Estimating a Dynamic Adverse Selection Model: Labor Market Experience and the Changing Gender Earnings Gap

George-Levi Gayle and Limor Golan
Tepper School of Business
Carnegie Mellon University

July 7, 2006

Abstract

This paper investigates the role of labor market attachment, on-the-job human capital accumulation, and occupational sorting in the narrowing gender earnings gap over the past three decades. Many of the observe changes, such as labor market attachment, experience and occupation choice are endogenous. We formulate of a dynamic general equilibrium model of earnings determination and estimate it using the PSID. In the model, statistical discrimination, labor market attachment, experience and occupational sorting are endogenous.

In our model, workers are heterogeneous with respect to characteristics affecting the degree of the labor market attachment and labor supply. These characteristics are privately to workers. It is costly for firms to hire new workers; these costs vary across occupation (for example, in some occupation there is higher level of investment in firm-specific training). The returns to labor market experience also differs across occupations. Employers observe age, experience, education, and other skills.

Based on observable characteristic, employers form beliefs regarding the length of the future employment spell when they offer employment contracts. Employment contracts consist of hours and earning. Thus, in equilibrium, statistical discrimination, based on degree of labor market attachment, may arise. We assume contracts cannot bind workers and employers to long-term employment. If firms believe that women are less likely to commit to long employment periods, they offer them lower earnings profiles. As a result, women work less hours, accumulate lower levels of human capital and sort into occupations which have lower returns to labor market experience.

We developed a new multi-step semiparametric estimation strategy that allows us to estimate the model without solving it. This is particularly important in this class model which may exhibit multiple equilibria. The idea is to estimate the employer’s problem in the first step, along with other inputs from the individual Euler equation of consumption. In the second step, these estimates are used to nonparametrically estimate the firms’

*We thank Zvi Eckstein, John Kennan, Bob Miller, Steve Spear, Yoram Weiss, the participants of SOLE 2005, Tel-Aviv summer workshop 2005, and the seminar participants in Purdue University, University of Minnesota, Northwestern University and Carnegie Mellon for comments and suggestions.
beliefs, the workers’ choice-specific probabilities and their derivatives. In the final step, these estimates are combined with a tractable alternative representation of the workers choice-specific valuation to form moment conditions for the estimation of the structural parameters of the agents utility function. The estimates of the structural parameters are $\sqrt{N}$ consistent (where $N$ is the sample size) and asymptotically normal although the second step is estimated nonparametrically.

The model allows us to estimate the gender wage gap accounting for the endogeneity of discrimination, labor market experience and occupational sorting. We then explore the effect of occupation specific aggregate productivity shocks, changes in cost of participation in the labor market (may result from improved home production technology), and demographic changes on the decline in the gender earnings gap.

1 Introduction

The gender earning gap in the U.S. has been declining since the late 1970s (log male/female wage differential decline from 0.47 to 0.33 between 1979-1988). During that period, labor force participation of women between the ages of 25 to 54 year increased from below 40% in 1965 to above 80% in 1996. The percentage of women holding a professional and administrative job went from 18% to 45% between 1976 and 1992 (Lewis 1996). Furthermore, women’s accumulated labor market experience has increased significantly during the 1980s when women’s relative wages increased the most (Kahn and Blau 1997 finds that the increase in accumulated labor market experience is the largest factor accounting for the decline in the gender earnings gap). The "unexplained" portion of the earnings gap, which is sometimes attributed to discrimination, has declined as well\(^1\). This paper investigates the role of labor market attachment, on-the-job human capital accumulation, occupational sorting and discrimination in the narrowing gender earnings gap. The main challenge in quantifying these effects is to account for the endogeneity of labor supply, discrimination and earnings. The contribution of this paper is three fold. First, we formulate a dynamic adverse selection model of labor supply and human capital accumulation in which gender discrimination and the earnings gap arise endogenously. Secondly, we develop a three-step estimation technique, and estimate it using the PSID. Thirdly, we decompose the changes in the gender earnings gap into the different components, and quantify the effect of the incomplete information on the changes in experience and the earnings gap.

There are two broad types of employers’ discrimination in the literature. The first type is taste based discrimination, formulated By Gary Becker (1971), and the second type is discrimination which results from incomplete information, pioneered by Arrow (1972), and Phelps(1973). Discrimination of the first type may not persist in a competitive environment, but frictions, such as search frictions, may lead to persistent group differentials in the long run equilibrium (Eckstein and Bowles (1998)). Such models, however, are more adequate to explain race based discrimination because their predictions do not match patterns of gender

---

related labor market differences\textsuperscript{2}. Our model belongs to the class of incomplete information discrimination model, and in particular, to the literature on statistical discrimination first analyzed by Coate and Loury (1993). Whereas the statistical discrimination literature focuses on the effect of beliefs on productivity differences across groups, in our model the uncertainty is about the turnover propensity of workers. In particular, if employers believe that women are less attached to the labor force and may have shorter employment spells than men, these beliefs may be self-confirming leading to women accumulating lower levels of labor market experience.

We incorporate statistical discrimination based on beliefs on employment spells length in a framework of a general equilibrium dynamic adverse selection model\textsuperscript{3}. In the model, workers are heterogeneous with respect to characteristics affecting disutility from working; these characteristics evolve according to a known Markov process. Every period, workers choose consumption, whether to participate in the labor market and how many hours to work in order to maximize lifetime utility. Utility is an increasing function of consumption, and "non-market" hours (leisure and hours spend on producing home goods). Every period there is a random utility shock for the utility of participating and not-participating. There is non-separability of leisure in the utility function. This allows for the possibility that, the "stock" of past non-market hours affect current disutility from hours worked. We assume assets markets are complete. Firms compete over workers, and there is a free entry of firms into the market. The returns to experience, hours worked and costs of hiring new workers, differ across occupations. The firm-specific cost of hiring new workers are also identical within occupations and different across occupations. Employers observe age, experience, education and other skills, but there are characteristics affecting the disutility of working which are the worker's private information. In particular, these characteristics evolve according to some known process, creating correlation in the worker's "type" over time. Based on observable characteristic, employers form beliefs regarding the workers future employment spell when they offer labor market contracts. Labor market contracts consist of hours and earning. That is, we solve for a contract posting equilibrium.

Because of the firm-specific cost of hiring new workers, employers make rents on workers after the first year they are hired. Since the market is competitive, when hiring workers firms make zero expected profits. In equilibrium, beliefs about future participation of the worker in the firm enters the earnings equations. Workers choose participation (extensive margin) and hours (intensive margin). If women face lower earnings, since there is disutility from working, they work less and sort into occupations with lower returns to experience and lower costs of hiring workers. Thus, on average they accumulate lower levels of labor market experience than men. Our model is a signaling model. The decision to participate and how many hours of work may provide information on the worker's type. In equilibrium, information on workers is

\textsuperscript{2}For example, these models predict that the discriminated group is less likely to work in professions in which they have extensive contact with customers. Women, however, tend to work in service industries.

\textsuperscript{3}Pay differences can be generated in our model if the groups of male and female are ex-ante identical. We allow, however, in the estimation stage for differences across groups and estimate it. Furthermore, characteristics which are exogenous in our model, such as education, can generate pay gaps. We are able, however, to estimate these effects separately.
revealed gradually over time (this is a typical feature of models of dynamic adverse selection with correction in the type over time, and incomplete contracts. See for example Tirole 1996). Over time employers update beliefs based on individual labor market history. Thus, working more today, may affect the worker’s potential earnings not only through accumulation of experience, but also because of the possible effect on the employers’ beliefs. Therefore the model predicts that the information employers have on experienced workers’ is more accurate than information on young workers. Light and Ureta (1992) find that among older women controlling for all observable characteristics, women are more likely to stay in a job and earn as much as men. Several other papers examined empirically the relationship between the pay gap and the length of employment spells. Light and Ureta (1992) find evidence that "stayers" are paid more than workers with short employment spells. They also find that women are more likely to move and that this difference becomes smaller when workers are older. Altonji and Paxton (1992) find that job mobility is strongly linked to hours change and women who face change in family responsibility adjust their hours and this may lead to lower earnings. Baron et al. (1993) develops a model in which employers expect women to have higher turnover rate, and therefore, women have lower training levels explaining the lower earnings. The, contribution of our model is to formulate a model in which differences in labor market experience, occupational sorting and attachment to the labor force are affected by the beliefs about the turnover probability and therefore, arise endogenously.

One goal of this paper is to account for changes in the earnings gap over time. The literature focuses on several factors that may have cause the change in relative earnings over time. First, over time, there were aggregate productivity shocks affecting overall wages in the economy. In our model such changes are driven by occupation specific (random) shocks to productivity. They raise productivity for all workers equally. Our model, however, predicts that if women’s participation is lower, positive productivity shocks may increase participation. If positive productivity shocks have a bigger effect on women’s labor supply than on men’s labor supply, firm’s beliefs about women’s employment spells length may increase, causing a relative wage gains for women. The second possible source of changes in relative wages is a decline in costs of producing home goods. In our model, there are fixed costs in the utility when an individual participates in the labor market. If costs of home production good declined over time, these costs should decline as well. It may affect the beliefs about women’s attachment to the labor force as well. The third factor is changes in education, marriage and fertility over time. Such changes may cause changes in labor supply behavior because they affect the disutility from working.

Our model exhibits multiplicity of equilibria. Different beliefs generate different equilibrium outcome. This is standard in models of statistical discrimination (see Moro(2003) and Antonovics (2004)). In the standard statistical discrimination model, an identification problem of the following form arises, given the observed wage distribution and human capital investment, an econometrician is trying to recover the preference parameters and the beliefs about an unobserved variable that affects productivity. The source of the identification problem is that there may be more than one combination of preferences parameters and beliefs that could have generated the observed wage distribution. Our model, however, does not have that problem.
This is because the beliefs in our model are about future participation probability, conditional on observed characteristics. Panel data, however, allows us to observe next period participation decision of the individual, this have to be correct across the population conditional on the observed characteristics. Therefore, we can nonparametrically identify these beliefs under the assumption that for each cohort there is one equilibrium played.

We developed a new multi-step semiparametric estimation strategy that allows us to estimate the model without solving it. This is particularly important in this class model which may exhibit multiple equilibria. There are several moment conditions that our model yields. First, there is a consumption Euler equation. The idea is to estimate the employer’s problem in the first step, along with other inputs from the individual Euler equation of consumption. From labor supply decisions we obtain an Euler equation for hours. In addition, workers face a discrete choice problem each period of whether to participate in the labor market or not. The free entry into the labor market yields another moment conditions as the employment contracts are offered so employers make zero expected profit from hiring a worker. We utilize these conditions for identification and estimation. First from the standard Euler equation, we identify the individual marginal utility from wealth used in the hours Euler equation. Secondly, the beliefs are identified non-parametrically from equilibrium condition: conditional on the current information available to the employer, the expected participation is correct. The zero profit condition identifies the cost of hiring employees in each occupation. This is identified over variation in beliefs of future participation in the earnings equation. The production function parameters in each occupation is also identified off this equation. Then we utilize the hours Euler equation and the choice probabilities to identify the rest of the utility parameters (Altug and Miller 1998). The estimates of the structural parameters are $\sqrt{N}$ consistent (where $N$ is the sample size) and asymptotically normal although the second step is estimated nonparametrically. Our estimate is akin to a number of estimator in the literature for the estimation of discrete games and single agents models (see Hotz and Miller(1993), Altug and Miller(1998), Pakes, Oskrovsky and Berry(2004), Pesendorfer and Schmidt-Denglar(2003), Bajari, Benkard and Levin (2004)); our estimator is different, however, in that we are estimating a dynamic adverse selection model with both discrete and continuous controls.

We find that the cost of replacing workers are higher in professional occupations compared to nonprofessional occupations. The returns to experience are higher in the professional occupations than they are in non-professional occupations. Our model predicts that earnings-gap should be smaller in occupations with low costs of hiring workers and that women will sort into occupations in which there is a "smaller" penalty to accumulating lower levels of labor experience. These predictions are confirmed by the estimation results. The estimation results do not support the hypothesis that changes in home production technology explain the increase in participation of women in the labor market. Such changes should have caused a decrease in the fixed cost of participating in the labor market (estimated as part of the utility function specification). These findings are consistent with Jones, Manuelli and McGratten (2003)\textsuperscript{4}. Our estimation results suggest that fixed costs of participation accounts for only

\textsuperscript{4}Jones, Manuelli and McGratten (1993) analyze a model of labor supply of women and quantify the different causes for the decline in the gender earnings gap of married women. Discrimination is exogenous in their model.
small insignificant decrease in the gender earnings gap. Employers’ beliefs on women participation have changed over time, that is, beliefs about probability of working in the proceeding periods increased. This caused an increase in current earnings, which increased women labor market participation. Increase in overall productivity caused an increase in overall wages in the economy (could be caused by aggregate technology shocks) this increase causes changes in beliefs in participation of women which cause an increase in wages of women relative to men. Our estimation results lend support to the hypothesis that there are complementaries in non-market work for women and that there is specialization in non-market hours for women relative men. That is, the results support the possibility that the equilibrium outcome in one group affects the equilibrium outcome of another group due to cross-groups complementarities in the utility function. Finally, a decomposition of the change in the gender earnings gap shows that labor market experience accounts for between 65 and 67 percent of the decrease in the gender earnings and that changes in the employers’ beliefs accounts for between 6 to 8 percent of the decrease. Further analysis shows that private information, demographic changes and aggregate shock to market productivity accounts for most the change in the accumulation of labor market experience by women over the period relative to men.

2 Theoretical Model

This section describes the basic structure of the economy and the dynamic general equilibrium model. In the labor market, firms compete over workers, and the gender wage gap and occupation sorting are determined together with worker’s labor supply. We then analyze the model’s predictions and empirical properties.

2.1 Worker’s problem

This section describes the agents in the economy and their preferences. The choice problem and the budget set is described in the equilibrium solution section. We assume that there exist a continuum of workers (men, and women) on the unit interval [0, 1]. Let \( g = \{f, r\} \) denote the worker’s gender (female and male respectively) and let \( a \) denote the worker’s age-cohort (all workers who were born at the same year, who have the same number of years of completed education). Workers are finitely lived (denote the worker’s age by \( \tau = \{18, ..., 65\} \)). The calendar year is indexed by \( t \) (\( t = 1975, ..., 1994 \)). Each individual has preferences over non-market hours (time in which the individual does not work) and consumption. The consumption allocation to individual \( n \) at date is denoted by \( c_{nt} \). The other choice variable is \( h_{nt} \), the time spent at work by agent \( n \) in period \( t \). There exist a fixed amount of time in each period that

\(^5\)Moro and Norman (2004) analyze a statistical discrimination model in which discrimination arises in equilibrium not only because of coordination failure, but as an asymmetric equilibrium resulting from complementarities in the production function. That is, there are cross-group complementarities, and the equilibrium in one group affects the equilibrium outcome of the other group. In Coate and Loury, in contrast, there is no connection between the equilibrium outcome across the groups.
is available for working, which implies that the amount of time worked in each period can be normalized as $0 \leq h_{nt} \leq 1$. If $h_{nt} = 0$, the agent does not work at time $t$. Otherwise, the agent works the fraction of time $h_{nt} > 0$. For notational convenience, a participation indicator $d_{nt}$ is defined, where $d_{nt} = 1$ if and only if $h_{nt} > 0$, 0 otherwise.

Preferences are additive in consumption and leisure both contemporaneously and over time. It is assumed that there are both observed and unobserved (by the econometrician) exogenous, time-varying characteristics that determine the utility associated with alternative consumption and leisure allocations. Denote the former by the $K \times 1$ vector $z_{nt}$ and the latter by the $3 \times 1$ vector $(\varepsilon_{0nt}, \varepsilon_{1nt}, \varepsilon_{2nt})'$. It is assumed that $z_{nt}$ is independently distributed over the population with known distribution function $F_{0ga}(z_{nt} + 1 | z_{nt})$; the vector $(\varepsilon_{0nt}, \varepsilon_{1nt}, \varepsilon_{2nt})'$ is independent across $(n, t)$ and drawn from the population with a distribution function $F_{1}(\varepsilon_{0nt}, \varepsilon_{1nt}, \varepsilon_{2nt})$.

