;sup>14</sup>Note that although  $\mathbf{P}_{t|t}$  can be calculated deterministically without observing  $x_{igs}(t)$ , women cannot "cheat" and update  $\mathbf{P}_{t|t}$  using equation (8) in periods they choose not to work. <sup>15</sup>Note that this is not true for the posterior variance of  $z_{igt}$ .

#### 3.3 Problem of the Household

Before I explain the problem solved by the household each period, it will be instructive to outline the timing assumptions within a given period and the state vector of the household. For ease of notation, I omit all household subscripts *i* in this section. As one can see from the timeline in Figure 6, both wage draws  $y_{mt}$  and  $y_{ft}$  are revealed to the household at the beginning of the period. Beliefs are then updated to the  $t^*$  stage, and the wife's participations decision is made. If she works, the vector of signals for the wife  $X_{ft} = \{x_{ft}(s) : s = 1, ..., S(\lambda_f)\}$  is observed, together with the signals for the husband  $X_{mt} = \{x_{mt}(s) : s = 1, ..., S(\lambda_m)\}$ . Beliefs are then updated once again the household solves the consumption-savings problem and makes the fertility decision (if applicable). Households in my model are allowed to save but not borrow.<sup>16</sup>

#### Figure 6: Within period timing



Define two state vectors for the household problem:

$$\Omega_t = \left\{ \lambda_m, \ \lambda_f, \ k_t, \ d_t, \ y_{mt}, \ y_{ft}, \ \widehat{\mathbf{S}}_{t|t}^m, \ \widehat{\mathbf{S}}_{t|t}^f, \ a_t, \ \mathbf{X}_{mt}, \ \mathbf{X}_{ft} \right\}$$
(10)
$$\Omega'_t = \left\{ \lambda_m, \ \lambda_f, \ k_t, \ d_t, \ y_{mt}, \ y_{ft}, \ \widehat{\mathbf{S}}_{t|t}^m, \ \widehat{\mathbf{S}}_{t|t}^f, \ a_t, \ \mathbf{X}_{mt} \right\}$$

where  $\Omega_t$  is the household state which includes the vector of signals for the wife  $X_{ft}$  while  $\Omega'_t$  is the state vector for a household who has not (yet) observed signals  $X_{ft}$ . I begin by describing the consumption-savings problem at time t for a household that had children at time  $\tau \leq t$ , both in the case when the wife works and when she does not. Then, I proceed to outline the wife's participation problem and finally, the fertility decision.

<sup>&</sup>lt;sup>16</sup>All the results in the paper remain in the version of the model where I allow household to borrow up to a realistic ad-hoc amount. This specification, however, induced a second non-concavity which lead to numerical inaccuracies in some areas of the parameter space. For this reason, I choose instead the no-borrowing specification as my benchmark. See the Appendix for details on the model with borrowing.

#### **Consumption-Savings Decision**

A couple who received their first child at time  $\tau \leq t$  and whose wife has chosen to work at time t solves:

$$\mathbb{V}_{t}^{w}\left(\Omega_{t}'\right) = \max_{c_{t}, a_{t+1}} u(c_{t}, P_{t} = 1; k_{t}) + \gamma E_{t} \left[\mathbb{V}_{t+1}\left(\Omega_{t+1} \mid P_{t} = 1, \widehat{\mathbf{S}}_{t|t-1}^{m}, \widehat{\mathbf{S}}_{t|t-1}^{f}\right)\right] \\
\text{s.t.} \ \widehat{c}_{t} + a_{t+1} = (1+r)a_{t} + y_{mt} + y_{ft} - \kappa \mathbb{I}_{\{k_{t} \le 6\}} \\
a_{T+1} \ge 0 \\
c_{t} = \frac{\widehat{c}_{t}}{e(k_{t})}$$
(11)

and Bayesian updating using equations (7) and (8) for  $X_{mt}$ ,  $X_{ft}$ 

where income consists of capital income from last period's assets  $a_t$  with return r, and the labor earnings of the husband and wife. Childcare costs  $\kappa$  must be paid if the wife works in a household where the youngest child is under the age of 6. Consumption expenditures  $\hat{c}_t$  are deflated by the McClements scale  $e(k_t)$  which is a function of the number of people in the household and their ages. Note that once the wife has decided to work, the household's state vector includes both the signals  $X_{mt}$  for the husband and  $X_{ft}$  for the wife.

If the wife has chosen to not work at time t, then after updating beliefs, the household solves:

$$\mathbb{V}_{t}^{w}\left(\Omega_{t}^{\prime}\right) = \max_{c_{t}, a_{t+1}} u(c_{t}, P_{t} = 0; k_{t}) + \gamma E_{t} \left[\mathbb{V}_{t+1}\left(\Omega_{t+1} \mid P_{t} = 0, \widehat{\mathbf{S}}_{t|t-1}^{m}, \widehat{\mathbf{S}}_{t|t-1}^{f}\right)\right]$$
s.t.  $\widehat{c}_{t} + a_{t+1} = (1+r)a_{t} + y_{mt}$ 

$$a_{T+1} \ge 0$$

$$c_{t} = \frac{\widehat{c}_{t}}{e(k_{t})}$$
(12)

and Bayesian updating using equations (7) and (8) for  $X_{mt}$ 

and 
$$\widehat{\mathbf{S}}_{t|t}^f = \widehat{\mathbf{S}}_{t|t^*}^f$$
 and  $\mathbf{P}_{t|t}^f = \mathbf{P}_{t|t^*}^f$ 

where the wife does not observe the signals  $X_{ft}$  and thus does not further update beliefs  $\widehat{\mathbf{S}}_{t|t^*}^f$ . The expected values in both the working and non-working optimization problems are taken over the income draws for the husband and wife,  $y_{m,t+1}, y_{f,t+1}$ , whose conditional distributions are given by equation (9).

#### Participation Decision of the Wife

Given that I've defined the household consumption-savings problem conditional on whether the wife participates or not in the labor market, I now proceed to define the participation problem. After observing wages  $y_{mt}, y_{ft}$ , the decision for whether the wife participates in period t is nade by comparing  $E_t [\mathbb{V}_t^w (\Omega_t)]$ to  $\mathbb{V}_t^n (\Omega_t')$ , where the expectation if taken over  $X_{mt}$  and  $X_{ft}$ . In other words, the participation decision is the solution to:

$$\mathbb{V}_t\left(\Omega_t'\right) = \max\left\{E_t\left[\mathbb{V}_t^w\left(\Omega_t \mid P_t = 1\right)\right], \quad \mathbb{V}_t^n\left(\Omega_t' \mid P_t = 0\right)\right\}$$
(13)  
with Bayesian updating of beliefs using  $y_{mt}, y_{ft}$ 

Note that the participation decision is made after beliefs  $\widehat{\mathbf{S}}_{t|t-1}^m$  and  $\widehat{\mathbf{S}}_{t|t-1}^f$  have been updated into  $\widehat{\mathbf{S}}_{t|t^*}^m$ and  $\widehat{\mathbf{S}}_{t|t^*}^f$  using equation (5).<sup>17</sup>

#### **Fertility Decision**

For a couple entering time  $t \leq t_F$  without children  $(k_t = 0)$ , the household consumption-savings problem, conditional on whether the wife works or not, is defined the same way as in equations (11) and (12). However, the continuation value  $\mathbb{V}_{t+1}$  is instead given by:

$$\mathbb{V}_{t+1}(\Omega'_{t+1}) = \max\left\{\mathbb{V}^{P}_{t+1}, \mathbb{V}^{NP}_{t+1}\right\}$$
(14)

$$= \max \left\{ \begin{array}{c} \mathbf{p}_{t} \mathbb{V}_{t+1}(\Omega'_{t+1} \mid k_{t+1} = 1) + (1 - \mathbf{p}_{t}) \mathbb{V}_{t+1}(\Omega'_{t+1} \mid k_{t+1} = 0), \\ \mathbb{V}_{t+1}(\Omega'_{t+1} \mid k_{t+1} = 0) \end{array} \right\}$$
(15)

where each value function  $\mathbb{V}_{t+1}$  is defined as in equation (13). If the couple chooses to become parents  $(\mathbb{V}_{t+1} = \mathbb{V}_{t+1}^P)$ , with probability  $p_t$ , they receive a child the period after and a second child two years after the first. The household  $k_t$  then becomes  $k_{\tau} > 0$ ,  $\forall \tau > t$  and they will not make any fertility decisions thereonafter. With probability  $1 - p_t$  however, the couple who chooses to become parents will not receive any children and  $k_{t+1} = 0$ . In this case, the couple could possibly choose to become parents again at a future  $\tau$ , as long as  $\tau \leq t_F$ . If the household chooses to not become parents this period then  $\mathbb{V}_{t+1} = \mathbb{V}_{t+1}^{NP}$ 

<sup>&</sup>lt;sup>17</sup>The discrete participation decision induces a non-concavity which can be smoothed out in the presence of enough uncertainty. I numerically check that this is the case. Full details on the computation are in the Appendix.

and  $k_{t+1} = 0$ .