The current-period utility function at date $t$ for individual $n$ is defined as

$$U_{nt} \equiv d_{nt}u_{0}(z_{nt}) + u_{1}(z_{nt}, 1 - h_{nt}) + u_{2}(z_{nt}, c_{nt}, \varepsilon_{2nt}) + (1 - d_{nt})\varepsilon_{0nt} + d_{nt}\varepsilon_{1nt},$$

where $u_{0}$ represents the fixed utility costs from working, which depends on observed individual-specific characteristic but may change over from period to period. We allow this to change over time in order to capture the possible changing home production technology. It takes the following form:

$$u_{0}(z_{nt}) = B_{ot} + z_{nt}'B_{0}.$$ 

$u_{2}$ represents the current utility from consumption (which depends on consumption and observed characteristics of the individual and the idiosyncratic shocks $u_{2}(z_{nt}, c_{nt}, \varepsilon_{2nt})$ is concave increasing in $c_{nt}$ for any $z_{nt}$ and $\varepsilon_{2nt}$. We assume the following functional form:

$$u_{2}(c_{nt}, z_{nt}, \varepsilon_{2nt}) = \exp(z_{nt}'B_{4} + \varepsilon_{2nt})c_{nt}^{\alpha}$$

and $E(\varepsilon_{2nt} | z_{nt}) = 0$. $u_{1}(z_{nt}, 1 - h_{nt})$ is the utility from non-market hours consumed. Intuitively, $u_{1}$ is the utility from consuming a greater fraction leisure or spending a greater fraction of time on home production (which varies with current time spent at work as well as observed individual specific characteristics). The utility from non-labor market hours is also a function of spouse characteristics such as education, income. It depends on past non-market hours and there is interaction between past hours. This formulation allows us capture the element of experience accumulation in home production

$$u_{1}(z_{nt}) = z_{nt}'l_{nt}BZ + \theta_{0y}l_{nt}^{2} + \sum_{s=1}^{p}\theta_{sg}l_{nt}l_{nt-s}$$

### 2.2 Firms’ problem

Employers are infinitely lived. Firms (employers) compete over workers. There are $M$ occupations, $m = 1, \ldots, M$. There are costs to the employer when a new worker is hired (for example, specific training). These costs are specific to the employer and vary across occupation. Within occupations the costs are the same for all employers. We denote employer’s cost of hiring a
new employee (switching costs) at time $t$ in occupation $m$ by $\gamma_m$. There is a free entry into the market. The number of hours can change from period to period. We assume that there is a homogeneous product with price normalized to 1.

Let $H_{nt-1}$ denote the worker’s labor market experience history at the beginning of period $t$ (participation and hours worked in every period), i.e. $H_{nt-1} = [h_{n1}, h_{n2}, \ldots, h_{nt-1}]$. Let $z^p_n$ denote a vector of worker characteristics that affects production. Production function and costs are identical within occupations, but vary across occupations. Output in period $t$ in occupation $m$ is denoted by: $y_m(h_{nt}, H_{nt-1}, z^p_n)$. Employers maximize life-time expected discounted profits. Each occupation has a large number of identical firms that competes to firm workers.

2.3 Timing and information structure of the game

Let $D_{nt-1} = \{d_{mn1}, d_{mn2}, \ldots, d_{mnt-1}\}_{m=1}^M$ denote the history of employment in occupations. $D_{nt-1}$ and Labor market experience $H_{nt-1}$ are common knowledge. The utility function parameters and functional forms are common knowledge as well. But outside employers do not observe past wage history of a worker. Some factors affecting labor supply, $z_{nt}$, are unobserved by potential employers in the labor market. The variables which are the worker’s private information, for example includes spouse income and wealth, marital status and number of kids. These variables affect labor market participation decisions and labor supply. Denote by $z^*_n$ the worker’s characteristics which are publicly observable at time $t$ and do not affect production (such as education, cohort, gender, individual "fixed-effect"). Notice that $z^*_n$ is constant over time, and the only information that employers have which changes over time is the labor market experience and participation parameters. Worker’s consumption is not observed by the employer (knowledge of $z_{nt}$, and $\varepsilon_{0nt}$ is sufficient to recover consumption, but both are unknown to the employer). Notice that the worker’s "type" here is characterized by $z_{nt}$, and $\varepsilon_{0nt}$. Whereas $\varepsilon_{0nt}$ is iid, $z_{nt}$ evolves over according to a known process $F_{ga}(z_{nt+1}|z_{nt})$, and therefore, there is a correlation between types over time (for example, number of children, marital status, spouse characteristics). Thus, the worker has better information about the probability of remaining in the firm in the future (the employer knows $F_{ga}(z_{nt+1}|z^*_n)$).

Time-line

To be exact the time-line of the actions is as follows.

1. At the beginning of the each period workers observe an i.i.d. preference shocks (this information is unavailable to potential employers). Every period, all workers draw from the same distribution.

2. Employers post employment contracts (wages and hours). Contracts depend on workers observable characteristics

3. Observing available employment contracts, workers decide whether to participate in the

\footnote{This assumption is made to simplify the off-equilibrium path analysis and to reduce notation.}

\footnote{Notice, that the important assumption is that workers have information potential employers do not have, as opposed to current employer. The reason is that wages are determined by the offers made by potential employers.}

\footnote{We assume, however, complete information in the consumption contingencies market. Further discussion is in the section on the solution to the worker’s consumption choice.}
labor market and hours to supply if they participate according to the wage contracts available to them.

4. Production occurs.
5. Employers and workers observe aggregate permanent productivity
6. Consumption for the period occurs.
7. Workers and employers observe aggregate permanent shock to the cost of participation (utility shock).  
8. This structure repeats itself in each period.

2.4 Labor market

2.4.1 Employment contract

Workers and firms can only commit to (non-contingent) spot contracts. Each firm offers one contract each period and commits to hours (for the current period) before posting the contract. Output is a function of human capital, hours, and other characteristics, $z_t^p$. Occupations may vary in the hours they offer. We assume that a firm in occupation $m$ only offers a contract with hours $h$ for worker with experience $H_{t-1}$ and characteristics $z_t^p$ if there exist no other occupation $m'$ such that $f_{m'}(h|H_{t-1}, z_t^p) - \gamma_{m'} > f_m(h|H_{t-1}, z_{mt}^p) - \gamma_m$.

This is a "sorting" assumption that implies that occupations "specialize" in hours in which they have advantage over other occupations.

This assumption implies that any choice of hours $h_{at}$ of any particular worker with characteristics $H_{t-1}, z_t^p$ (education and skill) "maps" into a unique occupation choice in period $t$. That is, for any given characteristics that affect production, each occupation offers contracts of $h_{mt}(H_{t-1}, z_t^p) \leq h_t(H_{t-1}, z_t^p) \leq \overline{h}_{mt}(H_{t-1}, z_t^p)$.

The profit function of the firm is simply the expected profits from hiring an individual worker. Thus, an optimal contract offer can be solved separately for each job (hours and salary). Denote by $\pi_{tm}$ the value of a vacancy in occupation $m$ in period $t$. Define participation in a firm in occupation $m$ by the indicator $d_{amt}$, where $d_{amt} = 1$ if and only if $h_{mt}(H_{t-1}, z_t^p) \leq h_{at}(H_{n,t-1}, z_{nt}^p) \leq \overline{h}_{mt}(H_{t-1}, z_t^p)$, and $d_{amt} = 0$ otherwise.

2.5 Equilibrium Analysis

Strategies

Worker's strategy is a choice of consumption, participation, hours in every period, for every possible state. We denote a strategy, as a probability of participation and leisure given

---

9 For example, home production technology shocks.
10 A more realistic assumption is that firms can commit to long-term contracts but workers cannot. The main feature, that contracts do not fully screen workers in such framework, can be maintained. See also Dionne and Doherty (1994) for derivation of optimal renegotiation proof contract with semi-commitment in an adynamic adverse selection model.
11 The assumption does not change qualitatively the theoretical prediction of the model. It simplifies, however, the estimation stage substantially.
the worker’s type, labor market experience and observable characteristics by \( \sigma(h_{nt-1}, d_{nt-1} \mid z_{nt-1}, H_{nt-2}, D_{nt-2}, z^*_n) \), and consumption by \( c_{nt} \).

Firm’s strategy is a choice of hours, \( h_{nt} \in [\underline{h}_m(H_{nt-1}, z^*_n), \overline{h}_m(H_{nt-1}, z^*_n)] \) and salary \( S_{tm}(h_{tm}, H_{nt-2}, D_{nt-2}, z^*_n) \) in every period for any observable characteristics and history of a worker.

**Firm’s beliefs**

At the beginning of period of each period firms form a common prior beliefs above an individual workers type, i.e.
\[ \mu_t(z_{nt} \mid D_{nt-1}, H_{nt-1}, z^*_n) \]. At the beginning of end of each period firms update their beliefs, about the worker’s type observing the worker’s actions in the previous period. Let \( D_{nt-1} = \{d_{m1}, d_{m2}, \ldots, d_{mnt-1}\}_{m=1}^M \) denote the history of employment in occupations, and \( H_{nt-1} = \{h_{m1}, h_{m2}, \ldots, h_{mnt-1}\}_{m=1}^M \) the history of past hours. We denote posterior beliefs, which are formed at the end of period \( t \), after observing the worker’s decision by \( \tilde{\mu}_t(z_{nt} \mid D_{nt}, H_{nt}, z^*_n) \). Notice that \( \tilde{\mu}_t(z_{nt} \mid D_{nt}, H_{nt}, z^*_n) \) is used to form the posterior beliefs in period \( t + 1 \), as the types are correlated, and evolve according to the Markov process we specified. Therefore, the worker’s behavior in time \( t \) may convey information about the worker’s future type. We also define the firms’ beliefs at the beginning of period \( t \) about future participation by \( \tilde{\mu}_{t+1} \). The beliefs about future participation will be derived from the beliefs about the worker’s type. We further discuss beliefs formation below.

Next, we define and characterize the equilibrium.

**Definition 1** A Perfect Bayesian Equilibrium consists of strategies \( (c_n, \sigma_n^0, S^0_m(h, H, D, z^*), h_m) \) where \( \sigma_n^0 \) is the probability of participation and hours workers in every period for every type in every state, \( S^0_m(h, H, D, z^*) \), are the contract offered in each occupation in each period for any observable characteristics and history, and beliefs system such that
1). Each player’s strategy is optimal given beliefs and other player’s strategy
2). The posterior beliefs \( \tilde{\mu} \) satisfies Bayes’ rule when possible.\(^{12}\)
\[
\tilde{\mu}_t(z_{nt} \mid H_{nt}, D_{nt}, z^*_n) = \frac{\mu_t(z_{nt} \mid .)\sigma(h_{nt-1}, d_{nt-1} \mid .)}{\int \mu_t(z_{nt-1} \mid .)\sigma(h_{nt-1}, d_{mnt-1} \mid .)d\tilde{z}_{nt-1}}
\]
(2)

3). At the beginning of period \( t + 1 \) firms form priors about the worker’s type at that period based on past history (types changed exogenously)
\[
\mu_{t+1}(\tilde{z}_{nt+1} \mid H_{nt}, D_{nt}, z^*_n) = \frac{f_{z}(\tilde{z}_{nt+1} \mid z_{nt})\tilde{\mu}_t(z_{nt} \mid .)\sigma(h_{nt}, d_{nt} \mid .)}{\int f_{z}(\tilde{z}_{nt} \mid z_{nt-1})\tilde{\mu}_t(z_{nt} \mid .)\sigma(h_{nt}, d_{mnt} \mid .)d\tilde{z}_{nt+1}}
\]
(3)

4). The probability of participation in the proceeding period is
\[
\overline{p}_{mn,t+1}(H_{nt-1}, D_{nt-1}, z^*_n) = E_t[E_{t+1}(\sigma(h_{nt+1}, d_{mt+1} \mid z_{t+1} \cdot)) \mid H_{nt-1}, D_{nt-1}, z^*_n]
\]

Next we characterize the equilibrium strategies of the workers and firms.

\(^{12}\)The restriction on the beliefs are stronger than the usual sense as it applies to updating in histories which are reached with probability zero. See Definition 8.2 of Fudenberg and Tirole(1996) for the formal description of the conditions of equilibrium.
2.5.1 Solution to the worker’s problem.

We allow men and women to have different distribution of $z_{nt}$ and a different utility parameter $\theta_s$. We do not, however, impose these differences.\textsuperscript{13}

Let $\beta \in (0, 1)$ denote the common subjective discount factor, and write $E_t(.)$ as the expectation conditional on information available to individual $n$ at period $t$. The expected lifetime utility of individual $n$ is then:

$$E_t \left[ \sum_{r=t}^{T} \beta^{r-t} U_{nt} \right]$$

We first solve for the individual worker’s consumption problem. To provide a tractable solution to the model we assume that assets markets are competitive and complete (CCM). Here the word "competitive" is synonymous with price taking behavior and " complete markets" means that there are no frictions in the markets for loans, a common interest rate facing borrowers and lenders, and a rich set of financial securities exists to hedge against uncertainty. Altug and Miller (1990,1998) have utilized this condition for estimating both males and females consumption and labor supply with aggregate shocks. Other papers that discuss complete markets and estimate frameworks based on this assumption include, Card(1990), Mace(1991), Townsend(1994), Altonji, Hayashi and Kotlikoff(1996), Miller and Seig(1997), Gayle and Miller(2003), among others.\textsuperscript{14}

The CCM assumption allows us to compactly write the lifetime individual budget constraint. Complete market implies that individuals can condition their choice at time $t$ on information that is publicly available at that time and can purchase contingent claims to consumption that pays off in each state of the world. This assumption allows us to rewrite the workers’ budget constraint in each period as

$$E_0 \left\{ \sum_{t=0}^{T_n} \beta^t \lambda_t [c_{nt} - S_{nt}] \right\} \leq w_n$$

Where $S_{nt}$ is the labor market income of the individual, $\lambda_t$ is the expected price of the contingent claim and $w_n$ is a exogenously determined quantity denoting bequests net of inheritances. In any state, $\tau$, the price of a contingent claim is

$$\bar{\lambda}_t(\tau) = \int_{\tau} \lambda(\tau_t) g(\tau_t) d\tau$$

\textsuperscript{13}Wage gap, in our model can arise even if men and women are identical with all respects.

\textsuperscript{14}Whereas to some the assumption of complete markets might be controversial, it is empirically tractable and serves as a useful benchmark , which allows us to focus our analysis on the primary source of asymmetry in our model. A popular alternative to the complete market assumption is to put wages directly into the utility function, this is an even stronger assumption than complete markets and can only to justified under two very strong assumptions; 1) complete markets with no unobserved heterogeneity or 2) no markets to borrow or save. Hence we feel that at least by assuming complete markets we know where our restrictions on behavior are coming from.
The states are determined by realizations of \( \varepsilon_{2nt}, z_{nt} \) (recall that the densities are \( F_0(z_{nt+1} \mid z_{nt}) \); \( F_1(\varepsilon_{2nt}) \)). They are independent and \( g(t) \) is the joint density.

The aggregate feasibility condition equates the sum of labor income by all individuals and the aggregate resource endowment \( \omega_t \)

\[
\int_0^1 [c_{nt} - S_{nt}] d\mathbb{L}(n) \leq \omega_t \quad t \in \{0, 1, \ldots\}.
\] (6)

In this expression, \( \mathbb{L}(.) \) is the Lebesgue measure which integrates over the population. Let \( \eta_n \) denote the Lagrangian multiplier associated with the budget constraint in Equation (5), then the optimal consumption that satisfies the necessary conditions

\[
\frac{\partial u_2(c_{nt}, z_{nt}, \varepsilon_{2nt})}{\partial c_{nt}} = \eta_n \lambda_t \quad (7)
\]

for all \( n \in [0, 1] \) and \( t \in \{0, 1, 2, \ldots\} \). With the contemporaneous separability of the consumption from the labor supply choices, the condition in (7) can be used to solve for the individual Frisch demand functions, which determine consumption in terms of the time-varying characteristics \( z_{nt} \), the idiosyncratic shocks to preferences \( \varepsilon_{2nt} \), and the shadow value of consumption \( \eta_n \lambda_t \).

Characterizing the optimal labor market participation and hours of work decisions is more involved. The conditional value functions associated with the work decision of individual \( n \) at time \( t \) are the building blocks of this analysis. Note that each person’s labor supply contributes an infinitesimal addition to aggregate output, define the conditional valuation functions associated with the decisions to work or not as

\[
V_{1nt} + \varepsilon_{1nt} \equiv \max_{\{h_{nr}\}_{t=0}^{T}} E_t \left\{ \sum_{r=t}^{T} \beta^{r-t} \left[ d_{nr} u_0(z_{nr}) + u_1(z_{nr}, 1-h_{nr}) + d_{nr} \varepsilon_{1nr} 
+ (1 - d_{nr}) \varepsilon_{0nr} \right] + \eta_n \lambda_t S(h_{nr}, z_{nr}^*, H_{nr-1}) \mid h_{nt} > 0 \right\} \] (8)

and

\[
V_{0nt} + \varepsilon_{0nt} \equiv \max_{\{h_{nr}\}_{t=0}^{T}} E_t \left\{ \sum_{r=t}^{T} \beta^{r-t} \left[ d_{nr} u_0(z_{nr}) + u_1(z_{nr}, 1-h_{nr}) + d_{nr} \varepsilon_{1nr} 
+ (1 - d_{nr}) \varepsilon_{0nr} \right] + \eta_n \lambda_t S(h_{nr}, z_{nr}^*, H_{nr-1}) \mid h_{nt} = 0 \right\} \] (9)

Equation (8) is the value function for an individual who chooses to participate in period \( t \) and behave optimally thereafter while equation (9) is the value function for an individual who choose not to participate in the labor force in period \( t \) and behave optimally thereafter.

We next analyze the worker’s labor supply decision.

We will define by \( h_{nt}^o \) the optimal labor supply decision for individual \( n \) in period \( t \) and by \( h_{nt}^* \in (0, 1) \), the optimal interior solution of the labor supply decision for individual \( n \) and period \( t \). Similarly, we define by \( d_{nt}^o \) the optimal participation decision for individual \( n \) in
period t. Using the above formulation, the necessary condition for an optimal interior solution for labor supply is

$$\frac{\partial V_{1nt}}{\partial h_{nt}} = 0,$$

(10)

The necessary condition for the optimal participation decision is

$$d^o_{nt} = \begin{cases} 1 & V_{1nt} + \varepsilon_{1nt} \geq V_{0nt} + \varepsilon_{0nt} \\ 0 & \text{otherwise} \end{cases}.$$ (11)

The expected value of $d^o_{nt}$ conditional on the observed (to the researcher) state variables is the conditional choice probability of participating in the labor force and can written as

$$p_{nt} \equiv E[d^o_{nt} \mid H_{nt-1}, z_{nt}, \lambda_n \lambda_t] = \int_{-\infty}^{V_{1nt} - V_{0nt}} (\varepsilon_{0nt} - \varepsilon_{1nt}) dF(\varepsilon_{0nt}, \varepsilon_{1nt}, \varepsilon_{2nt})$$ (12)

This equation basically characterizes the participation decision. Next we turn to the characterization of the decision of number of hours to supply to the market conditional on participating. The optimal choice is formulated in the following problem:

$$h^*_nt = \arg \max_{h_{nt} \in (0,1)} \left\{ u_0(z_{nt}) + u_1(z_{nt}, 1 - h_{nt}) + \lambda_n \lambda_t S_{nt}(h_{nt}, z_{nt}^*, H_{nt-1}) + \beta E_t[p_{nt+1}V_{1nt+1} + (1 - p_{nt+1})V_{0nt+1}] \right\}$$ (13)

First, consider the optimal hours decision, $h^o(L_{nt-1}, H_{nt-1}, \{S^0_m(h, \mu_t)\}_{m=1}^M)$, workers make given the contracts, $S_m(h, H_n, D_n, \mu)_{m=1}^M$, the worker is facing. The worker will be maximizing expected lifetime utility, i.e.

$$\max_{\{d_{nt}, h_{nt}\}_{t=1}^{\infty}} E \sum_{t=1}^{\infty} \beta^{t-1} d_{nt} \left\{ u_1(l_{nt}, L_{nt-1}, z_{nt}, \{S^0_{nt}(h, \mu_t)\}_{m=1}^M) + \varepsilon_{1nt} \right\} + (1 - d_{nt})\left\{ u_0(1, L_{nt-1}, z_{nt}) + \varepsilon_{0nt} \right\} | z_{nt}, L_{nt-1}$$ (14)

Lets define the conditional valuation functions of working as

$$V_1(L_{nt-1}, z_{nt}, \{S^0_{nt}(h, \mu_t)\}_{m=1}^M) = \max_{h_{nt} \in (0,1)} \left\{ u_1(l_{nt}, L_{nt-1}, z_{nt}, \{S^0_{nt}(h, \mu_t)\}_{m=1}^M) \right\}$$

$$+ E_t \sum_{s=1}^{\infty} \beta^{s-1} d^o_{ns} \left\{ u_1(l_{ns}, L_{ns-1}, z_{ns}, \{S^0_{ns}(h, \mu_s)\}_{m=1}^M) + \varepsilon_{1ns} \right\} + (1 - d^o_{ns})\left\{ u_0(1, L_{ns-1}, z_{ns}) + \varepsilon_{0ns} \right\} | z_{nt}, L_{nt-1}, h_{nt} > 0$$ (15)
similarly

\[ V_0(L_{nt}, z_{nt}) \]

\[
= u_0(1, L_{nt-1}, z_{nt}) + E_t \left( \sum_{s=1}^{T} \beta^{s-t} d_{ns}^p u_1(l_{ns}, L_{ns-1}, z_{ns}, \{S^0_{mns}(h, \mu_s)\}_{m=1}^M) + \varepsilon_{1ns} \right) \\
+ (1 - d_{ns}^p)\{u_0(1, L_{ns-1}, z_{ns}) + \varepsilon_{0ns}\} | z_{nt}, L_{nt-1}, d_{nt} = 0 \]

(16)

By Bellman principle we have

\[ V_1(L_{nt-1}, z_{nt}, \{S^0_{mns}(h, \mu_s)\}_{m=1}^M) \]

\[
= \max_{h_{nt} \in (0,1)} \left[ u_1(l_{nt}, L_{nt-1}, z_{nt}, \{S^0_{mns}(h, \mu_s)\}_{m=1}^M) \\
+ \beta E_t\{E_{t+1} [d_{ns}^p | z_{nt+1}, L_{nt}] V_1(L_{nt}, z_{nt+1}, \{S^0_{mns+1}(h^{0}, \mu_{t+1})\}_{m=1}^M) | z_{nt}, L_{nt-1}, h_{nt} > 0 \} \right] \\
+ \beta E_t\{(1 - E_{t+1} [d_{ns}^p | z_{nt+1}, L_{nt}] V_0(L_{nt+1}, z_{nt+1}) | z_{nt}, L_{nt-1}, h_{nt} > 0 \} \]

(17)

Thus, the optimal hours worked conditional on deciding to participate is the solution to the first order conditions of the value function, conditional on participation:

\[
\begin{align*}
& u_1(l_{nt}, L_{nt-1}, z_{nt}, \{S^0_{mns}(h, \mu_s)\}_{m=1}^M) \\
& + \beta E_t\{\partial [E_{t+1} [d_{ns}^p | z_{nt+1}, L_{nt}] V_1(L_{nt}, z_{nt+1}, \{S^0_{mns+1}(h^{0}, \mu_{t+1})\}_{m=1}^M) \\
& - V_0(L_{nt}, z_{nt})]/\partial h_{nt} \}
\end{align*}
\]

\[
+ 0 = 0
\]

(18)

Notice that by choosing hours, the worker selects into occupation and a salary that is paid in this occupation for the given number of hours worked.

The participation decision is simply derived from a comparison between the value of working and the value conditional on not working , given the contracts workers are facing \\{\{S^0_{mns}(h, \mu_s)\}_{m=1}^M\} is

\[
\begin{align*}
d_{ns}^p(L_{nt-1}, z_{nt}, \{S^0_{mns}(h^{0}, \mu_s)\}_{m=1}^M) \\
= \begin{cases} 
1 & \text{if } V_1(L_{nt-1}, z_{nt}, \{S^0_{mns}(h^{0}, \mu_s)\}_{m=1}^M) + \varepsilon_{1nt} \geq V_0(L_{nt}, z_{nt}) + \varepsilon_{0nt} \\
0 & \text{otherwise}
\end{cases}
\end{align*}
\]

2.5.2 Solution to the firm’s problem

Next, we characterize the optimal contracts firms offer.

Define \( I_m(z_{nt+1}, H_{nt}, z^*_n) \) \( \equiv I \{ h^m_m(H_{nt}, z^*_n) \leq h_{nt+1}(z_{nt+1}, H_{nt}) \leq h^m_m(H_{nt}, z^*_n) \) to be an indicator function if the worker’s hours are offered in occupation \( m \), and zero otherwise, and denote the probability that type \( z_{nt} \) with history \( H_{nt-1}, D_{nt-1}, \) and observables characteristics \( z^*_n \) work in that period by \( Q_t(z_{nt, .}) = E[d_{nt} | z_{nt, .}] \).

\
Proposition 2 (Optimal Contract) 1. The optimal contracts of the firms are

\[ S_m(h_{nt}, H_{nt-1}, z^*_n) = y_m(h_{nt}, H_{nt-1}, z^*_n) - \gamma_m + \beta \gamma_m \pi_{mn,t+1}(H_{nt-1}, D_{nt-1}, z^*_n) \]

for all \( h_{nt} \in [\bar{h}_m(H_{nt-1}, z^*_n), \bar{h}_m(H_{nt-1}, z^*_n)] \), where \( \pi_{mn,t+1}(H_{nt-1}, D_{nt-1}, z^*_n) \) are beliefs that the worker will work in the firm in the proceeding period.

2. \( \bar{\pi}_{mn,t+1}(H_{nt-1}, D_{nt-1}, z^*_n) \) is a function of the beliefs of the workers type (for example that beliefs in period 3 \( \mu_3(z_{n3} | H_{n2}, D_{n2}, z^*_n) \) depend on the beliefs of the type in period 2 \( \mu_2(z_{n2} | H_{n1}, D_1, z^*_n) \) and is derived by Bayes’ rule)

\[ \bar{\pi}_{mn,t+1} = \int Q_{t+1}(z_{nt+1}, H_{nt}, h_{t+1}) I_{t+1}(z_{nt+1}, H_{nt}, h_{t+1}) \mu_{t+1}(z_{nt+1} | H_{nt}, D_{nt}, z^*_n)dz_{nt+1} \]

3. Off-equilibrium path: if for all \( z_{nt}(h^t) \), \( \sigma(h_{nt}, d_{mnt} | D_{t-1}, H_{t-1}, z_{nt}) = 0 \forall z_{nt} \) then \( \mu_t(z_{nt} | H_{nt-1}, D_{n-1}, z^*_n) = \mu \) if \( h_{nt} < \bar{h}_m \), and \( \mu_t(z_{nt} | H_{nt-1}, D_{n-1}, z^*_n) = \bar{\mu} \) if \( h_{nt} > \bar{h}_m \). That is, if a worker’s hours are above the highest hours a worker of his type works he is believed to be the type who works the most hours, and if he works less, he is believed to be the lowest type given the observable characteristics.\(^{15}\)

Before we proceed to characterization of the equilibrium contract we demonstrate the nature of this repeated adverse selection problem. First note that "types" are correlated over time and: \( z \) evolves stochastically according to the density \( F(z_{t+1}|z_t) \); further, there are random shocks to the cost of participation every period. Typical to this class of dynamic adverse selection models, when only spot contracts are feasible, workers will not report truthfully their probability of participation in the future. More information is revealed over time and, in general, there is "semi-pooling". The payoff from working more hours today increases the future payoff through increase in experience and workers who are more likely to participate have higher future "returns" to current hours. In the symmetric information environment, workers who are more likely to participate will work more. Under asymmetric information, consider a worker who is less likely to participate next period, but has the same observable characteristics as a worker who is more likely to participate; the beliefs about future participation are based on publicly observable characteristics, and therefore, the increase in current salary for working more hours for the type with low probability of participation is now bigger than the increase under the symmetric information. This creates incentive to "pool". At the same time the current salary of the more attached worker is lower compared to the symmetric information case decreasing her incentive to work long hours. Secondly, the reward for providing more hours for a worker with the lower probability of future participation increases (compared to the full information case) because all future beliefs in different states about participation increase in the asymmetric information model. Therefore, the sequence of future wages increases as well. Workers who’s disutility from working is very high and have low probability of participation, however, will still choose to work less hours.

\(^{15}\)Note that the prior belief \( \bar{\mu}(z_{nt}|.) \) is always strictly positive. This is because given any possible history, \( F(z_{nt+1}|z_{nt}) > 0 \forall z_{nt+1}, z_{nt} \)
The above propositions characterize the optimal strategies of workers (given the available contracts), and the firm’s optimal contracts given their beliefs. Because of the free entry assumptions and the competitive nature of the labor markets all offers yield zero expected payoffs over time. The costs of hiring new employee imply that the beliefs on future participation enter the wage equation. These beliefs, in equilibrium, satisfy Bayes’ law and are consistent with worker’s labor supply behavior.

Next, we show existence. In order for an equilibrium to exist, one need to show that there exist solution to the worker’s problem given current and future expected earnings. Then we show that there exist beliefs about future participation that are self-fulfilled. Self-fulfilling beliefs mean that the expected probability of workers who accept a contract (given information firms currently have) will remain in the firm in proceeding period is indeed the proportion of workers who remain in the firm. Notice that firms and workers are forward looking: workers consider the effect of current choices on future earnings and firms predict the probability workers will remain in the firm in all future periods (until retirement). Therefore, there are system of equations in which the self-fulfilling beliefs are solved simultaneously.

**Proposition 3 (Existance) .**

1. For any state variables, there exist a unique solution the worker’s problem.
2. There exist \( \{ \tilde{p}_{mn,2}(h_1, z^*_n), ..., \tilde{p}_{mn,T}(z^*_n, H_{nT-1}, D_{nT-1}) \} \) which satisfy the equilibrium conditions

(proof is in the Appendix)

**2.6 Implications**

The model exhibits "discriminatory equilibria".\(^{16}\) Whereas, we allow men and women to be different, our model gives rise to discriminatory equilibria even if men and women are initially identical (same distribution of preferences and skills) in all aspects which affect labor supply, participation and production. As a benchmark, we describe these equilibria below. Suppose employers have different beliefs about likelihood of future labor market participation and labor supply; women can face lower wages than men (conditional on all other relevant characteristics). These beliefs are self-fulfilling in equilibrium and cause women and men to make different participation and labor supply decisions. One feature that generates the discriminatory equilibria in our model is the non-separability in leisure in the utility function (captured by \( \theta \)). Suppose \( \theta > 0 \).\(^ {17}\) That is, there is complementarity between "non-market" hours across time. Our model is also consistent with Becker (1965), and a story of home

---

\(^{16}\)See Tirole (1997) for a dynamic adverse selection and statistical discrimination. The difference between the model is because the matching in Tirole between firms and workers is random. In our framework, workers select into "contracts" offering different number of hours and earnings.

\(^ {17}\)We do not, however, impose this restriction on the empirical model. Instead, we estimate the utility function parameters and verify that there is complementarity in leisure. Whereas gender wage gap can occur when the two groups are ex-ante similar, differences in the utility function parameter (higher \( \theta_s \)), imply that for given wages, labor supply of women is smaller. Such utility differences may arise for example, if women’s productivity in home good is higher.
production division model (the statistical discrimination mechanism is similar to Coate and Loury (1993)). In particular, if married women face lower wages, the efficient division of home production hours is that women put more hours at home. The solution to the decision problem in which the worker decides whether to participate in the labor force and how many hours to work, if participate. This decision depends on the returns to working in the labor market (current an future expected earnings). If systematically, wages are lower for women, they may accumulate, on average, less labor market experience and more "home production" hours. Therefore, in equilibrium, employers beliefs on labor market participation are correct.

**Proposition 4.** Now suppose that the distribution of skills and preferences are identical for men and women. If the employers’ believe that the probability of staying in the firms stay in the proceeding periods is lower for women than for men That is, for some \( z_{nt} \) and \( m \)

\[
p_{1f}(z_{nt}^*) < p_{1f}(z_{nt}),
\]

then women face lower wages than men. These wage differences are larger in occupations with large \( \gamma_m \). Beliefs are self-fulfilling in equilibrium, and therefore women work less hours and are less likely to participate.