### Retirement

During retirement, the couples' problem is greatly simplified since there are no more participation or fertility decisions, the pension income is constant and learning about income parameters no longer takes place. The household problem can thus be written as as simple consumption-savings problem:

$$\mathbb{V}_{t}(\mathbf{b}, \ a_{t+1}) = \max_{c_{t}, a_{t+1}} u(c_{t}; k_{t}) + \gamma \mathbb{V}_{t+1}(\mathbf{b}, \ a_{t+1})$$
s.t.  $\hat{c}_{t} + a_{t+1} = Ra_{t} + \mathbf{b}$ 

$$a_{T+1} \ge 0 \qquad (16)$$

$$c_{t} = \frac{\hat{c}_{t}}{e(k_{t})}$$

for all periods  $t \ge t^R$  and terminal condition  $\mathbb{V}_{T+1} \equiv 0$ . The household's retirement benefits are constant and given by b.

## 4 Parametrization

In this section, I begin by discussing the set of parameters which will be set externally. These include some which are taken directly from preexisting estimates in the literature and others which will be estimated directly from data. Then, I proceed to outline the set of parameters which will be calibrated internally such that the model matches certain statistics in the NLSY79.

#### 4.1 Parameters Set Externally

#### **Income Process**

Recall that my model assumes that men and women face the same stochastic process for wages. Thus in order to bypass the selection problem with women's wages in the data, I choose to estimate these from wage data on male heads of household only. I use the 1968-1997 waves of PSID on earnings and hours worked for male household heads between the ages of 23 and 64.<sup>18</sup> In order to be included in the sample,

<sup>&</sup>lt;sup>18</sup>I restrict the PSID sample to the years in which it is of annual frequency.

the individual must be observed in at least 20 (not necessarily consecutive) years with i) positive earnings and hours, ii) between 520 and 5110 hours worked a year and iii) average hourly earning within certain bounds. The criteria is similar to the one followed in various other papers in the literature (see, for example, Abowd and Card (1989) and Heathcote, Storesletten, and Violante (2010)). Further sample selection details are outlined in the Appendix.

 Table 3: Estimated parameters for the age polynomial

		0	* 4
Estimate 0.35	55 -0.010	1.3e-4	-6.7e-7
Std. Error (0.04	(3) (0.002)	(2.4e-5)	(1.4e-7)

Following the procedure of Guvenen (2009), I first I regress observed log hourly wages for an individual j, observed at time t on a fourth degree polynomial on age:

$$\ln w_{jt} = \theta_0 + \theta_1 age_{jt} + \theta_2 age_{jt}^2 + \theta_3 age_{jt}^3 + \theta_4 age_{jt}^4 + \tilde{w}_{jt}$$

in order to obtain an estimate for the parameters of the common age component  $p(t; \Theta) \equiv \theta_1 t + \theta_2 t^2 + \theta_3 t^3 + \theta_4 t^4$ . These estimates and their standard errors are reported in Table 3. Given my assumptions on the functional form of low wages, I use the residuals  $\tilde{w}_{jt}$  from this regression to estimate the parameters which govern the distribution of unobserved individual wage parameters ( $\sigma_{\alpha}, \sigma_{\beta}, \sigma_{\alpha\beta}$ ) and the parameters of the stochastic component of wages ( $\sigma_{\nu}, \rho$ ). Thus I assume that the regression residuals  $\tilde{w}_{jt}$  take on the form:

$$\begin{split} \tilde{w}_{jt} &= \alpha_j + \beta_j \ age_{jt} + z_{jt} + \varepsilon_{jt} \\ z_{jt} &= \rho z_{j-1,t-1} + \eta_{jt} \\ \eta_{jt} \sim^{iid} \mathcal{N}(0, \sigma_{\nu}^2), \quad \varepsilon_{jt} \sim^{iid} \mathcal{N}(0, \sigma_{\varepsilon}^2) \end{split}$$

where  $\varepsilon$  is assumed to be iid measurement error. I estimate the parameters using the minimum distance estimator first proposed by Chamberlain (1984). This method seeks parameters which minimize the distance between the empirical covariance matrix of income residuals and the one obtained from simulating the income process outlined above. This choice of estimator is standard in the literature and its use and identification in this specific income process is described in detail in Guvenen (2009).

The results of this estimation are reported in Table 4. Following much of the literature, I allow for heteroskedastic innovations to  $z_t$  in the estimation process while assuming that the variance of the measurement error is invariant across time. The reported value for  $\sigma_{\eta}$  in Table 4 is an average of the estimated parameters for the 1983-1997. I choose these years because they are the most pertinent for my cohort of interest, when they are between the median ages of 23 and 37. Lastly, note that although this estimation process identifies the covariance structure of  $\alpha$  and  $\beta$ , it does not identify theirs means. Thus, I normalize  $\mu_{\alpha} = 1.5$  and set  $\mu_{\beta}$  to be the average growth of log income observed in my NLSY79 sample, 1.2% per year.

	-	Wage Profiles		Stochastic Process	
	$\sigma_{lpha}^2$	$\sigma_{eta}^2$	$\sigma_{lphaeta}$	$\rho$	$\sigma_\eta^2$
Estimate Std. Error	$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{c} 0.00031 \\ (0.00006) \end{array}$	-0.0025 (0.0028)	$\begin{array}{c} 0.809 \\ (0.022) \end{array}$	$0.045 \\ (0.005)$

Table 4: Estimated parameters for the residual wage process

After retirement, households in the model receive a constant pension which is a function of the husband's last observed income  $y_{mt_R}$ . The exact functional form follows the specification in Heathcote, Storesletten, and Violante (2010), and it mimics the US Social Security bendpoints.<sup>19</sup> Since I do not keep track of the wife's last *observed* wage, the household receives 1.5 times the husband's pension as calculated by the Social Security formula.<sup>20</sup>

<sup>&</sup>lt;sup>19</sup>To compute retirement benefits for a model household, I modify the approach used in Heathcote, Storesletten, and Violante (2010) in order to avoid keeping track of an individual's average earnings over their lifecycle. More specifically, I take the husband's last observed earnings  $y_{mt_R}$  and compute social security benefits as follows: 90% of  $y_{mt_R}$  up to a first threshold equal to  $0.38\bar{y}_{mt_R}$  where  $\bar{y}_{mt_R}$  is the average observed earnings in the economy, plus 32% of  $y_{mt_R}$  from this bendpoint to a higher bendpoint equal to  $1.59\bar{y}_{mt_R}$ , plus 15 percent of the remaining  $y_{mt_R}$  exceeding this last bendpoint.

<sup>&</sup>lt;sup>20</sup>Keeping track of the last observed wage for the wife would allow retirement benefits to also be a function of those earnings. This channel may be important for female labor supply, since this implies that her work decisions and observed wages will have an impact on retirement income. Doing so, however, requires one more state variable which, in earlier experiments proved to be too computationally time consuming. The importance of this channel remains open and I leave its examination to future research. Since the US social security rules imply that a household receives the maximum of either 1.5 times the husband's pension or the sum of the pensions of the husband and wife, another concern is whether my calibrated pension *levels* might be too low. In order to exclude these possible effects in robustness checks I experimented with giving households different proportions of the husband's pension. Even without recalibrating, the changes in any of the model results are negligible. For example, if instead the household were given twice the husband's pension, female LFP in the benchmark economy would drop by 0.2 percentage points.