The cost of hiring new employees are different across occupations. If the initial beliefs \( p_{1m} \) for men are lower than for women, then in equilibrium women face wages and expected future wages (over time and across the different states) which are lower than the wages men face. In equilibrium, these beliefs are self-fulfilling and generate different labor market histories for the two groups.

The Bayesian updating formula shows persistence of initial beliefs. Furthermore, these beliefs are self-fulfilling in equilibrium and, thus, create different experience and occupation choice among the two groups.

In particular, the optimal hours are characterized by the following Euler equation:

\[
h^*_n = \underset{h_{nt} \in (0,1)}{\text{arg max}} \left\{ u_0(z_{nt}) + u_1(z_{nt}, 1 - h_{nt}) + \lambda_n w_n(h_{nt}, z_{nt}, H_{nt-1})h_{nt} \right. \\
+ \beta E_t[p_{nt+1}V_{1nt+1} + (1 - p_{nt+1})V_{0nt+1}] \right\}
\]

(19)

Notice that for two "identical" workers with the same \( z_{nt} \), who differ only by gender, the first two elements in the equation are similar, but the last two elements are different. The change in beliefs

\[
\frac{\partial p_{nt}(h_{nts}, H_{Tnt-1}, z_{nt}^*)}{\partial h_{nt}}
\]

may be gender specific and it is part of the current earnings. Importantly, the last element: \( E_t \left\{ \frac{\partial V_{nt+1}}{\partial h_{nt}} \mid z_{nt}, h_{nt} = h_{nt}^0 \right\} \) which is the change in the continuation value from marginal change in hours workers may be different. If the returns to labor market experience are sufficiently lower for women, because they face lower steam of wages across states and time, then in order for the FOC to hold the current marginal dis-utility from working \( \frac{\partial u_1(z_{nt}, 1 - h_{nt})}{\partial h_{nt}} \) is smaller.

\[\text{We do not, however, observe hours work at home in the data and therefore do not model it explicitly.}\]
This is intuitive: the current salary for any level of hours is higher for men, and working more hours increase labor market experience which is rewarded when a worker is planning to work in the future. However, it is possible that for some workers in some states

\[
\frac{\partial p_{mnt}(h_{nts}, H_{Tnt-1}, z^{*}_{nt})}{\partial h_{nt}}
\]

is high and reduces the effect of the lower continuation value.

Whereas discrimination can be a result of pure coordination failure (in which case, if there exist a unique solution given fixed beliefs than if there is no multiplicity the two groups have the same outcomes), our model may exhibit discriminatory equilibrium due to a cross-group (gender) effects (Moro and Norman (2004) analyze a statistical discrimination model in which a discriminatory equilibrium may arise as a result of complementaries in production). That is, we have complementaries in the utility function. Once there exist complementarity between the hours (participation) women work and men (via the utility function), there could be a discriminatory equilibrium (asymmetric equilibrium) even if there is no coordination failure. If men are believed to participate more and earn more, women (married) have higher consumption and work less. This affect not only married women because the model is dynamic and single women may marry in the future, thus, it affect them through expectations.

Next, we discuss some implications of the model.

**Corollary 5** If women work less in equilibrium, they will sort into occupations which offer higher returns to working "part" time, reward less for labor market experience, and have lower costs to hiring new workers.

Occupation \( m \) "specialize" and offer hours \( h_t \) for given worker’s characteristics if no \( m' \) occupation satisfies:

\[
f_{m'(t|H_{t-1}, z^{p}_{t})} - \gamma_{m'} < f_{m(t|H_{t-1}, z^{p}_{nt})} - \gamma_{m} \forall m' \neq m.
\]

Therefore, women are more likely to choose occupations with "high returns" on part-time jobs (for example). Furthermore, wages in occupations with lower costs of hiring new employees, will have smaller differences in wages for men and women with the same observable characteristics, and gender wage gap in this occupations may be smaller. To sum, if employers believe that women are less likely to remain in the firm in the future, women may lower labor market experience, and thereafter, they will sort into occupation \( m \) that have lower costs of hiring workers and lower returns to experience.

**Corollary 6** The effect of the initial beliefs on wages declines over time. Thus, for a given cohort, conditional on all observable characteristics (to the employer), the gender wage gap (for men and women with the same observable characteristics) declines with experience.

This is an immediate implication of the Bayesian learning. Over time, more information about labor market participation and labor supply arrives, and therefore, the effect of the initial beliefs becomes smaller. The only wage component that generates difference in wages for equally productive men and women (with the same employment history) is the beliefs on future participation.
Changes in the wage gap

Corollary 7 According to our model, the following changes in exogenous (in the model) factors could account for the narrowing in the observed gender wage gap over time:

1. Differences in education across the different cohorts
2. Occupation-specific aggregate productivity shocks
3. Demographic changes which affect the distribution: \( F(z_{nt+1}|z_{nt}) \)
4. Changes in costs of production of domestic goods

The first factor explains differences across cohorts. Over time educational attainment of women increased, and therefore, beliefs about labor market participation of women increases. The other changes can occur within cohorts. Suppose that there is an increase in overall productivity within an occupation. The increase affects wages of all workers, but if men participation is already high, beliefs about women participation may increase women’s wages relative to men’s wages. This increase in wage will result in a bigger increase in labor supply and participation of women. The third factor, changes in demographics (such as decline in fertility and increase in divorce rate) affect beliefs about future participation. Lastly, if home production becomes cheaper over time (due to technological changes), it will make the cost of participation in the labor market smaller, and may increase participation. Since women are less likely to participate than men, changes in costs of participation may affect the relative wage because changes of beliefs about participation will raise women’s wages more than men’s wages.

3 Identification and Estimation

There are two types of multiple equilibria in this model which we have to consider before taking it to data. The first is the standard multiple equilibria that arise because it is a model of adverse selection, which means that a given equilibrium is supported by specific off equilibrium beliefs and different off equilibrium beliefs may give rise to a different equilibrium. We never observed off equilibrium behavior but the earnings equation holds for any beliefs of histories that are off-equilibrium path. Hence we will be estimating a particular equilibrium outcome observed in the data. The second type of multiple equilibria is more central to the estimating of our model. Statistical discrimination means that employers choose to believe that one group is less attached that other and these beliefs are then self fulfilling in equilibrium. For any given belief there is a different equilibrium. In the standard statistical discrimination model, an identification problem of the following form arises, given the observed wage distribution and human capital investment, an econometrician is trying to recover the preference parameters and the beliefs about an unobserved variable that affects productivity. The source of the identification problem is that there may be more than one combination of preferences parameters and beliefs that could have generated the observed wage distribution (see Moro and Antonovics for detail of
this problem). Our model, however, does not have that problem. This is because our beliefs is about next period participation probability conditional on observed characteristics. Panel data, however, allows us to observe next period participation decision of the individual, this have to be correct across the population conditional on the observed characteristics. Therefore we can nonparametrically identify these beliefs if the following assumption holds:

A2: (Equilibrium Selection): The data for each age-education cohorts is generated by only one equilibrium.

This assumption is standard in the literature for estimating dynamic discrete games. It only rules out the possibility that for any given age-education cohort the data is generated by a mixture of two or more equilibria. It does not select any one equilibrium or restricted the type of equilibrium played across age-education cohort.

This reduces our identification problem to recovering the preferences and technology parameters given the observed choices, conditional wage distribution and labor supply choices. As shown in Gayle and Miller (2006) these model are fully nonparametrically identified.

3.1 Outline of Estimation Strategy

Estimation proceeds in three steps. In the first step, we estimate the earnings and consumption equations using a panel data Generalized Method of Moment (GMM) estimation strategy. In the second step, we estimate the conditional choice probability and the firm’s beliefs using a kernel density based nonparametric estimation strategy. In the third step, we use the estimates of the marginal utility of wealth, production functions and switching cost from the first step estimation, along with estimates of the conditional choice probabilities and the firms’ beliefs in a set of moment conditions derived from the workers’ optimization problem. We then minimize this GMM criterion function to obtain estimates of the utility function. We will describe the estimation and discuss identification in further details in what follows below.

Before we describe in details the three steps of estimation in details, we begin with a background description. Let $L_{nt-1}$ denote the entire work history of worker $n$ at the beginning of period $t$. The optimal participation decision of the workers given the contracts the workers face at any state,

$$d^o_{nt}(L_{nt-1}, z_{nt}, \{S^o_{mt}(h^o_{nt}, \mu_t)\}_{m=1}^M, \varepsilon_{0nt} \varepsilon_{1nt}) = \begin{cases} 1 & \text{if } V_1(L_{nt-1}, z_{nt}, \{S^o_{mt}(h^o_{nt}, \mu_t)\}_{m=1}^M) + \varepsilon_{1nt} \geq V_0(L_{nt}, z_{nt}) + \varepsilon_{0nt} \\ 0 & \text{otherwise} \end{cases}, \quad (20)$$

implies that the equilibrium choice probabilities of working $p(d^o_{nt} | \omega_{nt})$ must satisfy:

$$p(d^o_{nt} | \omega_{nt}) = \Pr\{V_1(\omega_{nt}) + \varepsilon_{1nt} \geq V_0(\omega_{nt}) + \varepsilon_{0nt}\}. \quad (21)$$

The state variable $\omega_{nt}$ in all observed state to the econometrician, which in includes the firms beliefs that the worker will remain in the firm next period. A very useful insight of the seminal work of Hotz and Miller (1993) applies to our model: there is a one-to-one relationship between...
the equilibrium choice probabilities and the difference between the choice-specific value function $V_1(\omega_{nt}) - V_0(\omega_{nt})$.\footnote{This equation is central to estimation in a number of papers including, Hotz, Miller, Sanders and Smith(1993), Altug and Miller(1998), Aguirregabiria and Mira(2003), Gayle and Miller(2003), Pesendorfer and Schmidt-Donglar(2003), Bajari, Benkard and Levin (2004), Pakes, Ostrovsky and Berry (2004), Bajari and Hong( 2005), among others.} We let $Q : R \rightarrow (0, 1)$ denote the mapping from the choice-specific value function to the conditional choice probabilities, i.e.

$$p(d^o_{nt} \mid \omega_{nt}) = Q(V_1(\omega_{nt}) - V_0(\omega_{nt}))$$

(22)

and the inverse exist and gives

$$V_1(\omega_{nt}) - V_0(\omega_{nt}) = Q^{-1}(p(d^o_{nt} \mid \omega_{nt})).$$

(23)

This mapping, $Q(.)$, is only a function of the unobserved state variables, $\varepsilon_{0nt}$, and $\varepsilon_{1nt}$. Proposition 1 of Hotz and Miller(1993) also states that there exist a mapping $\varphi_k : [0, 1] \rightarrow R$, which measures the expected value of the unobservable in the current utility, conditional on action $k \in \{0, 1\}$:

$$\varphi_k(p(\omega_{nt})) \equiv E[\varepsilon_{knt} \mid \omega_{nt}, d^o_{nt} = k].$$

(24)

Given these two results we can obtain the difference between the two conditional valuation functions once we know the distribution function of the unobserved state variables. However, we are interested in reformulating the problem as just a function of the primitives of our original problem, i.e., the utility functions and the equilibrium wage equation. Hence, we need in addition a representation of our conditional valuation functions as a function of only the model’s primitives and the conditional probabilities. To do this we need the following assumption which states that the production and utility functions only depend on the most recent work history.

**A1**: (Finite State Dependence): The production and utility functions only depend on a finite history of past work decision, i.e. $H^p_{nt-1} = (h_{nt-\rho}, ..., h_{nt-1})'$.

This assumptions states that the payo\-ff relevant history is finite. However, it still leaves the possibility that the information relevant history includes the workers’ complete work history. Note that the firm’s beliefs is an aggregator of information, and hence helps to reduce the state space for the problem. Since the information relevant history only enter into the firms beliefs, and the beliefs are an aggregator that enters into the workers state space, the workers state space has finite dependence also.

First define vectors $\omega_{ont}^{(r)}$ and $\omega_{1nt}^{(r)}$, as

$$\omega_{ont}^{(r)} = (h_{nt-\rho+r}, ..., h_{nt-1}, 0, ..., 0, z'_{nt+r}, \tilde{p}_{1n,t+1+r}, \ldots, \tilde{p}_{Mn,t+1+r})',$n

(25)

and

$$\omega_{1nt}^{(r)} = (h_{nt-\rho+r}, ..., h_{nt-1}, h^s_{nt}, 0, ..., 0, z'_{nt+r}, \tilde{p}_{1n,t+1+r}, \ldots, \tilde{p}_{Mn,t+1+r})'.$n

(26)
for \( r = 0, \ldots, \rho \). The vector \( \omega_{ont}^{(r)} \) is the state for an individual at date \( t + r \) who has accumulated the work history \( (h_{nt-\rho+r}, \ldots, h_{nt-1})' \) up to period \( t \), and then chooses not to participate at date \( t \) and for \( r - 1 \) periods following period \( t \). Similarly, \( \omega_{ont}^{(r)} \) is the vector for an individual who has accumulated work history \( (h_{nt-\rho+r}, \ldots, h_{nt-1})' \) up to period \( t \) and chooses to participate and supply hours \( h_{nt}^* \) at date \( t \), but chooses not to participate for \( r - 1 \) periods following period \( t \).

A restatement of proposition 1 of Altug and Miller (1998) pp. 64 in our context gives us the following representation. Suppose we define the primitives of our problem as follows:

\[
U_k(\omega_{nt}) = \begin{cases} 
  u_1(z_{nt}, 0) & \text{for } k = 0 \\
  u_0(z_{nt}) + u_1(z_{nt}, 1 - h_{nt}^*) + \eta_n \lambda_t S(z_{nt}, h_{nt}^*, \bar{p}_{nt+1}) & \text{for } k = 1
\end{cases}
\]

and let \( p_{knt}^{(r)} \equiv \Pr(d_{nt+1}^{(r)} = 1 | \omega_{knt}^{(r)} = \omega) \) for \( k \in \{0, 1\} \) and \( r = 0, \ldots, \rho \). Then for \( k \in \{0, 1\} \), the conditional valuation functions can be expressed as

\[
V_k(\omega_{nt}) = U_k(\omega_{nt}) + E_t\left\{ \sum_{r=1}^{\rho} \beta^r [U_0(\omega_{knt}^{(r)}) + \varphi_0(\omega_{knt}^{(r)}) + p_{knt}^{(r)}(Q^{-1}(p_{knt}^{(r)}) - \varphi_1(\omega_{knt}^{(r)}) - \varphi_0(\omega_{knt}^{(r)}))] + \beta^{r+1}[V_0(Z_{knt+1}) + \varphi_0(\omega_{knt+1}) + p_{knt}^{(r+1)}(Q^{-1}(p_{knt}^{(r+1)}) + \varphi_1(\omega_{knt}^{(r+1)}) - \varphi_0(\omega_{knt}^{(r+1)}))]ight\}
\]

The difference between the conditional valuation functions that characterize the participation decision does not depend on the unknown valuation function any more since \( \omega_{nt}^{(r+1)} \) for \( k \in \{0, 1\} \) are both equal to \((0, \ldots, 0, z_{nt+\rho}^*, \bar{p}_{1n,t+1+\rho}, \ldots, \bar{p}_{Mn,t+1+\rho})' \), hence the Hotz-Miller proposition shows that the conditions for optimal participation decision depends on state probabilities and payoffs for \( 2\rho \) dates into the future. Note that the valuation function now only depends on the utility function, the distribution of the utility shocks and conditional choice probabilities in finite number of counterfactual states. We can now apply an estimation strategy that does not require the computation of the valuation functions and hence does not have a problem with the possibility of multiplicity of equilibria. Below we outline our estimation strategy.

### 3.2 First Step: Estimation of Consumption and Earnings Equations

Suppose the econometrician has access to a panel data of a age-education of \((n = 1, \ldots, N, t = 1, \ldots, T)\) cohort of individuals. During each time period, the econometrician observes the actions and states variable except any idiosyncratic shocks to utility or any individual specific-time invariant state variables \((S_{nt}, d_{nt}, d_{nmt}, h_{nt}, c_{nt}, z_{nt}, z_{nt}^*)\). Note that we are assuming that the econometrician observes the private information \( z_{nt}^* \).