#### Other external parameters

The model period is one year. Agents enter the the working stage at age 23 (t = 1) as a married couple. Both individuals of a household retire at age 62, which implies  $t_R = 39$  and die with certainty at the end of age 92 (T = 67). Given that estimates of relative risk aversion in the micro literature range between one and two (see Attanasio (1999) for a survey), I set  $\sigma = 1.5$ . The interest rate is chosen to be 3% a year, following Heathcote, Storesletten, and Violante (2010).

Households can only choose to become parents before the age of 40; thus  $t_F = 18$ . This is not a very restrictive assumption given the evidence presented about high infertility after the age of 40. Also, in the data, less than half a percent of women have children beyond this age. The probability of not receiving a child if the couple decides to become parents is set to 0 before the age of 30 ( $p_t = 0$ ,  $\forall t \leq 7$ ), and the values of  $p_t$  afterwards are given by the probability of conception presented in van Noord-Zaadstra et al. (1991) (graph reproduced earlier in this paper), regardless of the child's health. Children deflate household consumption expenditures according the McClements scale. This scale is within the range of the economies of scale in household consumption estimated by Fernandez-Villaverde and Krueger (2007) using data from the Consumer Expenditure Survey (CEX).

Another parameter set externally is B which governs the proportion of agents in low versus high-learning occupations. In the model, all those agents with  $\beta_{ig} \leq B$  are assigned to the low learning occupation while those with  $\beta_{ig} > B$  are assigned to the high learning occupation. I set B so that 79.8% of men in my model work in the occupation with low learning and 20.2% work in the occupation with high learning these are the proportions in the NLSY79 data. Finally, recall that agents are born with priors over their hidden state  $\widehat{\mathbf{S}}_{t}^{ig}$  defined by a multivariate normal distribution with mean  $\widehat{\mathbf{S}}_{1|0}^{ig}$  and variance  $\mathbf{P}_{1|0}$ . For an agent with occupation  $\lambda$ , these will be given by  $\widehat{\mathbf{S}}_{1|0}^{ig}(\lambda) = [\bar{\alpha}(\lambda), \ \bar{\beta}(\lambda), \ 0]'$ 

$$\mathbf{P}_{1|0} = \begin{bmatrix} \sigma_{\alpha}^2 & \sigma_{\alpha\beta} & 0 \\ \sigma_{\alpha\beta} & \sigma_{\beta}^2 & 0 \\ 0 & 0 & \sigma_{\nu}^2 \end{bmatrix}$$

where  $\mathbf{P}_{1|0}$  is simply given by the population variances estimated from the data and  $\bar{\alpha}(\lambda), \bar{\beta}(\lambda)$  are the average  $\alpha$  and  $\beta$  in each occupation  $\lambda \in \{L, H\}$ . Note that all the agents within an occupation have the same prior over their wage parameters. I do this in order to isolate the effects of different wage

draws/signals from the effects of different priors on the evolution of beliefs.<sup>21</sup> Since agents know their own occupation, my assumption that  $\mathbf{P}_{1|0}$  is equal to the population variances implies that agents' uncertainty about their own  $\alpha$  and  $\beta$  is not reduced by observing their occupation. This is not an innocuous assumption for the amount of uncertainty that households face may possibly have effects on the role of the wife's labor supply as a channel of insurance. However, as shown by Attanasio, Low, and Sanchez-Marcos (2005), these effects on female participation are very small and mostly focused on the later years of the life-cycle (after age 40) when  $\mathbf{P}_{1|0}$  has been greatly reduced for most of the agents in the model, namely for all husbands who work every period.

### 4.2 Parameters Set Internally

The last set of parameters will be set internally such that the model reproduces certain statistics from the NLSY79 data. This set includes those parameters that govern the disutility from work, utility from children and childcare costs, some that affect wage levels and dynamics and others that define agents' speed of learning. While all parameters are codetermined in the model, one can draw an approximate mapping between each set of calibrated parameters and their "closest" targets.

Utility and Fertility Parameters: The parameters which govern the work disutility of the wife ( $\psi$  and  $\psi_k$ ), the utility derived from the presence of children under 18 in the household ( $\chi$ ) and childcare costs ( $\kappa$ ) are calibrated so that the model generates:

- (i) The average participation rate of women between the ages of 25-45
- (ii) The average participation rate of mothers with children under the age of 6
- (iii) The proportion of women who never had children by the age of 40
- (iv) The proportion of mothers who give birth before the age of 24

**Discount factor:** As it is usual in the literature (see, for example, Storesletten, Telmer, and Yaron (2004)), the discount factor  $\gamma$  will be set such that the model reproduces the ratio of average wealth to average pre-tax earnings in the data, 3.94. This is the value calculated by Heathcote, Storesletten, and Violante (2010) using the 1992 Survey of consumer finances for households up to the 99th percentile of wealth. This yields a value of 0.981 for  $\gamma$ .

<sup>&</sup>lt;sup>21</sup>In the Appendix I discuss the results from a robustness check in which I give agents different priors over their wage parameters. All the model results remain the same.

Parameters			Value
Disutility from labor		$\psi$	0.090
Added disutility from labor		$\psi_{k}$	0.049
Childcare		$\kappa$	4.298
Utility from children		$\chi$	1.155
Discount factor		$\gamma$	0.981
Scale of female to male wages		$y_{0}^{f}/y_{0}^{m}$	0.774
Wage growth penalty for early career interruptions		δ	0.046
Variance of the noise to signal		$\sigma_{ u}^2$	0.097
Number of signals received by low learning occupation		$S_L$	62
Proportion with perfect info	Low learning	$\Upsilon_L$	12.29
	High learning	$\Upsilon_H$	3.84
Targets		Model	Data
Participation of women 25-45		82.37	83.74
Participation of mothers with children $\leq 6$		63.44	61.97
Proportion of non-mothers		11.31	11.27
Proportion of mothers by age 24		13.15	15.95
Wealth-to-Income ratio up to 99th ptile 3.45		3.94	
Observed gender wage ratio, ages 25-45		0.79	0.81
Motherhood wage penalty		0.17	0.17
Low learning group	Median SVP	1	1
	Standard Deviation	20.12	19.57
	Median SVP	10	10
High learning group	Standard Deviation	10.25	9.84

Table 5: Internally Calibrated Parameters

Wage parameters: Both male and female wages include a time-invariant gender-specific intercept term  $y_g$ . Their ratio  $y_f/y_m$  will be calibrated such that the model reproduces the ratio of average female to average male wages observed in the data for those between the ages of 25-40 in my sample (0.81). Recall that women who interrupt their careers during the first  $\bar{t}$  years of work suffer a penalty  $\delta$  to their wage growth. I set  $\bar{t} = 15$ , in order to reflect the importance of early career decisions on wages. The penalty  $\delta$  will be calibrated such that in the model, the women who have  $P_{\tau} = 0$  for any  $\tau \leq 15$  have observed wages upon return which are, on average, 17% lower than they would have been had they not quit the labor market. This is the value of the estimated wage gap in my sample between women who work continuously and those who have at least one career interruption during the first 15 years after they have completed their education.

Information Parameters: The last set of internally calibrated parameters govern aspects of the structure of agents' information acquisition. These parameters are the proportion of agents in each occupation who know their true wage parameters  $\alpha, \beta$  and  $z_t$  for all t ( $\Upsilon_L, \Upsilon_H$ ), the number of signals received upon working  $(S_L, S_H)$ , and the variance of the signal noise  $(\sigma_{\nu})$ , which is common across occupations. — proceed in the following way. First, I normalize the number of signals received by those in the high learning occupation to one (i.e.  $S_H = 1$ ). I then internally calibrate  $\sigma_{\nu}^2$ ,  $S_L$ ,  $\Upsilon_L$  and  $\Upsilon_H$  such that the median and standard deviation of the number of periods that an agent in each occupation must work before he/she learns his/her true  $\beta$  are equal to the median and standard deviation of SVP observed in the data for those in occupations in the low and high learning groups, respectively.