In the first step we use the Euler equation for consumption(7) to form the moment condition:

\[
E \left[ \frac{\partial u_2(c_{nt}, z_{nt}, e_{2nt}, \theta_e)}{\partial c_{nt}} - \eta_n \lambda_t z_{nt} \right] = 0
\]

(29)
Here we are assuming that the functional form of $u_2()$ is known up to a finite dimensional parameter vector $\theta_c$. Based on equation(29), the econometrician can estimate $\theta_c$ and $\eta_n\lambda_t$ up to a proportional constant. In fact we assume that

$$u_2(c_{nt}, z_{nt}, \varepsilon_{2nt}, \theta_c) = \exp(z_{nt}^T B_4 + \varepsilon_{2nt}^T c_{nt}^0/\alpha)$$

Let $\Delta$ denote the first-difference operator. Taking logarithms of both sides of this expression, differencing and rearranging implies

$$(1 - \alpha)^{-1} \Delta \varepsilon_{2nt} = \Delta \ln(c_{nt}) - (1 - \alpha)^{-1} \Delta z_{nt}^T B_4 + \Delta (1 - \alpha)^{-1} \ln(\lambda_t)$$

Let $\Theta_c$ denote the $(K + T - 1)$ dimensional vector of parameters to be estimated, defined:

$$\Theta_c = \begin{pmatrix}
(1 - \alpha)^{-1} B_4 \\
\Delta (1 - \alpha)^{-1} \ln(\lambda_2) \\
\vdots \\
\Delta (1 - \alpha)^{-1} \ln(\lambda_T)
\end{pmatrix}$$

We also define $Y_n = (\Delta \ln(c_{n2}), ..., \Delta \ln(c_{nT}))'$ as a vector of endogenous variables, and $Z_n$ the exogenous variables as

$$Z_n = \begin{bmatrix}
\Delta z_{nt}^T & D_2 & ... & 0 \\
... & ... & ... & ...
\end{bmatrix}$$

where $D_t$ denotes a time dummy for $t \in \{2, \ldots, T\}$. The assumptions in Section 2 imply that the unobserved variable $\varepsilon_{2nt}$ is independent of individual specific characteristics. Therefore $E((1 - \alpha)^{-1} \Delta \varepsilon_{2nt} | z_{nt}) = 0$. Substituting for $(1 - \alpha)^{-1} \Delta \varepsilon_{4nt}$ using equation(??) one can obtain a set of orthogonality conditions:

$$E\left[ (Y_n - Z_n \Theta_c) Z_n \right] = 0$$

which can be exploited here to estimate $\Theta_c$ using a optimal instrumental variable estimation technique.

We use a traditional fixed effect estimator to estimate $(1 - \alpha)^{-1} \ln(\eta_n)$. Let $T_1$ be the number of time periods for which the marginal utility of consumption equation is estimated. Let:

$$(1 - \alpha)^{-1} \ln(\eta_n) \equiv \sum_{t \in T_1} [\ln(c_{nt}) - (1 - \alpha)^{-1} z_{nt}^T B_4 + (1 - \alpha)^{-1} \ln(\lambda_t)]/T_1$$

The fixed effects estimates of $(1 - \alpha)^{-1} \ln(\eta_n)$ are obtained as the simple time averages of the estimated residuals of the consumption, which correspond to sample counterparts of $(1 - \alpha)^{-1} \ln(\eta_n)$ defined above. In order to form the sample counterpart of (31) we need an estimate of $\{ (1 - \alpha)^{-1} \ln(\lambda_t) \}_{t=1}^{T_1}$, from the estimate of $\Theta_c$, however, we can only obtain estimates of $\{ \Delta(1 - \alpha)^{-1} \ln(\lambda_2) \}_{t=2}^{T_1}$. This necessitates us making an additional assumption that $E_n[\eta_n |
We now have estimates of \( \{ (1 - \alpha)^{-1} \ln(\lambda_t) \}^{T_1}_{t=1} \) and \( (1 - \alpha)^{-1} \ln(\eta_n) \). We are then able to recover \( \alpha \) in the third step of our estimation.

In the first step we also use the free entry condition from Proposition 1 (Optimal Contract) to form the following orthogonality conditions

\[
E_t\{d_{nmt}[S_{nt} - y_m(h_s, H_{s-1}, z_{nt}^p, \theta_e) + \gamma_m - \beta \gamma_m d_{n,m,t+1}]|z_{nt}^*, H_{n,t-1}, h_{nt}^*\} = 0. \tag{32}
\]

Again we are assuming that the functional form of \( y_m() \) is known up to a finite dimensional parameter vector \( \theta_e \). Based on equation (32), the econometrician can estimate \( \theta_e \) and \( \beta \gamma_m \).

We assume the following functional form for our production function,

\[
y_m(h_s, H_{s-1}, z_{nt}^p, \theta_e) = \pi_{0nt} + \pi_{m1}h_{nt} + \pi_{m2}h_{nt}^2 + \sum_{r=1}^{n} \pi_{m3r}h_{nt-r} + z_{nt}^*B_{m5} + \nu_{nt}
\]

This production function allows for occupation specific aggregate shocks and human capital is general it’s rate of return is different across occupation. There is an unobserved individual specific component, \( \nu_{nt} \), which is completely general in nature. Since all the information set in equation (32) is public at period \( t \) then we have

\[
E_t\{d_{nmt}d_{nmt-1}[\Delta S_{nt} - \Delta \pi_{0nt} - \pi_m \Delta HC_{nt} - \Delta z_{nt}^*B_{m5} - \beta \gamma_m \Delta d_{n,m,t+1}]|z_{nt}^*, H_{n,t}, h_{nt}^*\} = 0. \tag{33}
\]

where \( \Delta HC_{nt} = (\Delta h_{nt}, \Delta h_{nt}^2, \Delta h_{nt-1}, \ldots, \Delta h_{nt-\rho})' \) and \( \pi_m = (\pi_{m1}, \pi_{m2}, \pi_{m31}, \ldots, \pi_{m3p}) \).

Let \( \Theta_{em} \) denote the \((2 + K + \rho + T)\) dimensional vector of parameters to be estimated, defined:

\[
\Theta_{em} = \left( \begin{array}{c}
\pi_m \\
B_{m5} \\
\beta \gamma_m \\
\Delta \pi_{0m2} \\
\vdots \\
\Delta \pi_{0m2}
\end{array} \right)
\]

We also define \( Y_{mn} = (d_{nm2}d_{nm1} \Delta S_{n2}, \ldots, d_{nmT}d_{nmT-1} \Delta S_{nT})' \) as a vector of endogenous variables, and \( X_n \) the exogenous variables as

\[
X_{mn} = \left[ \begin{array}{ccc}
\Delta x_{m2}' & D_2 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots \\
\Delta x_{MT}' & 0 & \cdots & D_T
\end{array} \right]
\]

where \( \Delta x_{mnt} = d_{nmt}d_{nmt-1}(\Delta h_{nt}, \Delta h_{nt}^2, \Delta h_{nt-1}, \ldots, \Delta h_{nt-\rho}, \Delta z_{nt}^*, \Delta d_{n,m,t+1}) \). Let \( Z_n \) be the
vector of conditioning variable

\[ Z_n = \begin{bmatrix} z'_{n2} & H_{n2} & h_{n2} \\ \vdots & \vdots & \vdots \\ z'_{nT} & H_{nT} & h_{n2T} \end{bmatrix} \]

and using equation (33) one can obtain a set of orthogonality conditions:

\[ E \left[ (Y_{mn} - X_{mn}\Theta_{em}) Z_n \right] = 0 \]

which can be exploited here to estimate \( \Theta_{em} \) using an optimal instrumental variable technique similar method to the regression procedures that estimated the consumption function. The aggregate effect and fixed effect in the earnings equation is estimated in a similar way to those in the consumption equation.

### 3.3 Second Step: Nonparametric Estimation of Choice Probabilities and Beliefs

The equilibrium earnings equation and the alternative representations for the Euler and participation conditions require estimates of the beliefs and the conditional choice probabilities. First, \( \tilde{p}_{mnt} \) are computed as nonlinear regression of next period participation index \( d_{mnt+1} \) on today’s public information variables \( z'_{n'} \) and work histories \( H_{nt-1} \) and hours worked, \( h_{nt} \), conditional on working today in occupation \( m \). Let \( X_{nt} = (z'_{nt} H_{nt-1}, h_{nt}, \nu_n) \) and define \( J_1(\delta^{-1}_{1N}(X_{nt} - X_{n's})) \) as the normal kernel, \( \delta_N \), is the bandwidth associated with each argument. The nonparametric estimate of \( \tilde{p}_{mnt} \) denoted \( \tilde{p}^N_{mnt} \) is computed using the kernel estimator

\[ \tilde{p}^N_{mnt} = \frac{\sum_{n'=1, n' \neq n}^N \sum_{s=1}^{T-1} d_{mnt's} d_{mnt's+1} J_1(\delta^{-1}_{1N}(X_{nt} - X_{n's}))}{\sum_{n'=1, n' \neq n}^N \sum_{s=1}^{T-1} d_{mnt's} J_1(\delta^{-1}_{1N}(X_{nt} - X_{n's}))}. \] (34)

The derivative is then estimated using the standard nonparametric derivative kernel estimator (see Pagan and Ullah(1999)). Theoretically, \( H_{nt-1} \) includes the complete working history of the worker, including number of hours worked and occupation. In practice, for older workers the history could be very large, and therefore, our estimator would suffer from the curse of dimensionality. In order to obtain the optimal history length for information we use a cross validation procedure to find the optimal information history dependence.

The conditional choice probability \( p_{nt} \) are computed as nonlinear regressions of the participation index \( d_{nt} \) on the current state \( Z^N_{nt} \equiv (z'_{nt}, H_{nt}, w_{nt}, \lambda^N_{nt})' \), where the N superscript denotes estimated quantity. Define \( J(\delta_N(Z^N_{nt} - Z^N_{n's})) \) as the normal kernel, \( \delta_N \), is the bandwidth associated with each argument. The nonparametric estimate of \( p_{nt} \) denoted \( p^N_{nt} \) is computed using the kernel estimator

\[ p^N_{nt} = \frac{\sum_{n'=1, n' \neq n}^N \sum_{s=1}^{T} d_{n't's} J(\delta^{-1}_{1N}(Z^N_{nt} - Z^N_{n's}))}{\sum_{n'=1, n' \neq n}^N \sum_{s=1}^{T} J(\delta^{-1}_{1N}(Z^N_{nt} - Z^N_{n's}))}. \] (35)
The conditional choice probabilities \( p_{knt}^{(r)} \) are also computed as nonlinear regressions of a participation index on the appropriate state variables. Define
\[
d_{knt}^{(r)} = [1 - (1 - k)d_{nt-r} - k(1 - d_{nt-r})] \prod_{t=1}^{r-1} (1 - d_{nt-t}), \quad k \in \{0, 1\}
\]
Notice that \( d_{nt}^{(r)} = 1 \) if the person participated at \( t - r \) but then did not participate for \( r - 1 \) periods. Similarly, \( d_{0nt}^{(r)} = 1 \) if the person did not participate between \( t - r \) and \( t - 1 \). Thus, \( d_{knt}^{(r)} \)

The conditional choice probabilities \( p_{knt}^{(r,N)} \) are computed as
\[
p_{knt}^{(r,N)} = \frac{\sum_{n'=1}^{N} \sum_{n''=n}^{T} d_{n's} d_{knt}^{(r)} J[\delta_N^{-1}(Z_{knt}^{(r,N)} - Z_{n's})]}{\sum_{n'=1}^{N} \sum_{n''=n}^{T} d_{knt}^{(r)} J[\delta_N^{-1}(Z_{knt}^{(r,N)} - Z_{n's})]} \tag{36}
\]
where \( Z_{knt}^{(r,N)} = (H_{knt}^{(r)}, z_{nt+1}'r, \tilde{p}_{nt+1+r}'r, ..., \tilde{p}_{nt+Mt+r}'r, \eta_n \lambda_{t+r})' \) for \( k \in \{0, 1\} \) is the counter-factual state vector for individual \( n \).

To evaluate the term \( \partial p_{1nt}^{(r)} / \partial h_{nt} \) which appears in the Euler equation, define
\[
f_{1nt}^{(r)} = f_1(Z_{1nt}^{(r)} | d_{nt+r} = 1),
\]
as the probability density function for \( Z_{1nt}^{(r)} \), conditional on participating at date \( t + r \). Likewise, let \( f_{nt}^{(r)} = f(S_{1nt}^{(r)}) \) be the related probability density that does not condition on participating in period \( t + r \) for \( r = 1, ..., \rho \), denote their derivatives with respect to \( h_{nt} \) by \( f_{nt}^{(r)} \) respectively. It is straightforward to show that
\[
\frac{\partial p_{1nt}^{(r)} \partial h_{nt} = \left[ f_{1nt}^{(r)} - f_{nt}^{(r)} \right] / f_{nt}^{(r)}}{p_{1nt}^{(r)} \quad r = 1, ..., \rho.} \tag{38}
\]
We could derive this expression using the fact that \( p_{1nt}^{(r)} \) can be written as \( p_{1nt}^{(r)} = \Pr(d_{nt+r} = 1 | Z_{1nt}^{(r)}) = \Pr(d_{nt+r} = 1)f_{1nt}^{(r)} / f_{nt}^{(r)} \). Differentiating this expression with respect to \( h_{nt} \) yields the above expression. The nonparametric estimates \( f_{1nt}^{(r)} \) and \( f_{nt}^{(r)} \) are defined, respectively, as the numerators and denominators of \( p_{knt}^{(r,N)} \) in equation (36). The estimates of \( f_{1nt}^{(r)} \) and \( f_{nt}^{(r)} \) are obtained from the derivatives of the estimates \( f_{1nt}^{(r,N)} \) and \( f_{nt}^{(r,N)} \) with respect to \( h_{nt} \). (See Pagan and Ullah(1999)).

### 3.4 Third Step: Estimation of the Structural Parameter

Assume that \( (\varepsilon_{0nt}, \varepsilon_{1nt}) \) is distributed as a Type I extreme value with variance parameter \( \sigma^2 \) and mean zero. This distributional assumption for the preference shocks implies that \( Q(p) = \sigma \ln[p/(1-p)] \), \( \varphi_0(p) = \xi - \sigma \ln[(1-p)] \) and \( \varphi_1(p) = \xi - \sigma \ln[p] \), where \( \xi \) is the Euler constant. Combining equation (23) and (28) along with the above expressions of \( Q(p), \varphi_0(p) \)
and \( \varphi_1(p) \) we obtain that:

\[
\sigma \ln[p_{nt}/(1 - p_{nt})] = B_{0t} + z'_{nt}B_1 - z'_{nt}h_{nt}B_2 - \theta_0(1 - l^2_{nt})
- \sum_{s=1}^{\rho} \theta_s h_{nt}(l_{nt-s} + \beta^s) + \sigma E_t[\sum_{s=1}^{\rho} \beta^s \ln \left( \frac{1 - p_{snt}}{1 - p_{0nt}} \right)]
+ \eta_t \lambda_t \sum_{m=1}^{M} d_{mnt}[y_m(h_{nt}, H_{nt-1}, z^p_{nt}, \theta_e) - \gamma_m + \beta \gamma_m \bar{p}_{mn,t+1}] - \gamma_{mnt} \sum_{m=1}^{M} d_{mnt}[y_m(h_{nt}, H_{nt-1}, z^p_{nt}, \theta_e) - \gamma_m + \beta \gamma_m \bar{p}_{mn,t+1}] \quad (39)
\]

Finally, combining the Euler equation for hours (18) and the alternative representation of value (28) along with the above expressions of \( Q(p) \), \( \varphi_0(p) \) and \( \varphi_1(p) \) we obtain that:

\[
E_t \{ d_{nt}[\sigma \sum_{s=1}^{\rho} \beta^s (1 - p_{snt})^{-1} \nabla h_{nt} p_{1nt}^{(s)} - z'_{nt}B_2 - 2\theta_0 l_{nt} - \sum_{s=1}^{\rho} \theta_s l_{nt-s} + \beta^s] + \eta_t \lambda_t \sum_{m=1}^{M} d_{mnt}[\pi_{m1} + 2\pi_{m2} l_{nt} + \beta \gamma_m \nabla h_{nt} \bar{p}_{mn,t+1}] \} = 0 \quad (40)
\]

where \( \nabla h_{nt} \bar{p}_{mn,t+1} \) is derivative of the beliefs with respect to current hours.