The implied values for the internally calibrated parameters are shown in the top panel of Table 5. The values of the targeted statistics in the model and in the data are shown in the bottom panel of the same table. As one can see, the model does quite a good job matching most of the target statistics from the data. However, the model's wealth-to-income ratio is somewhat lower than what is observed in the data. This is mainly due to the fact that there is no uncertainty in the model during the retirement stage while individuals in the data face significant health related risk in that period.<sup>22</sup>

The calibrated value of  $y_0^f/y_0^m$  is lower than the observed gender wage ratio, which implies that, on average, women's selection into work is positive in the model. The calibrated parameter for  $\psi$ , the utility cost from the woman's participation, is equivalent to 35.9% of consumption. For those households with at least one child under the age of 6, the same statistic is 57.8% of consumption. Although this value may seem large, children bring significant utility for the household, equivalent to 53.9% of increased in per period consumption, for every period that there is at least one child under the age of 18 in the household.

The calibrated value for childcare costs is equivalent to 31.1% of the average female wages in the model during the ages of 23-32. These childcare costs are less than half of those implied by the calibrated model of Attanasio, Low, and Sanchez-Marcos (2008). The reasons for this are twofold: first, because those authors target the participation rates of mothers with children under 3, their participation target for

<sup>&</sup>lt;sup>22</sup>Introducing health shocks into the model's retirement stage would increase the agent's willingness to accumulate assets well into old age. This is commonly done in the literature which focuses on later life decisions (see, for example, De Nardi, French, and Jones (2010)).

mothers is much lower than the one I used (47% versus 62% in mine). Second, unlike Attanasio, Low, and Sanchez-Marcos (2008), my model explicitly includes an added disutility from labor if the woman has a child under the age of 6 at home. Thus my model relies on a combination of monetary and psychic costs in order to generate the low participation of mothers.

### 4.3 Quantifying income and learning

Although agents are heterogeneous in both  $\alpha$  and  $\beta$ , the estimated values for these parameters imply that the key determinant of an agent's position in the wage distribution for most of his life is his value of  $\beta$ . This coupled with the fact that the estimated  $\sigma_{\alpha\beta} < 0$ , implies that although a steep wage profile may begin at a level which is relatively lower, the growth dynamics take over a few years into one's career. A consequence of this is that by age 45, when most life-cycle wage profiles reach their peak, the agents at the top of the distribution are those who had higher wage growth, and not necessarily those who had higher intercepts.

To demonstrate this in my model, I plot the log of  $\tilde{y}_{igt}$  paths for 9 possible combinations of  $(\alpha, \beta)$ , where  $\alpha \in \{\alpha_{25}, \alpha_{50}, \alpha_{75}\}$  and  $\beta \in \{\beta_{25}, \beta_{50}, \beta_{75}\}$ , the 25th, 50th and 75th percentile of those parameters, respectively. These wage paths are shown in Figure 7. The triangles denote  $\alpha_{75}$ , circles  $\alpha_{50}$  and squares  $\alpha_{25}$ . The dark set of dashed lines are those with  $\beta_{75}$ , while the dark set of solid lines plots wages for those with  $\beta_{25}$ . The set of lighter lines in the middle are wage paths with  $\beta_{50}$ .

As one can see from the figure, during the first few years in life until age 26, the set of triangle lines  $(\alpha_{75})$  always lie on at the top of the wage distribution. Thus, the  $\alpha$  parameter is key in determining the agent's position in the distribution of wages early in life. Between the ages of 26 and 35, however, the dark dashed lines start moving towards the top of the wage distribution. After 35 those paths with  $\beta_{75}$  (dark dashed lines) lie above all others, independently of their  $\alpha$ . Furthermore, this pattern is repeated for the wage paths with  $\beta_{50}$ , which always lie above those with  $\beta_{25}$ . Note that  $\alpha$  is not completely inconsequential; within each  $\beta$  level, wages are monotonically increasing with  $\alpha$ . To better understand this, consider an agent with low growth but high intercept ( $\alpha_{75}, \beta_{25}$ ). The estimated parameters for the income process



Figure 7: Log wage paths for different combinations of  $\alpha, \beta$ 

imply that  $\beta_L \approx -\frac{1}{28} \alpha_H$ . The expected value of  $\tilde{y}_t$  conditional on the agent's true parameters is given by:

$$E(\ln \tilde{y}_t | \alpha_H, \beta_L) = \alpha + \beta age$$

$$\approx \alpha_H - \left(\frac{1}{28}\alpha_H\right) age \tag{17}$$

From the equation (17), one can see that for all age > 28, the second term  $\beta_L = -\frac{1}{28}\alpha_H$  dominates the expected value. Thus, conditional on the true parameters, most of late life wages will be determined by the agent's  $\beta$ .<sup>23</sup>

In view of its importance in the determination of life-cycle wages, one may want to ask how quickly do agents in the high-learning occupation learn about their own  $\beta$ .<sup>24</sup> It turns out that learning in this setting takes place very gradually. Even though working agents in high-learning occupations update their beliefs twice every period (once from the wage draw  $y_t$  and again from the noisy signal  $x_t(1)$ ), learning may take much longer than the calibrated median length of 10 years. A key reason for this is the fact that agents are learning about a growth parameter whose contribution to wages is very low in the beginning of life (as can be inferred from Figure 7) but growing over time. At later ages, however, the moderate

 $<sup>2^{3}</sup>$  The same can be seen for an agent with high growth but low intercept  $(\alpha_{25}, \beta_{75})$ , where  $\beta_H \approx \frac{1}{14} \alpha_L$ . Since  $E(\ln \tilde{y}_t | \alpha_L, \beta_H) \approx \alpha_H + (\frac{1}{14} \alpha_H) age$ , for all age > 14, the growth term  $\beta_H \approx (\frac{1}{14} \alpha_H)$  dominates the conditional expected value.

<sup>&</sup>lt;sup>24</sup>Recall that my calibration choices imply that learning takes place very quickly for those in the low-learning occupation.

persistence of shocks  $z_t$  confounds and slows down learning.



Figure 8: Evolution of beliefs

Moreover, there is significant heterogeneity in learning times and belief paths within the high learning occupation. Figure 8 plots the evolution of beliefs by age for two male agents (i = 1, 2) in the high-learning occupation, where agent 1 has a true  $\beta$  which is much larger than the average  $\beta$  observed for that occupation group, while agent 2 has a true  $\beta$  which is very close to the mean. Although both agents have the same mean beliefs at age 23, (equal to  $\bar{\beta}(H)$ ), they immediately diverge after the first period. Interestingly, although agent 1 learns about his true parameter almost 10 years before agent 2 does, the beliefs of the former barely change during the first five periods of his life. Because agent 2's parameter is much further away from his initial prior, his early wage draws are much more informative than the wage draws of agent 1 about the true  $\beta$ . In other words, the Kalman gain  $(\ln \tilde{y}_t - \mathbf{H}'_t \mathbf{\hat{S}}_{t|t-1})$  in equation (5) is much bigger for agent 2 during the first 5 periods of work.

### 5 Benchmark Results

In this section I present the main results of the model and discuss its central implications. The left panel in Figure 9 plots the participation profiles generated by the model: the solid line corresponds to those in the low learning occupation while the dashed line is for those in the high learning occupation. The corresponding data series are given by the filled (high learning) and white (low learning) dotted lines. Even though I only targeted the average participation rate for ages 25-45 and the average participation of women with children under 6, the model is able to generate differently shaped participation profiles for women in different occupations. In particular, it is upward sloping for those in low learning and downward sloping for those in high. Furthermore, the model does a very good job at matching the distribution of mother's ages at first birth seen in the data. As one can see from the right panel in Figure 9, 13.9% of the women in low learning occupations become mothers right away at age 23, but the proportions are lower for all ages thereafter. On the other hand, a steady 9% of women in high learning occupations choose to become mothers for the first time during each of the first 6 periods of the model.



Figure 9: Women's work and fertility behavior. Model vs data

Given that I only targeted a few aggregate statistics, what mechanisms in the model are responsible for these different behavior patterns across women? In order to better understand these effects I next discuss some simple comparative statics.