Note that from the first step estimation we have estimates of \( \pi_{m1}, \pi_{m2}, \beta \gamma_m \) and all the other parameters of the production function. In addition, from the first step estimation we have estimate of \( \phi_{nt} \), which is

\[
\phi_{nt} = (1 - \alpha)^{-1} \ln(\eta_t \lambda_t).
\]

The second step estimation yields estimates of \( p_{nt}, p_{1nt}^{(s)}, \bar{p}_{mn,t+1}, \nabla h_{nt} p_{1nt}^{(s)} \) and \( \nabla h_{nt} \bar{p}_{mn,t+1} \). We can estimate the moment conditions

\[
m_{1nt}(\Theta u, \Theta_e^{(N)}, \Theta_e^{(N)}, \psi^{(N)}) = \sigma \ln[p_{nt}^{(N)}/(1 - p_{nt}^{(N)})] - B_{0t} - z'_{nt}B_1 + z'_{nt}h_{nt}B_2
+ \theta_0(1 - l^2_{nt}) + \sum_{s=1}^{\rho} \theta_s h_{nt}(l_{nt-s} + \beta^s)
- \sigma \sum_{s=1}^{\rho} \beta^s \ln \left( \frac{1 - p_{snt}^{(N)}}{1 - p_{0nt}^{(N)}} \right)
- \exp((1 - \alpha)\phi_{nt}^{(N)}) \sum_{m=1}^{M} d_{mnt}[y_m(h_{nt}, H_{nt-1}, z^p_{nt}, \theta_e^{(N)}) - \gamma_m^{(N)} + \beta \gamma_m^{(N)} \bar{p}_{mn,t+1} + 1]
\]

(41)
and
\[
   m_{2nt}(\Theta u, \Theta_c^{(N)}, \Theta_e^{(N)}, \psi^{(N)}) = d_{nt} [\sigma \sum_{s=1}^{\rho} \beta^s (1 - P_{1nt}^{(N)})^{-1} \nabla h_{nt} P_{1nt}^{(N)} - 2 z_{nt}' B_2 - 2 \theta_0 l_{nt} - \sum_{s=1}^{\rho} \theta_s (l_{nt-s} + \beta^s) + \exp((1 - \alpha) \phi_{nt}^{(N)}) \sum_{m=1}^{M} d_{mnt} (\pi_{m1}^{(N)} + 2 \pi_{m2}^{(N)} h_{nt}) + \beta \gamma_{m}^{(N)} \nabla h_{nt} \tilde{P}_{mn,t+1}^{(N)}] \]

where \(\psi^{(N)} = (\tilde{P}_{nt}^{(N)}, \tilde{P}_{0nt}^{(N)}, \tilde{P}_{1nt}^{(N)}, \tilde{P}_{mn,t+1}^{(N)})\), is the nonparametric second step estimates and \(\Theta u = (\sigma, \alpha, \beta, B_{01}, ..., B_{0T}, B_1, B_2, \theta_0, ..., \theta_{\rho})\) are the structural parameters left to be estimated.

There are now two sources of errors in evaluating the sample counterpart of (39) and 40. The first is the forecast errors from replacing the expectations of future variables with their realizations. The second error is the approximation error that arises due to replacing the true values of condition choice probabilities, condition expectation and the time-invariant individual specific effects with their estimates. Let us define the 2×1 vector \(m_{3nt}(\Theta u, \Theta_c^{(N)}, \Theta_e^{(N)}, \psi^{(N)}) \equiv m_{1nt}(\Theta u, \Theta_c^{(N)}, \Theta_e^{(N)}, \psi^{(N)}), m_{2nt}(\Theta u, \Theta_c^{(N)}, \Theta_e^{(N)}, \psi^{(N)})\)' and let \(T_3\) denote the set of periods for which the hours and participation equations are valid. Define the vector \(m_{3n}^{(N)}(\Theta u, \Theta_c^{(N)}, \Theta_e^{(N)}, \psi^{(N)}) \equiv (m_{3n1}(\Theta u, \Theta_c^{(N)}, \Theta_e^{(N)}, \psi^{(N)}), ..., m_{3nT_3}(\Theta u, \Theta_c^{(N)}, \Theta_e^{(N)}, \psi^{(N)}))'\) as the vector of the idiosyncratic errors for given individual over time. Define \(\Omega_{nt}^{(N)} = E_t[m_{3nt}(\Theta u, \Theta_c^{(N)}, \Theta_e^{(N)}, \psi^{(N)})m_{3nt}(\Theta u, \Theta_c^{(N)}, \Theta_e^{(N)}, \psi^{(N)})']\), the off diagonal elements of \(\Omega_{nt}^{(N)}\) are zero because
\[
   E_t[m_{3nt}(\Theta u, \Theta_c^{(N)}, \Theta_e^{(N)}, \psi^{(N)})m_{3nr}(\Theta u, \Theta_c^{(N)}, \Theta_e^{(N)}, \psi^{(N)}')] = 0 \text{ for } r \neq t, r < t. \]

The 2×2 conditional heteroscedasticity matrix \(\Omega_{nt}^{(N)}\) associated with the individual-specific errors \(m_{3nt}(\Theta u, \Theta_c^{(N)}, \Theta_e^{(N)}, \psi^{(N)})\) is evaluated using a nonparametric estimator based on the estimated moments \(m_{3nt}(\Theta_{1u}^{(N)}, \Theta_c^{(N)}, \Theta_e^{(N)}, \psi^{(N)})\) derived from an initial consistent estimate of \(\Theta_{1u}^{(N)}\). The optimal instrumental variables estimator for \(\Theta u^{(N)}\) satisfies
\[
   \Theta u^{(N)} = \arg \min_{\Theta u} \left[ 1/N \sum_{n=1}^{N} \{ m_{3n}^{(N)}(\Theta u, \Theta_c^{(N)}, \Theta_e^{(N)}, \psi^{(N)}) (\Omega_{nt}^{(N)})^{-1} m_{3n}^{(N)}(\Theta u, \Theta_c^{(N)}, \Theta_e^{(N)}, \psi^{(N)}) \} \right]^{-1} \Omega_{nt}^{(N)} \}
\]

3.5 Semiparametric Asymptotic Variance

It is well known in the econometric literature that under certain regularity condition pre-estimation does not have any impact on the consistency of the parameters in subsequent steps of multi stage estimation (see. Newey 1984, Newey and McFadden(1994) and Newey(1994)). The Asymptotic Variance, however, is affected by the pre-estimation. In order to conduct inference in this type of estimation, one has to correct for the asymptotic variance for the pre-estimation. The method used for correcting the variance in the final step of estimation depends
on whether the pre-estimation parameters are of finite or infinite dimension. Unfortunately, our estimation strategy combines both finite and infinite dimensional parameters. Combining results from two sources (Newey, 1984 and Newey and McFadden, 1994), however, allows us to derive the corrected asymptotic variance for our estimator.

Following Newey (1984), we can write the sequential moments conditions for the first step and the third step estimation as a set of joint moment conditions:

\[
m_n(\Theta_u, \Theta_c, \Theta_e, \psi) = \begin{bmatrix}
(Y_n - Z_n \Theta_c) Z_n \\
(Y_{1n} - X_{1n} \Theta_{e1}) Z_n \\
\vdots \\
(Y_{Mn} - X_{Mn} \Theta_{eM}) Z_n \\
m_{3n}(\Theta_u, \Theta_c, \Theta_e, \psi)
\end{bmatrix}
\]

where \((Y_n - Z_n \Theta_c) Z_n\) is the orthogonality condition from the estimation of the consumption equation, \((Y_{mn} - X_{mn} \Theta_{em}) Z_n\) is the orthogonality condition from the estimation of the earnings equation and \(m_n(\Theta_u, \Theta_c, \Theta_e, \psi)\) is the moment conditions from the third step estimation. Let \(\Theta = (\Theta_u, \Theta_c, \Theta_e)'\), with the true value denoted by \(\Theta_0\). For expositional simplicity consider first a standard Taylor series expansion for \(\Theta^{(N)}\) around \(\Theta_0\):

\[
\sqrt{N}(\Theta^{(N)} - \Theta_0) = -\sum \nabla_{\Theta} m_n(\Theta) \Omega_n^{-1} \nabla_{\Theta} m_n(\Theta) \Psi^{(N)} - \frac{1}{\sqrt{N}} \sum \nabla_{\Theta} m_n(\Theta) \Omega_n^{-1} m_n(\Theta)
\]

(43)

where \(m_n(\psi^{(N)}) \equiv m_n(\omega, \Theta_0, \psi^{(N)})\), \(\nabla_{\Theta} m_n(\psi^{(N)}) \equiv m_n(\omega, \Theta_0, \psi^{(N)})\) and \(\Theta\) is the intermediate value between \(\Theta^{(N)}\) and \(\Theta_0\). The idea is to show the convergence in distribution for the term \(N^{-\frac{1}{2}} \sum \nabla_{\Theta} m_n(\psi^{(N)}) \Omega_n^{-1} m_n(\psi^{(N)})\). As described in details in the Appendix, we first obtain an appropriate linearization for the moment condition \(m_n(\omega, \Theta_0, \psi)\) in \(\psi\). In particular, there exist a functional \(\Gamma(\omega, \psi)\) that is linear in \(\psi\) such that for \(\psi\) close enough to \(\psi_0\)

\[
\|m_n(\omega, \Theta_0, \psi) - m_n(\omega, \Theta_0, \psi_0) - \Gamma(\omega, \psi - \psi_0)\| \leq \Psi(\omega) \|\psi - \psi_0\|^2
\]

The conditions below ensure that \(\psi^{(N)}\) is close enough to \(\psi_0\) for \(N\) large enough, in particular that \(\sqrt{N} \|\psi^{(N)} - \psi_0\|^2\) converges to zero

**A3:** There is a version of \(\psi_0(z)\) that is continuously differentiable of order \(\tau\), greater than the dimension of \(z\) and \(\psi_1(z) = f_z(z)\) is bounded away from \(0\).

**A4:** \(\int K(u) du = 1, \text{ and for all } j < \tau \int K(u)(\bigotimes_{s=1}^{j} u) du = 0\)

**A5:** The bandwidth \(\delta_N\) satisfies \(N^{-\frac{2}{2}} (\log(N))^2 \rightarrow \infty\) and \(N \delta_N^{2\tau} \rightarrow 0\)

**A6:** There exists a \(Y(\omega), \epsilon > 0\) such that

\[
\|\nabla_{\Theta} m_n(\omega, \Theta, \psi) - \nabla_{\Theta} m_n(\omega, \Theta_0, \psi_0)\| \leq \Psi(\omega) [\|\Theta - \Theta_0\|^\epsilon + \|\psi - \psi_0\|^\epsilon]
\]
and $E[\Upsilon(\omega)] < \infty$. and

**A7:** $\Theta^{(N)} \rightarrow \Theta_0$ with $\Theta_0$ in the interior of its parameter space.

**Theorem 1:** Under A1-A7 and $\Upsilon(\omega)$ is in the Appendix. Then

$$\sqrt{N}(\Theta^{(N)} - \Theta_0) \Rightarrow N(0, V(\Theta_0))$$

where

$$V(\Theta_0) = \mathbb{E}[\nabla_\Theta m_n(\omega) \Omega_n^{-1} \nabla_\Theta m_n(\omega)]^{-1}$$

$$\times \mathbb{E}[\nabla_\Theta m_n(\omega) \Omega_n^{-1} \{m_n(\omega) + \Upsilon(\omega)\} \{m_n(\omega) + \Upsilon(\omega)\}' \Omega_n^{-1} \nabla_\Theta m_n(\omega)']^{-1}$$

Assumption A3-A7 is standard in the semiparametric literature, see Newey and McFadden (1994) for details. One can now use theorem 1 to calculate the standard for all the parameters in our estimation.

3.6 Data

The data for this study are taken from the Family-Individual File, Childbirth and Adoption History File and the Marriage History File of the Michigan Panel Study of Income Dynamics (PSID). The variables used in the empirical study are $h_{nt}$, the annual fraction of hours work by individual $n$ at date $t$; $w_{nt}$, the reported real average hourly earnings at $t$; $c_{nt}$, real household food consumption expenditures; $FAM_{nt}$, the number of household members; $YKID_{nt}$, the number of children less than six years of age; $OKID_{nt}$, the number of children of ages between six and fourteen; $AGE_{nt}$, the age of the individual at date $t$; $EDU_{nt}$, the years of completed education of the individual at time $t$; $NE_{nt}$, $W_{nt}$, $SO_{nt}$, which are region dummies for northeast, west, and south, respectively, and $MS_{nt}$, denoting whether a woman is married or not. We also use variables of the variables that are related to the spouse of the individual if the individual is married. $SP:EDUC_{nt}$ and $SP:INCOME_{nt}$, the completed years of education and the labor market income of the of individual $n$ spouse in period $t$ respectively. Individual are classified into two different occupation categories, Professional and Nonprofessionals. We only keep white individuals between the ages of 18 and 65 in our sample. After eliminating missing variables we are left with 6453 individuals over the years 1968 to 1993 of which 46% are women.

4 Empirical Results

The purpose of estimating the consumption equation is obtaining estimates of the marginal utility of wealth for our main estimation equations, and therefore, we do not focus the discussion on these results. Table 1, contains the results from this estimation. The results very standard and consistent with estimates of these parameters in previous literature (see Gayle and Miller (2004) and Altug and Miller (1998) for similar estimates). Furthermore, we obtain
"reasonable" estimates for our risk aversion parameter which is normally a problem in estimation of consumption equations (See Altug and Miller (1990) and Gayle and Miller (2004) for a discuss of these problems).

As can be seen from Figure 1, there has been a significant increase in aggregate productivity in both occupations. This increase, however, was much larger in the professional occupations than in the nonprofessional occupations. Our theoretical model implies that productivity shocks should have a more significant effect on female labor force participation than on men's participation, and therefore lead to reduction in the gender earnings gap (see Corollary 7 (i)). These shocks changes should also lead to increase sorting of women into the professional occupations, as is documented in the literature (see Lewis (1996) for more details). We will later compute directly the effect of productivity shocks on the changes in the gender earnings gap (see Table 8).

The estimation results of the earnings equation is reported in Table 2. Consistent with corollary 5, the professional occupations have significantly higher returns to labor market experience than the nonprofessional occupations. These results are consistent with the empirical fact that more women sort into the nonprofessional occupations than men. Another, evidence in support of corollary 5 is that there is a larger cost of hiring a new worker, along with the higher returns to working less hours (part time) in the nonprofessional occupation. This can be seen by comparing the linear and quadratic terms in current hours.

Table 3 contains the estimates of the fixed cost of labor force participation. Here we found our first evidence in support of Proposition 4. The difference in the estimated coefficients for number of kids (both young and old kids) for men and women, is suggestive evidence that more women "specialize" in non-market work relative to men. This is consistent with a theory of Becker (1965) of home production division of labor. It should be noted that this evidence is only suggestive because we do not observed time spent in home production directly in our data; our results, using the number of kids as a proxy for the amount of home production hours, however, supports this theory. There is no significant difference in the cost of participation of men and women with the same years of completed education. A larger number years of completed education raises the likelihood of working for men and women equally. The effect of marital status is highly nonlinear and depends on the education level of one's spouse. A married individual is more likely to participation in the labor force. A married women who is married to a more educated man, however, is less likely to participation, whereas a man who is married to a more educated woman is more likely to participate in the labor force.

Table 4 contains the results of our estimates on the utility of leisure (or non-market production function). Again the estimates support the idea of specialization, however, the results are a little more counter intuitive. For example, whereas women with kids are less likely to work in the labor market, conditional on working they are more likely to work more hours; the opposite is true for men. The opposite is true for education. That is, educated women are more likely to work in the labor market, but education has no significant effect on the likelihood of working more hours conditional on working in the labor market. Education, on

\footnote{In our model unemployment is interpreted as a decision not to work. This is in keeping with the labor literature on female labor supply.}
the other hand, does not have any significant effect on the likelihood of participation of mean man, but it does increase the likelihood of working more hours conditional on working in the labor market. Lastly, conditional on working in the labor market, marital status decreases the likelihood of working more hours. Whereas conditional on working, being married to a more educated male increases the likelihood of working more hours for women but not for men. These results are supportive as well of the idea of cross-group complementarities in the utility function (asymmetric equilibrium).