#### Husband's Wages and Assets

Husband's wages and household assets have standard roles with respect to the wife's work decision. At any time t, ceteris paribus, a woman who has a husband with a higher wage  $y_{mt}$  will be less likely to work; the same holds for the woman with higher household assets  $a_t$ . In particular, as  $y_{ft}$  increases, the maximum of  $E_t \left[ \mathbb{V}_t^w \left( y_{mt} | \Omega_t^- \right) \right]$  and  $\mathbb{V}_t^n \left( y_{mt} | \Omega_t^- \right)$  switches once from working to not working. Again, the same holds for assets.<sup>25</sup> With regard to the fertility decision in the model, ceteris paribus, households with higher (expected) wealth and higher male wages are more likely to choose to become parents at earlier ages. The intuition for this is straightforward: these households expect higher utility from consumption in the future so they would like to supplement their utility now, when consumption levels are low, through the presence of children in the household.

#### Women's Beliefs and Wages

The effects of the wife's wage draws and beliefs on the fertility and female work decision are a little less straightforward and merit some discussion. Figure 10 plots the work histories, belief paths and age at first birth for four different women. The two panels on the left correspond to two women with low wage growth  $\beta$  in the low-learning occupation, who are identical in all aspects (i.e.  $\alpha, \beta$ , husband, initial assets...) but hold different beliefs  $\hat{\mathbf{S}}_{t|t}$  for all  $t \geq 1$ . In particular, those of the "high beliefs" type converges to her true  $\beta$  from above while those of the "low beliefs" woman converge to to her true  $\beta$  from below (see top left panel). The right side plots the corresponding series for two women with high  $\beta$  in the high-learning occupation. Examining first the women with low  $\beta$  (left panels), the one with a lower  $\hat{\beta}_{t=1}$  chooses to have children right away and not work during the 8 periods when she has a child under the age of 6 at home. Her beliefs evolve slowly during this non-working period. On the other hand, the woman with high belief at time t = 1 chooses to work for the first three periods of life, and quickly learn about her true  $\beta$ . As her beliefs approach her true parameter, she becomes a mother at 25 and quits the labor market for 10 periods. Both women (return to) work after their children are older, although the woman who has children later stays out of the labor market for longer. This is household assets are higher due to her past labor participation before childbirth.

As one can see from the panels on the right, the women with high wage growth  $\beta$  have different belief paths and different labor and fertility behaviors than their low-learning counterparts. First, and as previously discussed in the right panel of Figure 9, these women give birth for the first time at later ages: the one with low beliefs becomes a mother at age 27 while her counterpart with high beliefs does not do so until she is 36. One interesting feature is that the woman with high wage growth and low beliefs becomes a mother earlier but does not stop working while she has young children in the household.

 $<sup>^{25}\</sup>mathrm{The}$  notation  $\Omega^-_t$  denotes the state vector  $\Omega$  minus the variable  $y_{ft}.$ 



Figure 10: Work and Fertility Histories for women with low  $\beta$  (left panels) and high  $\beta$  (right panels)

This is due to the interaction between beliefs and wage draws. The woman with more pessimistic beliefs becomes a mother at age 27 anticipating not working while she has young children at home because she does not expect high wage draws in the future. However, next period, when her first child arrives and she draws a new wage  $y_{ft}$ , that wage is higher than expected and she instead chooses to work. This occurs all throughout motherhood as she revises her pessimistic beliefs. On the other hand, the woman who began life with a much more optimistic belief about her wage growth  $\beta$  chooses to delay fertility and work all throughout her 20s. As her beliefs converge to her true  $\beta$ , she then chooses to become a mother. This revision of beliefs combined with high household assets leads her to exit the labor market while she has small children at home. She does not return to work after after her children grow up because of a combination of high household assets and large wage penalties. In particular, the calibrated value of  $\delta$ implies that her expected wage upon return is 12% lower than her last observed earnings. Comparing her wage loss to the equivalent loss of 1.9% for the women with low  $\beta$  analyzed in the left panels helps explain why these women have very different participation profiles.<sup>26</sup> For this analysis it appears that

<sup>&</sup>lt;sup>26</sup>Recall that the wage loss associated with a career interruption ( $\delta$ ) affects wage growth  $\beta$  multiplicatively in the log wage

both learning and wage penalties play significant roles in shaping women's work and fertility. I next turn to analyzing each of their separate contributions to these outcomes.

### 5.1 Learning versus wage penalties

In order to quantify the contributions of learning and wage penalties on women's work and fertility choices, I use the benchmark calibrated model to perform the two following counterfactual experiments. In the first experiment, I give all households in the model perfect information about both the husband's and wife's wage parameters  $\alpha, \beta$  and  $z_t$  for all periods  $t \ge 1$  while keeping all other features of the model intact. In a second experiment, I set the wage loss due to a career interruption in the benchmark model to zero ( $\delta = 0$ ), thus agents now face imperfect information but non-participation does not lead to wage penalties.<sup>27</sup> Figures 11 and 12 plot the participation and fertility behaviors for experiment (i) in the left panel and for experiment (ii) on the right panel.

Figure 11: Participation profiles in model with perfect information (left panel) and  $\delta = 0$  (right panel)



As can be seen from the participation profiles plotted in Figure 11 both the long periods required for learning and the wage penalty are key in generating the downward sloping participation profile of the women in occupations with high learning. Without either of these features, the model still does a fairly good job at matching the (upward sloping) participation profile of women in low-learning occupations,

specification.

<sup>&</sup>lt;sup>27</sup>The total proportion of women who become mothers is pretty much unchanged from the benchmark in both of these experiments. It is 88.76% in the first experiment and 89.10% in the second experiment (versus 88.69% in the benchmark). Also, in each experiment I recalibrate the discount fact  $\gamma$  such that the wealth-to-income ratio generated by the model matches the target in the data.

but it instead generates either a flat or an U-shaped participation profile for the women in high-learning occupations. On the other hand, examining the distributions of age at first birth plotted in Figure 12, one finds that while wage penalties have very small effects on women's fertility timing decisions (the right panel), imperfect information about wage parameters are key in generating the fertility patterns observed in the data for women in the high-learning occupations (left panel).

In the first experiment, when households are given perfect information about the wage parameters of the husband and the wife, 27.3% of the mothers in the high-learning occupation choose to have children at age 23. When agents know about their wage growth prospects, the women who expect high future income are more likely to have children early because the opportunity cost of non-participation due to young children (foregone wages) is lower at youth. However, most of these early mothers still choose to remain in the labor market in order to avoid the large wage losses associated with a career interruption. Participation rates for these women are only slightly lower during the ages of 23 to 30 when compared to the benchmark model: 91.0% on average versus 96.4%. Participation rates increase thereafter until they peak between the ages of 35 and 40, at around 95%. The model with perfect information has higher participation rates for those in the high-learning occupation between ages 35 and 45 because more of them choose to have children younger such that, by their time their children grow up, household consumption is still low and there are higher gains to working. These changes in work behavior generate participation profiles for women in low and high-learning occupations which are similarly hump-shaped. Overall, when households are given perfect information about wage parameters, women's participation rates do not change by more than 10 percentage points at any given period, but their fertility behavior is strikingly different from the benchmark.

The second experiment, on the other hand, generates participation rates which are significantly different from those in the benchmark model. As the right panel in Figure 11 shows, the women in high-learning occupations have a participation rate between the ages of 23 and 33 which is on average 12.6 percentage points lower than in the benchmark model. For the women in low-learning occupations, average participation rates during those same ages is 4.4 percentage points lower in this experiment. These differences are entirely due to changes in the participation of mothers with young children. If career interruptions are not associated with wage penalties, these women who face high disutility from labor due to the presence of young children at home choose instead to not work during those years and work after



Figure 12: Distribution of mother's age at first birth in model with perfect information (left panel) and  $\delta = 0$  (right panel)

the children grow up. Wage penalties do not play a significant role in women's fertility behavior (see the right panel in Figure 12), although the model without them does generate a 1.2 percentage point per period increase in the proportion of mothers in the high-learning occupation who have children during the ages of 23 to 26.

To summarize, while learning is key in generating differences in fertility timing behavior across women in different occupations, the model implies that a combination of different learning speeds and wage penalties associated with career interruptions is necessary in order to explain the differences in labor force participation profiles of women in different occupations.