Table 5 contains the estimates for the time non-separability in non-market hours. They show that there are significant complementarities between non-market hours across time for women. The results on complementarities between non-market hours across time for men are mixed. In particular, non-market hours for men are compliments one period back; however, they become substitutes two periods back. This further suggests that our model does not require a coordination failure in order to exhibit a "discriminatory equilibria". This could be the results of just an asymmetric equilibria that would generate self-fulfilling beliefs and different labor market history between men and women as noted in Proposition 4.

The fact that our structural estimates are consistent with the prediction of our model does not mean that private information is quantitatively important or that the gender earning gap is driven by discriminatory equilibria. This is even more a problematic given that the estimated switching cost is not very high ($3,000 in professional and less $1,000 in nonprofessionals). Whereas these numbers may be reasonable they are still small relative to the earnings gap.

To investigate this, we decompose the earnings gap into four components: human capital, i.e. current and past hours worked in the market, firms’ beliefs; the fixed effects, and other, i.e. education and age composition. The results are reported in Table 6. Table 6 has the median earnings of a woman over the median earnings of a man. The earnings gap for our sample is 87% and 76% for professionals and nonprofessionals, respectively. Our model predicts an earnings gap of 92 and 81 percent for professional and nonprofessional respectively. Of the 92 percent predicted earnings gap in the professional occupations, 60 percent is due to difference in human capital, 12 percent is due to differences in firms’ beliefs, 4 percent is due to differences in the estimated fixed effects, and 10 percent is due to differences in education and age composition between men and women. In the nonprofessional occupation, of the 81 percent earnings gap predicted by our model, 56 percent is due to differences in human capital, 9 percent is due to differences in employers’ beliefs, 7 percent is due to differences in the estimated fixed effect and 4 percent is due to differences in the education and age composition between men and women. Given how the fixed effect is estimated, one may be worried that it is capturing implicit discrimination which is not in our model; given that it accounts, however, for only a small fraction of the predicted earnings gap, we can safely ignore these other sources of possible discrimination.

Given that our model performs reasonable well in explaining the earnings gap we are now in a position to look at the sources of the change in the earnings gap over two disjoint time periods; 1974-1978 and 1984-1988. To do this we calculate the median earnings gap in both time periods and expressed the difference as a percentage of the median earnings gap in the first period. The results are reported in Table 7. The raw changes in the earnings gap over the period are 30 and
25 percent for professionals and nonprofessionals, respectively. Our model predicts a change of 28 and 22 percent for the two occupations, respectively. Therefore, our model is doing slightly better at predicting the change in the earnings gap than the levels. We then decomposed the predicted changes into changes due to differences in human capital, firms’ beliefs and education and age composition (labelled "other" in table 7). Changes in the differences in human capital over the two periods account for 67 and 65 percent of the changes in our two occupations, respectively. Changes in firms’ beliefs account for 8 and 6 percent of the changes in professional and nonprofessional respectively. Whereas education and age composition accounted for 25 and 29 percent of the changes in professionals and nonprofessionals, respectively.

Human capital is by far the most important factor in explaining both the earnings gap and the changes in the earnings gap over time. Human capital accumulation, however, is endogenous to our model. Hence, the effect of private information is compounded into the human capital effect. In order to disentangle these different effects on human capital accumulation we need to solve the model. The problem with solving the model is that with private information there is the possibility of multiple equilibria. Therefore, there are no guarantees that the equilibrium we solve for is the one actually played in the data, since the same parameter values of our model may be consistent with many different equilibria. In order to get around this problem we solve the model under two different situations. In both situations we know that there is a unique equilibrium. We then compare the results from our solution with the actual data to obtain measure of the effect of private information on the effect of human accumulation on the earnings gap. The first situation is when the switching cost, \( \gamma_m \), is equal to zero. The second is when firms observed all the private information of the individual. Under both scenarios we can solve the model using backward induction.

We then calculate five sources of the changes in human capital accumulation on the earnings gap. Below is explained in details how we calculate each one. These results are reported in Table 8,

1. The scenario under which the switching costs are zero is calculated as follows. We take the two disjoint time periods and calculate the average of all the inputs to our model for each period. These inputs includes, the demographic characteristics, aggregate shocks, marginal utility of wealth, fixed effects and the estimated transition probabilities of marital status and number of kids. We then solve the model in both time periods, setting \( \gamma_m \), equal to zero. Then we calculate the implied changes in the earnings gap. We decompose these changes as in table 7. The we expressed the amount of the change due to human capital as a percentage of the change due to human capital in table 7. This accounts for 38 and 32 percent of this change in the professional and nonprofessional occupations, respectively.

2. The symmetric information case is calculated as described in 1; however, instead of setting \( \gamma_m \) equal to zero, we solve the model backward calculating the actual probability of working next period in the the same occupation conditional on working today in that occupation. This calculation is done conditioning on all the information (private or otherwise) known by the worker today. This accounted for 12 and 13 percent of the
changes in the profession and nonprofessional occupation, respectively.

3. We again set $\gamma_m$ equal to zero, holding all the demographic characteristic-marital status, number of kids, years of completed education, spouse education level- and there transition probabilities at the 1974-1978 levels in both period and repeat the calculations in 1. This accounts for 28 and 34 percent of the changes in the two occupations, respectively, over and above those found in 1.

4. We again repeat the exercise in 3, but this time we hold aggregate shock to home production fixed at their 1974-1978 levels in both time periods, while allowing demographic characteristics to vary across the two periods. This only accounted for 2 and 3 percent of the changes over and above the 38 and 34 percent of the changes accounted for by the measure in 1.

5. We again repeat the same exercise as in 3, but this time holding aggregate shock to market productivity constant across the two time period, while allowing all other input to vary across the two time period. This accounted for 18 and 11 percent of the changes over and above the 38 and 34 percent of changes accounted for in 1.

From this exercise we learn that private information, switching cost demographic changes and changes in aggregate productivity in the market sector account for most of the changes in human capital. In term, human capital accounts for most of the changes in the explained gender earnings gap over the period..

5 Conclusions

We investigate the role of labor market attachment, on-the-job human capital accumulation and statistical discrimination in the narrowing gender wage gap. In particular, we explore the effect of the overall increase in productivity, improvement in home production, and changes in family structure on the decline in the wage gap. Many of the observe changes, such as labor market experience and occupation choice are not exogenous. Therefore, based on empirical regularities we find, we formulation of the model of labor market attachment and experience.

Our model is a dynamic adverse selection model of earnings determination and statistical discrimination based on expectations of future labor market participation. In our model, workers have heterogeneous preferences with respect to the degree of the labor market attachment. Preferences are known privately to workers. It is costly for firms to hire new workers. The rate of on-the-job human capital accumulation, and costs of hiring new workers vary across occupations. Thus, labor market experience, and attachment to the labor force importance varies across occupation. Weak labor market attachment may be more costly at different stages of a worker’s career, and may vary with workers’ skills and experience. Based on observable characteristic, employers form beliefs regarding the degree of labor market attachment when they offer wage contracts. Wage contracts consist of hours and earning. Workers maximize life time utility. Utility is an increasing function of consumption, leisure and/or home production.
The estimation results do not support the hypothesis that changes in home production technology explain the increase in participation of women in the labor market. Such changes should have caused a decrease in the fixed cost of participating in the labor market (estimated as part of the utility function specification). Our estimation results suggest that fixed costs of participation have increased slightly (or have not changed) in periods when women increase labor market participation. Increase in overall productivity cause an increase in overall wages in the economy (could be caused by aggregate technology shocks) this increase causes changes in beliefs in participation of women which cause an increase in wages of women relative to men.

Further extensions of our framework will include exploring the role of changes in family structure on the gender wage gap. We find that a significant factor that drives the change in beliefs is the changes in family structure. In particular, there has been a significant increase in participation of married women. Divorce rate has gone up, causing the probability of married women to be single in the future to increase. Single women and single mothers are more attached to the labor force. Thus, beliefs about future participation of women has gone up. Furthermore, fertility of married women declined; this caused an increase beliefs of future labor market participation. In our model these changes in family structure are assumed to be exogenous, and therefore, are identified as factors causing changes in beliefs, increase in participation rate and decline in the gender wage gap. Whereas, our empirical findings suggest that changes in family structure may be important to further understanding of the changes in the gender wage gap, these changes are endogenous to changes in earnings. Therefore, we cannot disentangle the causality relations. Inferring causality is beyond the scope of this paper.

6 Appendix

6.0.1 Appendix A: Theoretical Results

Proof of Optimal Contract. This proofs establishes that given the strategies and beliefs of the players the optimal contract is of the form specified in the proposition. The free entry condition implies that in equilibrium, the expected value of a vacancy is zero. We use this to derive the optimal contract.

At time $t = T$ (worker’s final year) free entry condition that implies that for a new employer, expected profit from offering a worker a contract is zero. The expected profit for from offering a contract $s(h_t, H_{t-1}, D_{t-1}, z^*_t)$ is

$$\pi_T = f_m(h_T, H_{T-1}, z^*_t) - S_T(h_T, H_{T-1}, z^*_{mT}) - \gamma_m = 0 \quad (44a)$$

As long as $f_m(h_T, H_{T-1}, z^*_t) - \gamma_m < 0$, it is possible to raid the worker (who accepts the highest offer). Therefore,

$$S_T(h_T, H_{T-1}, z^*_{mT}) = f_m(h_T, H_{T-1}, z^*_t) - \gamma_m$$

Denote the (current) employers’ profit $(d_{m,T-1} = 1)$ at time $T$ by $\pi^e_T$. Recall that if the worker participates there is no additional costs and if she does not then the value of a vacancy is zero.
Consider potential employer making an offer at time $t = T - 1$, outside employers anticipate that in the proceeding period they will earn a profit of $\gamma_m$ if the worker participate, therefore, they make an offer such that the expected profits over the two periods are zero. Consider a free entry condition at $T - 1$:

$$\pi_{T-1} = f_m(h_{T-1}, H_{T-2}, z_{nT-1}^*) - \gamma_m - S(h_{T-1}, H_{T-2}, z_{T-1}^*)$$

$$+ \beta \pi_T T^m (z_{T-1}^*, H_{T-2}, h_{T-1}) = 0$$

substituting $\pi_T$, and given the worker’s strategy to accept the highest offer (remain with current employer if tie)

$$\pi_{T-1} = f_m(h_{T-1}, H_{T-2}, z_{nT-1}^p) - \gamma_m - S(h_{T-1}, H_{T-2}, z_{T-1}^*)$$

$$+ \beta \gamma_m T^m (z_{T-1}^*, H_{T-2}, h_{T-1}) = 0$$

Thus,

$$S(h_{T-1}, H_{T-2}, z_{T-1}^*) = f_m(h_{T-1}, H_{T-2}, z_{nT-1}^p) - \gamma_m (1 - \beta T^m (z_{T-1}^*, H_{T-2}, h_{T-1}))$$

Notice that if a workers is hired at period $T - 2$, the expected profits if the worker works in period $T - 1$ is

$$\gamma_m (1 - \beta T^m (z_{T-1}^*, H_{T-2}, h_{T-1}))$$

In period $T$ if the worker is employed the profit is $\gamma_m$, thus, the expected profits over both periods is $\gamma_m (1 - \beta T^m (z_{T-1}^*, H_{T-2}, h_{T-1}')) + \beta T^m (z_{T-1}^*, H_{T-2}, h_{T-1}) \gamma_m = \gamma_m$. Solving backwards to period : one can show that at any period $s < T$ that the free entry condition implies

$$[f_m(h_s, H_{s-1}, z_{ns}^p) - \gamma_m - S(h_s, H_{s-1}, z_{s-1}^*)]$$

$$+ T_{mns+1} \beta \gamma_m = 0$$

$$S(h_s, H_{s-1}, z_s) = f_m(h_s, H_{s-1}, z_{ns}^p) - \gamma_m$$

$$+ \beta \gamma_m T_{mns+1} (z_{s}^*, H_{s-1}, h_{s-1})$$

Given the beliefs and worker’s strategy to accept the highest offer, and other firm’s strategies, the contract above is optimal.

Next, we show that no possible deviation from the competitive rate can be profitable to any player. Each occupation offers all hours in the occupation’s range. No worker accepts lower wage ($d = 0$ if $S'(h_s, H_{s-1}, z_s) < S(h_s, H_{s-1}, z_s)$). By construction, offering higher wage, for any given hours, and observables, and the inability to commit of both firms and workers, the expected profit is negative. Notice that every contract for every hour in any occupation is offered. They all yield zero profit and firms are indifferent. Any hours contract that is not offered provides an opportunity for a new firm to enter and attract a worker at a wage which yields positive profits. The worker’s choice of participation and optimal hours is described in (20).and(7) The proof that the beliefs above satisfy Bayes’ law is in (???)

36
Next, we argue that the worker’s strategy, given any hours she chooses to work is to accept the highest salary. Higher salary increases utility today and does not affect beliefs or salaries tomorrow (recall salary is not part of the observed employment history). Thus, for any hours worked, if offered a higher salary by a firm in the occupation, the worker will take the highest offer (switching employers within occupation does not affect beliefs). Off-equilibrium-path, workers who work less than the minimal hours receive lower salaries and there are no gains from deviation (beliefs will not increase), there is no deviation which is profitable for working more the largest amount of hours, as beliefs are not adjusting to be higher than the beliefs if worked the highest hours optimal on the equilibrium-path Q.E.D. ■

**Proof of Existence.** 1. Follows immediately from continuity and strict concavity of the utility function that there is a solution to the worker’s problem, for any contracts offers.

2. First notice that since choice of hours worked and participation depend on current salary, they are affected by the firm’s beliefs. Consider the final period $T$, the salaries are independent of the firm’s beliefs as all factors affecting production are observable the following functions summarize the worker’s strategy (participation decision and hours worked respectively):

$$Q_T(z_{nT}, H_{nT-1} (S_{nT-1}(h_{nT-1}, \bar{p}_{mn,T}), ..., S_{n1}(h_{n1}, \bar{p}_{mn,2})), h_T^*)$$

$$I_T(z_{nT}, H_{nT-1} (S_{nT-1}(h_{nT-1}, \bar{p}_{mn,T}), ..., S_{n1}(h_{n1}, \bar{p}_{mn,2})), h_T^*)$$

In period $T-1$

$$\bar{p}_{mn,T} = \int QT(z_{nT}, H_{nT-1}(\bar{p}_{mn,T}), h_T) I_T(z_{nT}, H_{nT-1}(\bar{p}_{mn,T}), h_T) \mu_T(z_{nT} | H_{nT-1}, D_{nT-1}, z_{nT}^*) dz_{nT}$$

Next, in order to solve for $\bar{p}_{mn,t}$ notice that it appears in $Q_T$ and $I_T$ on the right hand side, for any period $t$. In particular, $\bar{p}_{mn,t+1}$ is the solution to

$$\bar{p}_{mn,t+1} = \int \{ \int f(z_{nt+1} | z_{nt}) \{ Q_{t+1}(z_{nt+1}, H_{nt}(\bar{p}_{mn,t+1}), h_{nt+1}(\bar{p}_{mn,t+1}))

I_{t+1}(z_{nt+1}, H_{nt}(\bar{p}_{mn,t+1}, \bar{p}_{mn})), h_{nt+1}^*(\bar{p}_{mn,t+1}), h_{nt+1}^*(\bar{p}_{mn,t+1}), H_{nt}(\bar{p}_{mn,t+1}, \bar{p}_{mn}))) dz_{nt+1} \}$$

$$\mu_t(z_{nt} | H_{nt-1}, D_{nt-1}, z_{nt}^*) dz_{nt}$$

A) Note that $\bar{p}_{mn,t+1} : [0, 1]$ , and that the left hand side is also defined on the interval $[0, 1]$ . Thus, continuity of the RHS suffices to guarantee a solution to each one of the equations separately. We then show that there exist a solution to the system of beliefs $\bar{p}_{mn,t+1}, ...\bar{p}_{mn,2}$ simultaneously.