# 6 Welfare Implications

Given the importance of learning in shaping women's work and fertility choices I next proceed to quantify the welfare losses from households having imperfect information about wage parameters. I measure these losses as the percentage change in per-period consumption required to give each household in the benchmark model the same average lifetime utility they would get if they lived in a world where the income parameters  $\alpha, \beta$ , and  $z_t$  for the husband and wife were known (denote this world by "PI"). Thus, I compute  $\omega_i$  for each household *i* in the benchmark model such that

$$\sum_{t=1}^{T} \gamma^{t} u(\omega_{i} c_{it}^{*}, P_{it}^{*}; k_{it}^{*}) = \sum_{t=1}^{T} \gamma^{t} u(c_{it}^{PI^{*}}, P_{it}^{PI^{*}}; k_{it}^{PI^{*}})$$

where \* and  $PI^*$  denotes the optimal allocations for  $c_{it}$ ,  $P_{it}$  and  $k_{it}$ , for household *i* at age *t* in the benchmark model and in the world with perfect information, respectively.



Figure 13: Percent welfare loss from imperfect information by decile of wife's  $\beta$ 

On average, there is a welfare loss equivalent to 1.0% of per-period consumption; these losses, however, are not homogenous across the population. Figure 13 plots these losses across the deciles of the distribution of wives' wage growth  $\beta$ . The key feature in the figure is that while these losses are negligible for most of the economy, they are quite significant for those households with wives in the top deciles. The largest losses are at the top, where a household in the 10th decile loses 4.8% of per-period consumption due to imperfect information. For all households whose wives are in the bottom 8 deciles of the  $\beta$  distribution, the welfare losses are less than 1% of consumption.

The reasons for higher welfare in PI are closely linked to the behavioral changes analyzed in Figures 11 and 12, for the experiment in which households have perfect information. First, because women in occupations with high-learning have high wage growth, the opportunity cost of small children is lower when these women are young. When given perfect information about their wage parameters, they would rather have children early and (possibly) forego relatively lower wages due to high disutility from working with young children. A second and perhaps more interesting channel of welfare losses comes from the interaction between slow learning and wage penalties. Since the probability of conception declines with age, some women in the high-learning occupation choose to have children before they fully learn their

wage parameter  $\beta$ . Furthermore, these also tend to be extremely productive women whose true parameter lies very far from her prior. Upon childbirth, some of these women choose to exit the labor market for a few years, incurring large wage losses from this career interruption. One consequence of this behavior is that the model generates what can be interpreted as a glass ceiling in wages. While the overall ratio of female to male wages during the ages of 45 to 50 is 0.68, the same statistic for those in the high-learning occupation is much lower, 0.55. This difference mostly reflects the fact that the distribution of observed female wages at the ages of 45-50 has a much shorter upper tail than the distribution of male wages. For example, the top female wage at those ages is only 34.9% of the top male wage, while the corresponding statistic at the ages of 25-30 was 79.3%.

# 7 Concluding Remarks

In this paper, I use an incomplete markets life-cycle model in order to examine possible drivers for the work and fertility patterns of women across different occupations. In particular, I investigate whether these differences in behavior can be explained by differences across occupations in: (i) the number of years that it takes for an individual to learn her true productivity and (ii) the wage penalties that accompany work interruptions. The key feature of the model is that agents do not know their own productivity parameters but instead learn about them by observing occupation specific wage draws and productivity signals. The model calibrated to a few aggregate statistics from the NLSY79 data can capture the different participation and fertility patterns for women in different occupations.

I find that both imperfect information and wage penalties have equally significant effects on the participation rates of women in high learning occupations, but that learning is crucial in explaining the differences in the timing of fertility. The differential impact of learning on work versus fertility choices is due to the differences in dynamic links of those decisions. With respect to differences in participation, my model without either learning or wage penalties generates an average participation rate for women in the high-learning occupation which is around 4 percentage points higher than the benchmark. These features, however, have negligible effects on the participation of women in low-learning occupations (less than 0.5 percentage points). Most importantly, in the absence of either of these features, life-cycle participation profiles are hump-shaped for women in both high and low learning occupations. Reflecting the importance of learning, I also find that welfare losses due to imperfect information are equivalent to almost 5% of

consumption for the households whose wives are in high-learning occupations. Finally, I also show that learning may have a significant role in generating what looks like a glass ceiling on female wages.

Although the two features examined in this paper (learning and wage penalties) mostly pertain to the behavior of women in high-learning occupations, the fact that these women are very productive professionals implies that their work decisions may have a potentially significant economic impact. One question for future research would be to examine the extent that these occupational features are related to the fact that married women's labor force participation rates have stagnated at under 80% since the 1990s, after growing from around 5% at the turn of the 20th century. As women surpass men in higher education and more of them join the ranks of high-powered professionals, it is crucial to understand how these occupational choices interact with fertility behavior in order to allow women to fully take advantage of labor market opportunities.

# References

- Abowd, J. and D. Card (1989). On the covariance structure of earnings and hours changes. *Econometrica* 57, 411–445.
- Albanesi, S. and C. Olivetti (2009). Gender roles and medical progress. working paper.
- Altonji, J. G. and L. M. Segal (1996). Small sample bias in gmm estimation of covariance structure. Journal of Business and Economics Statistics 109, 659–684.
- Anderson, D. J., M. Binder, and K. Krause (2002). The motherhood wage penalty: Which mothers pay it and why? The American Economic Review, Paper and Proceedings 92(2), 354–358.
- Attanasio, O. (1999). Consumption. In J. B. Taylor and M. Woodford (Eds.), Handbook of Macroeconomics, Volume 1B. Amsterdam: Elsevier Science.
- Attanasio, O., H. Low, and V. Sanchez-Marcos (2005). Female labor supply as insurance against idiosyncratic risk. *Journal of the European Economic Association* 3(2-3), 755–764.
- Attanasio, O., H. Low, and V. Sanchez-Marcos (2008). Explaining changes in female labor supply in a life-cycle model. *The American Economic Review* 98(4), 1517–1552.
- Bailey, M. (2006). More power to the pill: The impact of contraceptive freedom on women's life cycle labor supply. *Quarterly Journal of Economics* 121(1), 289–320.
- Bertrand, M., C. Goldin, and L. F. Katz (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics* 2, 228–255.
- Black, S. E. and C. Juhn (2000). The rise of female professionals: Are women responding to skill demand? The American Economic Review, Paper and Proceedings 90(2), 450–455.
- Budig, M. and P. England (2001). The wage penalty for motherhood. *American Sociological Review 66*, 204–225.
- Caucutt, E. M., N. Guner, and J. Knowles (2002). Why do women wait? matching, wage inequality, and the incentives for fertility delay. *Review of Economic Dynamics* 5, 815–855.
- Chamberlain, G. (1984). Panel data. In Z. Griliches and M. Intriligator (Eds.), Handbook of Econometrics, Volume 2, Chapter 22. North Holland.
- Cox, D. (1984). Panel estimates of the effects of career interruptions on the earnings of women. *Economic Inquiry 22*(3), 386–403.
- De Nardi, M., E. French, and J. B. Jones (2010). Why do the elderly save? the role of medical expenses. Journal of Political Economy 118(1), 39–75.

- Erosa, A., L. Fuster, and D. Restuccia (2002). Fertility decisions and gender differences in labor turnover, employment, and wages. *Review of Economic Dynamics* 5, 856–891.
- Fernández, R. (2011). Cultural change as learning: The evolution of female labor force participation over a century. working paper.
- Fernandez-Villaverde, J. and D. Krueger (2007). Consumption over the life cycle: facts from consumer expenditure survey data. *The Review of Economic and Statistics* 89(3), 552–565.
- Fogli, A. and L. Veldkamp (2011). Nature or nurture? learning and the geography of female labor force participation. *Econometrica* 79(4), 1103–1138.
- Goldin, C. and L. F. Katz (2000). Career and marriage in the age of the pill. *The American Economic Review, Paper and Proceedings* 90(2), 461–465.
- Goldin, C. and L. F. Katz (2002). The power of the pill: Oral contraceptives and women's career and marriage decisions. *Journal of Political Economy* 110(4), 730–770.
- Greenwood, J., A. Seshadri, and M. Yorukoglu (2005). Engines of liberation. *Review of Economic Studies* 72(1), 109–133.
- Guvenen, F. (2007). Learning your earning: Are labor income shocks really very persistent? The American Economic Review 97(3), 687–712.
- Guvenen, F. (2009). An empirical investigation of labor income processes. *Review of Economic Dynamics* 12(1), 58–79.
- Hamilton, J. D. (1994). Time Series Analysis. Princeton University Press.
- Heathcote, J., K. Storesletten, and G. L. Violante (2010). The macroeconomic implications of rising wage inequality in the united states. *Journal of Political Economy* 118(4), 681–722.
- Herr, J. L. and C. Wolfram (2010). Work environment and "opt-out" rates at motherhood across high-education career paths. working paper.
- Jones, L., R. Manuelli, and E. McGrattan (2003). Why are married women working so much? working paper.
- Kambourov, G. and I. Manovskii (2009). Occupational specificity of human capital. International Economic Review 50(1), 63–155.
- Knowles, J. (2007). High-powered jobs: Can contraception technology explain trends in women's occupational choice? working paper.
- Meyers, P. B. and A. M. Osborne (2005, September). Proposed category system for 1960–2000 census occupations. U.S. Bureau of Labor Statistics Working Paper No. 383.

- Mincer, J. and S. Polachek (1974). Family investments in human capital: Earnings of women. *Journal of Political Economy* 82(2), 76–108.
- Mulligan, C. B. and Y. Rubinstein (2008). Selection, investment and women's relative wages over time. *Quarterly Journal of Economics* 123(3), 1061–1110.
- Olivetti, C. (2006). Changes in women's hours of work: The role of returns to experience. *Review of Economics Dynamics* 9(4), 557–587.
- Rendall, M. (2010). Brain versus brawn: The realization of women's comparative advantage. working paper.
- Storesletten, K., C. I. Telmer, and A. Yaron (2004). Consumption and risk sharing over the life cycle. Journal of Monetary Economics 51(3), 609–633.
- Topel, R. H. and M. P. Ward (1992). Job mobility and the careers of young men. Quarterly Journal of Economics 107(2), 439–479.
- van Noord-Zaadstra, B. M., C. W. Looman, H. Alsbach, J. D. Habbema, E. R. te Velde, and J. Karbaat (1991). Delaying childbearing: Effect of age on fecundity and outcome of pregnancy. *BMJ 302*, 1361–1365.
- Waldfogel, J. (1997). The effect of children on women's wages. American Sociological Review 62(2), 209–217.
- Waldfogel, J. (1998). Understanding the "family gap" in pay for women with children. *Journal of Economic Perspectives 12*, 137–156.
- Weiss, Y. (1986). The determination of life-time earnings: A survey. In O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics*. Amsterdam: North-Holland.
- Weiss, Y. and R. Gronau (1981). Expected interruptions in labour force participation and sex-related differences in earnings growth. *Review of Economic Studies* 48(607-619).
- Wilde, E. T., L. Batchelder, and D. T. Ellwood (2010, December). The mommy track divides: The impact of childbearing on wages of women of differing skill levels. NBER Working Paper No. 16582.
- Wolff, E. (2000). Recent trends in wealth ownership, 1983–1998. Working Paper 300, Jerome Levy Economics Institute, New York.
- Wood, R. G., M. E. Corcoran, and P. N. Courant (1993). Pay differences among the highly paid: The male-female earnings gap in lawyers' salaries. *Journal of Labor Economics* 11(3), 417–441.

# A Appendix: Three Occupation Groups

My decision to split individuals into only 2 occupational learning groups was grounded in the fact that an earlier data and model specification in which I had three different training categories showed very similar work and fertility behaviors between women in two of the groups. In particular, Figure 14 plots the equivalent of Figure 2 when I divide women into three occupational groups: (i) those in occupations with low SVP (under 2 years), (ii) those in occupations with medium SVP (between 2 and 7 years) and finally, (iii) the women in occupations with high SVP (over 7 years). As one can see from Figure 14, the life-cycle participation profiles are very similar for the women in low and medium SVP and the same applies about thei distribution of age at first birth.





Thus, given these similarities and for simplicity's sake, I chose to focus only on the distinction between the women in high-learning occupations (over 7 years of SVP) and those in occupations with under 7 years of SVP (the low-learning group in my benchmark specification).

# **B** Appendix: Data

In this section I outline details about my sample selection procedures.

### B.1 NLSY79 Sample

The NLSY79 is panel dataset of a representative sample of 12,686 young men and women in the US who were between the ages of 14-22 when the survey began in 1979. This is the main dataset used in this paper for employment, occupations, hourly wages and fertility history. The NLSY79 provides data on weekly work history beginning in 1978 (like most surveys, the work history questions pertain to the year the previous year). I consider a woman to be working if she reports non-zero working hours in that year. Each individual is assigned to an occupational learning category based on the occupational code associated with the job in which they worked the most hours during that year. As I explained in the text, after the age of 35, only 0.6% of my sample switches between occupations which belong to different learning categories (low versus high learning). Since my model abstains from occupational switches, I throw out those individuals who are observed switching after the age of 35 and I assign each individual to the first occupation I observe them in after 35.

In order to compute the log hourly wages for men plotted in Figure 1, by dividing total labor earnings by total work hours in the previous calendar year. Once again, occupational category is assigned based on the job with the highest number of hours that year. I deflate nominal wages to 1992 dollars using CPI for all urban consumers. Finally, the age of the woman at firth birth is constructed using a variable which codes the year in which the woman gave birth to her first child.

### B.2 PSID Sample

I use the PSID for the estimation of the parameters of the income process. The main variable for labor earnings in the PSID is given by the labor earnings for the head. I include all male heads of households from the waves from 1968-1997 from the PSID into the sample. I deflate the earnings variable to 1992 dollars using CPI for all urban consumers. I exclude exclude individuals in the the Latino, SEO and immigrant samples. I also drop observations from people younger than 25 and people older than 65 years old and those who report being self-employed. I choose only individuals with at least 20 (not necessarily consecutive) observations. Furthermore, I drop individuals with missing, top-coded and zero earnings, those with less than 520 hours or more than 5110 annual work hours and those with hourly labor earnings which are less than half the legal minimum wage in that year. Individuals with changes in log earnings greater than 4 or less than -2 are also eliminated from the sample. This leaves me with 1292 individuals in my main PSID sample.

# C Appendix: Estimation and Solution Method

### C.1 Minimum Distance Estimation

Using the measure of annual earnings  $(w_{e,t})$  for an individual from the selected PSID sample which ranges from year t = 1, ..., T with  $age_{i,t}$  observed at time t, I regress:

$$\ln w_{i,e,t} = \mathbb{D}_t + \theta_1 age_{i,t} + \theta_2 age_{i,t}^2 + \theta_3 age_{i,t}^3 + \theta_4 age_{i,t}^4 + \tilde{w}_{i,e,t}$$

where  $\theta_j$ , j = 1, 2, 3, 4 are the age polynomials used in the paper and  $\mathbb{D}_t$  denotes the year dummy. Given my assumptions, I parametrize the residual earnings from that equation  $\tilde{w}_{e,t}$  for each agent *i* as:

$$\begin{split} \tilde{w}_{i,e,t} &= \alpha_i + \beta_i e_{i,t} + z_{i,e,t} + \varepsilon_{i,e,t} \\ z_{i,e,t} &= \rho z_{i,e-1,t-1} + \eta_{i,e,t} \\ (\alpha_i, \beta_i) \sim^{iid} \mathbf{N}\left(\mathbf{0}, \Sigma\right), \quad \varepsilon_{i,e,t} \sim^{iid} \mathbf{N}(0, \sigma_{\varepsilon,t}^2), \quad \eta_{i,e,t} \sim^{iid} \mathbf{N}(0, \sigma_{\eta,t}^2), \quad z_{i,0,t} = 0 \\ (\alpha_i, \beta_i) \perp \varepsilon_{i,e,t} \perp \eta_{i,e,t} \end{split}$$

and

$$\Sigma = \begin{pmatrix} \sigma_{\alpha}^2 & \sigma_{\alpha\beta} \\ \\ \sigma_{\alpha\beta} & \sigma_{\beta}^2 \end{pmatrix}$$

Define the parameters to be estimated as  $\Omega = \left\{\rho, \sigma_{\alpha}^2, \sigma_{\alpha\beta}, \sigma_{\beta}^2, \{\sigma_{\eta,t}^2, \sigma_{\varepsilon,t}^2\}_{t=1}^T\right\}$ . Thus, the simulated moments of the wage process will be functions of  $\Omega$ . Let the covariance matrix for income residuals  $\tilde{w}_{i,e,t}$  be denoted by  $\widehat{\mathbf{M}}$  with a typical element given by

$$\widehat{m}_{j,k,n} = \frac{1}{I_{jkn}} \sum_{i=1}^{I_{jkn}} \widetilde{w}_{i,j,k} \widetilde{w}_{i,j+n,k+n}$$

Define  $\mathbf{M}(\Omega)$  and  $m_{j,k}(\Omega)$  as their simulated counterparts. The moment conditions to be estimated are given by:

$$\mathbf{E}\left[\mathbb{I}_{i,j,k,n}(\widehat{m}_{j,k,n} - m_{j,k}(\Omega))\right] = 0$$

where  $\mathbb{I}_{i,j,k,n}$  is an indicator function for if the individual *i* is present in both cells (e = j, t = k) and (e = j + n, t = k + n).

Define  $\overline{\mathbf{M}}(\Omega)$  and  $\widehat{\overline{\mathbf{M}}}$  as vectors of the stacked *unique* entries in those respective matrices. Then the parameters  $\widehat{\Omega}$  such that

$$\min_{\boldsymbol{\Omega}} \left[ \widehat{\overline{\mathbf{M}}} - \overline{\mathbf{M}}(\boldsymbol{\Omega}) \right]' \tilde{\boldsymbol{W}} \left[ \widehat{\overline{\mathbf{M}}} - \overline{\mathbf{M}}(\boldsymbol{\Omega}) \right]$$

where  $\tilde{W}$  is a weighting matrix. Following Altonji and Segal (1996) who show that the optimal weighting matrix leads to small sample bias, I choose the weighting matrix to be the identity matrix.

### C.2 Solution of Household Problem

I begin by outlining the algorithm to the solution of the household consumption/savings problem:

- Using the population distribution parameters estimated directly from the data, draw 10,000 types of  $(\alpha_i, \beta_i)$  from  $\mathbf{N}([\mu_{\alpha}, \mu_{\beta}], \Sigma)$ .
- Divide the spanned 2-dimensional space defined in  $(\alpha, \beta)$  into 6 areas. Within each area j = 1, ..., 6, calculate the average  $(\overline{\alpha}, \overline{\beta})_j$ . Since the husband and the wife have the same distribution over  $(\alpha_i, \beta_i)$ , there will be 36 possible combinations of  $(\overline{\alpha}, \overline{\beta})_{wife=j}, (\overline{\alpha}, \overline{\beta})_{husb=k}$  across households. The model will be solved once for each combination j, k, in parallel, using one processor per combination. In each solution, all belief and income paths will be drawn as functions of  $(\overline{\alpha}, \overline{\beta})_{wife=j}, (\overline{\alpha}, \overline{\beta})_{husb=k}$ . Although the value function does not explicitly depend on the individual's true type, attempts at solving the model only once lead to large numerical errors. By solving the model for these groups separately, the grid for the state space becomes a direct function of the true wage generating parameters. The steps that follow will be implemented in each solution j, k; for simplicity, I omit the subscripts.
- Draw 100 income paths for  $(\overline{\alpha}, \overline{\beta})_{wife}$  and another 100 for  $(\overline{\alpha}, \overline{\beta})_{husb}$ . Draw 100 signal paths for the husband and wife, respectively.
- Compute all the possible evolutions of  $\widehat{\mathbf{S}}_{t|t}$  for the husband and the wife. Note that for the wife, I need to take into account all the possible combinations of working versus not working at every period and its effects on the updating of beliefs using the signal. I then use the simulated belief paths in order to determine the bounds for my grids for beliefs. The household problem is solved each point

of the independent cartesian grids for  $\widehat{\alpha}_t^m, \widehat{\beta}_t^m, \widehat{z}_t^m, \widehat{\alpha}_t^f, \widehat{\beta}_t^f$ , and  $\widehat{z}_t^f$ . I used grids with 5 points in each dimension. I simulated the model using the calibrated parameters from the benchmark and 10 grid points in each dimension; there was no noticeable difference in my results.

When computing the expected value at time t for time t+1, I use the conditional distributions for  $y_{t+1}^m, y_{t+1}^f$  given by equation (9), but truncated at 3 standard deviations. I do this because, if the agents use the distribution of  $y_{t+1}^m, y_{t+1}^f$  implied by their true generating process, there are some wage draws which will be very far away from their beliefs at some given points. As with Guvenen (2007), these draws occur only rarely, but when they do, the numerical errors are very large.

The discrete participation and fertility decisions in the model introduce non-concavities, which remains even after conditioning on the participation of the wife today. I alter the approach of Attanasio, Low, and Sanchez-Marcos (2008) and verify that, once I hold all the other states fixed, there are 3 possible levels of  $a_k^*$ , k = 1, 2, 3 such that conditional on the participation decision today, the participation-fertility decisions tomorrow will change. This will hold as long as between each combination of 2 values functions (Work + have children, not work + not have children....) a single-crossing condition exists, i.e. conditional on decisions today and tomorrow, the value function is "sufficiently" concave. I verify numerically that this is the case.

## D Appendix: Robustness Checks

**Pension:** The US social security rules imply that a household receives either 1.5 times the husband's pension or the sum of the husband and the wife's pensions. Given that I've calibrated the pensions in my model to 1.5 times the husband's pension at all times, I experiment with giving the household retirement benefits which correspond to 1.8 and 2 times the husband's pension. As mentioned in the text, even without recalibrating the model, the average LFP in the economy only drops by 0.21 percentage points (from 76.51 to 76.30). Recalibrating the model using the new pension values, almost all the quantitative results of the paper go through with changes which are within one-tenth of their original values.

**Initial Priors:** I calibrated the benchmark model with a common prior for each occupation group  $\widehat{\mathbf{S}}_{1|0}(\lambda)$  because I wanted to isolate the effects from different wages/signals from the effects of different priors. In

order to evaluate the implications that this assumption could have had in driving my results, I recalibrate the benchmark model in a setting where each agent draws  $\hat{\alpha}_{1|0}$ ,  $\hat{\beta}_{1|0}$  from occupation specific distributions given by

$$\widehat{\alpha}_{1|0}(\lambda) \sim \mathcal{N}(\bar{\alpha}(\lambda), \sigma_{\alpha}^2), \quad \widehat{\beta}_{1|0}(\lambda) \sim \mathcal{N}(\bar{\beta}(\lambda), \sigma_{\beta}^2)$$
(18)

Once again, even without recalibrating the model most of the aggregate statistics targeted in the internal calibration change by less than 1%. The predictions concerning the labor profiles for the women in each of the two groups remains the same and the quantitative implications of the counterfactual experiments are also unchanged.

**Borrowing Constraints:** One may have concerns that some of the results in my paper are driven by the no-borrowing constraint that I imposed on my households. In a different version of the model, I allow all households to borrow up to an ad-hoc amount  $\underline{a}$ , i.e. every period households face the borrowing constraint  $a_{t+1} > -\underline{a}$  and a terminal solvency condition given by  $a_{T+1} \ge 0$ . I calibrated  $\underline{a}$  internally such that the model generates the same proportion of people who have 0 or negative wealth as in the data (15.5% in 1983 according to Wolff (2000)). The implied value for  $\underline{a}$  corresponds to 32% of average per-period male wages. This specification did not yield results which were very different from the benchmark. The reason for this is twofold: while households in low-learning occupations quickly learn their wage profiles and begin borrowing against their expected future earnings, they do not go heavily into debt because they do not expect very high earnings. On the other hand, the households in the high-learning occupations who expect high future wages also do not borrow large amount because it takes them a long time to learn whether they can pay their debt off. Thus the average borrower household in this specification only has debts which are around 11% of the mean household income during the ages of 23 to 33.