B) To show continuity: let $z_{nt+1}$ be the "marginal type" for which $h_{nt+1}^*(H_{nt}, z_{nt}^*) \equiv h_{nt+1}(H_{nt}, z_{nt+1}^*)$, and $\bar{z}_{nt+1}$ the type for which $h_{nt+1}(H_{nt}, z_{nt+1}, \bar{z}_{nt+1}) \equiv h_{nt+1}(H_{nt}, z_{nt+1})$. $h_{nt+1}(z_{nt+1}, ...)$ is continuous and invertible in $z_{nt+1}$ as the utility function is continuous. Where $z_{nt+1}$ and $z_{nt+1}$ are the lowest (highest), $z$ so that a worker with a history $H_{nt}$, and characteristics $z_{nt+1}$ chooses optimally to work $h_{nt+1}$. Thus we can write: $h_{nt+1}^*(H_{nt}, z_{nt+1}) = z_{nt+1}$ and $z_{nt+1}$

37
\( \equiv \overline{h}_{mt+1}(H_{nt}; z_n^*) \). Since \( I_{t+1}(z_{nt+1}; H_{nt}(\overline{p}_{mn,t+1}, \ldots, \overline{p}_{mn2}), h_{nt+1}^*, S_{nt+1}(h_{nt+1}^*)) \) is an indicator function which takes the value 1 when \( h_{nt+1}^* \in [\overline{h}_{nt+1}(H_{nt}; z_n^*), \overline{\overline{h}}_{nt+1}(H_{nt}; z_n^*)] \) and 0 otherwise, we can rewrite the integral

\[
\int_{z_{nt}} f(z_{nt+1} \mid z_{nt}) \{Q_{t+1}(\cdot)I_{t+1}(\cdot)dz_{nt+1} = \int_{z_{nt}} f(z_{nt+1} \mid z_{nt}) \{Q_{t+1}(h_t(\overline{p}_{mn,t+1})).dz_{nt+1}
\]

C) Since \( h_t(\overline{p}_{mn,t+1}) \) is continuous in \( p_{mn,t+1} \) and \( Q_{t+1}(h_t(\overline{p}_{mn,t+1})). \) is continuous in \( p_{mn,t+1} \). Let \( m - 1, m + 1 \) denote the occupations for which the worker works if she chooses hours below (above) \( \overline{h}_{nt+1}(H_{nt}; z_n^*), \overline{h}_{nt+1}(H_{nt}; z_n^*) \). Since these hours are determined by the following conditions:

\[
f_m(h_{nt+1}^*(H_{nt}; z_n^*)) - f_m-1(h_{nt+1}^*(H_{nt}; z_n^*)) - \gamma_m + \gamma_{m-1} = 0
\]
\[
f_m(\overline{h}_{nt+1}(H_{nt}; z_n^*)) - f_m+1(\overline{h}_{nt+1}(H_{nt}; z_n^*)) - \gamma_m + \gamma_{m+1} = 0
\]

From the continuity of production function in each occupation in all factors of production, \( h_{nt+1}^*(H_{nt}; z_n^*), \overline{h}_{nt+1}(H_{nt}; z_n^*) \) are continuous in \( h_{nt} \). Therefore, there exists a solution to every period beliefs separately. Next we show that there exists a beliefs sequence which solves the following simultaneous equation systems (where \( g \) represent the rhs of equations):

\[
\overline{p}_{mnT} = g(\overline{p}_{mnT}, \ldots, \overline{p}_{mn2})
\]
\[
\overline{p}_{mnT-1} = g(\overline{p}_{mnT-1}, \ldots, \overline{p}_{mn2})
\]
\[
\ldots
\]
\[
\overline{p}_{mn2} = g(\overline{p}_{mn2})
\]

The matrix is diagonal and nonsingular, therefore, solution exists. Q.E.D. ■

6.1 Appendix B: Derivation Asymptotic Variance

The proof of Theorem 1 will follow from checking the conditions for Theorem 8.12 in Newey and McFadden(1994). We Assume A1-A7 along with the following additional Assumptions. Note that each element of \( \psi \) is a conditional expectation, let redefine each element as \( \psi^j(z^j) = f_{z^j}(z^j)E[\hat{d}_{nt} \mid z^j] \) where hence \( \hat{d}_{nt} = [1, d_{nt}]' \) for the estimation \( p_{nt} \), \( \hat{d}_{nt} = [d^{(r)}_{n't}, d^{(r)}_{nt}]' \) for the estimation of \( p^{(r)}_{n't} \) and \( \hat{d}_{nt} = [d_{n't}, d_{mnt}d_{mnt+1}]' \) for the estimation of \( \overline{p}_{mnt+1} \). Therefore

\[
\psi^j(N)(z^j) = \frac{1}{N} \sum_{n=1}^N \hat{d}_{nt}^i K_{\delta N}(z^j - z^j_n).
\]
A8: (Boundedness) (i) Each element of $m_n(\Theta, \psi)$ is bounded almost surely so that

$$E[\|m_n(\Theta, \psi)\|^2] < \infty; \quad \text{(ii)} \quad E[Z_m^T Z_n] < \infty, \quad E[X_m^T X_n] < \infty, \quad E[\exp((1-\alpha)\phi_{nt})] < \infty, \quad E[\gamma_m] < \infty, \quad E[\nabla_{\theta_{mn}} \bar{p}_{mn,t+1}] < \infty, \quad E[X_{mn}] < \infty \text{ for } m = 1, \ldots, M \quad \text{(iii)} \quad \bar{p}_{nt}, \bar{p}_{mnt}, \bar{p}_{mnt+1}, \in (0, 1), \text{ for } k \in \{0, 1\}, \ r = 1, \ldots, \rho \text{ and } m = 1, \ldots, M \quad \text{(iv)} \quad E[\nabla_{\theta_{mn}} (z_i^j)] < \infty \quad \text{and} \quad E[\nabla_{\theta_{mn}} (d_n^2 | z)] < \infty$$

**Proof of Theorem 1.** We first check the various boundedness requirement of theorem 8.12 in Newey and McFadden (1994). By assumption A8(i) we have that $E[\|m_n(\Theta, \psi)\|^2] < \infty$. It obvious by inspection that $m_n(\Theta, \psi)$ is continuously differentiable in $\Theta$ and by A8(ii-iv) that $E[\nabla_{\theta} m_n(\Theta, \psi)] < \infty$. Additionally, $\nabla_{\psi} m_n(\Theta_0, \psi_0)$ is also bounded so that $E[\|\nabla_{\psi} m_n(\Theta_0, \psi_0)\|^2] < \infty$.

Second, consider a pointwise Taylor expansion for the $j^{th}$ element of $m_n$

$$m^j(\omega, \psi) = m^j(\omega, \psi_0) + \nabla_{\psi} m^j(\omega, \psi_0)(\psi(z) - \psi_0(z)) + (\psi(z) - \psi_0(z))^T \nabla_{\psi} m^j(\omega, \psi_0)(\psi(z) - \psi_0(z))$$

$$ + o(\|\psi(z) - \psi_0(z)\|^2)$$

where the norm over the $\psi$ is the sup-norm. Next, note that

$$\left| m^j(\omega, \psi) - m^j(\omega, \psi_0) - \nabla_{\psi} m^j(\omega, \psi_0)(\psi(z) - \psi_0(z)) \right| \leq \| (\psi(z) - \psi_0(z))^T \nabla_{\psi} m^j(\omega, \psi_0)(\psi(z) - \psi_0(z)) \|$$

$$ + o(\|\psi(z) - \psi_0(z)\|^2)$$

$$\leq \| \psi - \psi_0 \|^2 \| \nabla_{\psi} m^j(\omega, \psi_0) \| + o(\|\psi - \psi_0\|^2)$$

using the triangle inequality and the Cauchy-Schwartz inequality. Therefore for $\|\psi - \psi_0\|$ small enough

$$\left| m^j(\omega, \psi) - m^j(\omega, \psi_0) - \nabla_{\psi} m^j(\omega, \psi_0)(\psi(z) - \psi_0(z)) \right| \leq \| \psi - \psi_0 \|^2 \| \nabla_{\psi} m^j(\omega, \psi_0) \|$$

So that

$$\| m(\omega, \psi) - m(\omega, \psi_0) - \nabla_{\psi} m(\omega, \psi_0)(\psi(z) - \psi_0(z)) \| \leq \| \psi - \psi_0 \|^2 \| \nabla_{\psi} m(\omega, \psi_0) \|$$

$$\| m(\omega, \psi) - m(\omega, \psi_0) - \nabla m(\omega, \psi_0)(\psi(z) - \psi_0(z)) \| \leq \| \psi - \psi_0 \|^2 \| \nabla_{\psi} m(\omega, \psi_0) \|$$

Hence $\Gamma(\omega, \psi - \psi_0) = \nabla_{\psi} m(\omega, \psi_0)(\psi(z) - \psi_0(z))$ and $\Psi(\omega) = \| \nabla_{\psi} m(\omega, \psi_0) \|$. It follows that both $\Gamma(\omega, \psi - \psi_0)$ and $\Psi(\omega)$ from the boundedness conditions established above. Next we establish the form of the influence function. Note that we have

$$\int \Gamma(\omega, \psi) F_0(d\omega) = \int f(z) E[\nabla_{\psi} m(\omega, \psi_0) \mid z] \psi(z) dz$$

$$= \int \nu(z) \psi(z)$$

39
where $v(z) = f_z(z)E[\nabla \psi m(\omega, \psi_0) \mid z]$. So by the arguments on p.2208 of Newey and McFadden(1994) we have the influence function for $m(\omega, \psi^{(N)})$ to be

$$\Upsilon(\omega) = v(z) - E[v(z)\tilde{d}]$$

$$= f_z(z)E[\nabla \psi m(\omega, \psi_0) \mid z] - E[f_z(z)E[\nabla \psi m(\omega, \psi_0) \mid z]\tilde{d}]$$

Again by the boundedness of $\nabla \psi m(\omega, \psi_0)$ it follows that $\int \|v(z)\|dz < \infty$. Finally assumption A7 guarantees that the Jacobian term converges.

**References**


[18] Fudenberg D. and J. Tirole (1996), Game Theory, MIT Press,


Table 1: Consumption Equation
\[ \ln(c_{nt}) = 1/(1 - \alpha) \left[ z_{nt}' B_4 - \ln(\lambda_n \lambda_t) + \epsilon_{2nt} \right] \]

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>PARAMETER</th>
<th>ESTIMATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>RISK AVERTION</td>
<td>(\alpha)</td>
<td>0.636 (\text{E} - 04)</td>
</tr>
<tr>
<td>Socioeconomic variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(FAM_{nt})</td>
<td>((1 - \alpha)^{-1} B_{41})</td>
<td>0.0253 (\text{E} - 04)</td>
</tr>
<tr>
<td>(YKID_{nt})</td>
<td>((1 - \alpha)^{-1} B_{42})</td>
<td>0.0014 (\text{E} - 04)</td>
</tr>
<tr>
<td>(OKID_{nt})</td>
<td>((1 - \alpha)^{-1} B_{43})</td>
<td>-0.0013 (\text{E} - 04)</td>
</tr>
<tr>
<td>(AGE_{nt}^2)</td>
<td>((1 - \alpha)^{-1} B_{24})</td>
<td>-1.20 (\text{E} - 04) (\text{E} - 05)</td>
</tr>
<tr>
<td>Region Dummies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(NE_{nt})</td>
<td>((1 - \alpha)^{-1} B_{45})</td>
<td>-0.0076 (\text{E} - 04)</td>
</tr>
<tr>
<td>(SO_{nt})</td>
<td>((1 - \alpha)^{-1} B_{46})</td>
<td>-0.0041 (\text{E} - 04)</td>
</tr>
<tr>
<td>(W_{nt})</td>
<td>((1 - \alpha)^{-1} B_{26})</td>
<td>-0.0023 (\text{E} - 04)</td>
</tr>
</tbody>
</table>
Estimated Aggregate Shock to Productivity
<table>
<thead>
<tr>
<th>Variable</th>
<th>Professional</th>
<th>Nonprofessional</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{nt}$</td>
<td>183392</td>
<td>100688</td>
</tr>
<tr>
<td></td>
<td>(2560)</td>
<td>(967)</td>
</tr>
<tr>
<td>$h_{nt}^2$</td>
<td>-251162</td>
<td>-88891</td>
</tr>
<tr>
<td></td>
<td>(4908)</td>
<td>(2152)</td>
</tr>
<tr>
<td>$h_{nt-1}$</td>
<td>14252</td>
<td>12394</td>
</tr>
<tr>
<td></td>
<td>(808)</td>
<td>(340)</td>
</tr>
<tr>
<td>$h_{nt-2}$</td>
<td>6086</td>
<td>3969</td>
</tr>
<tr>
<td></td>
<td>(730)</td>
<td>(330)</td>
</tr>
<tr>
<td>$age_{nt}^2$</td>
<td>-36</td>
<td>-13</td>
</tr>
<tr>
<td></td>
<td>(1.5)</td>
<td>(0.7)</td>
</tr>
<tr>
<td>$age_{nt} \times education_{nt}$</td>
<td>-23</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>(14)</td>
<td>(6.6)</td>
</tr>
<tr>
<td>Switching cost</td>
<td>3032</td>
<td>875</td>
</tr>
<tr>
<td></td>
<td>(171)</td>
<td>(70)</td>
</tr>
</tbody>
</table>
Table 3 Fixed Cost to Labor Participation

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>ESTIMATE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Socioeconomic variables</strong></td>
<td></td>
</tr>
<tr>
<td>$FAM_{nt}$</td>
<td>$-0.0625$</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
</tr>
<tr>
<td>$YKID_{nt}$</td>
<td>$-0.713$</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
</tr>
<tr>
<td>$YKID_{nt} \times male dummy_{nt}$</td>
<td>$0.863$</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
</tr>
<tr>
<td>$OKID_{nt}$</td>
<td>$-0.413$</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
</tr>
<tr>
<td>$OKID_{nt} \times male dummy_{nt}$</td>
<td>$0.477$</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
</tr>
<tr>
<td>$AGE_{nt}$</td>
<td>$0.163$</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
</tr>
<tr>
<td>$AGE_{nt}^2$</td>
<td>$-0.003$</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
</tr>
<tr>
<td>$EDUC_{nt}$</td>
<td>$0.08$</td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
</tr>
<tr>
<td>$EDUC_{nt} \times male dummy_{nt}$</td>
<td>$-0.03$</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
</tr>
<tr>
<td>$MS_{nt}$</td>
<td>$0.205$</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
</tr>
<tr>
<td>$SP.EDUC_{nt} \times MS_{nt}$</td>
<td>$-0.088$</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
</tr>
<tr>
<td>$SP.EDUC_{nt} \times MS_{nt} \times male dummy_{nt}$</td>
<td>$0.145$</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
</tr>
<tr>
<td>Variable</td>
<td>Estimate</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>$l_{nt}$</td>
<td>-4.4558</td>
</tr>
<tr>
<td>$FAM_{nt} \times l_{nt}$</td>
<td>0.082</td>
</tr>
<tr>
<td>$YKID_{nt} \times l_{nt}$</td>
<td>-0.1033</td>
</tr>
<tr>
<td>$YKID_{nt} \times l_{nt} \times \text{male dummy}_{nt}$</td>
<td>0.933</td>
</tr>
<tr>
<td>$OKID_{nt} \times l_{nt}$</td>
<td>0.098</td>
</tr>
<tr>
<td>$OKID_{nt} \times l_{nt} \times \text{male dummy}_{nt}$</td>
<td>-0.141</td>
</tr>
<tr>
<td>$AGE_{nt} \times l_{nt}$</td>
<td>-0.045</td>
</tr>
<tr>
<td>$AGE^2_{nt} \times l_{nt}$</td>
<td>0.0005</td>
</tr>
<tr>
<td>$EDUC_{nt} \times l_{nt}$</td>
<td>0.0504</td>
</tr>
<tr>
<td>$EDUC_{nt} \times l_{nt} \times \text{male dummy}_{nt}$</td>
<td>-0.225</td>
</tr>
<tr>
<td>$MS_{nt} \times l_{nt}$</td>
<td>0.198</td>
</tr>
<tr>
<td>$MS_{nt} \times SP.EDUC_{nt} \times l_{nt}$</td>
<td>-0.0398</td>
</tr>
<tr>
<td>$MS_{nt} \times SP.EDUC_{nt} \times l_{nt} \times \text{male dummy}_{nt}$</td>
<td>0.0956</td>
</tr>
</tbody>
</table>
Table 5 Non-Separability in Utility of Leisure/Home Production

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{nt}^2$</td>
<td>-0.214</td>
<td>42553 (12376)</td>
</tr>
<tr>
<td>$l_{nt}$</td>
<td>2.423</td>
<td></td>
</tr>
<tr>
<td>$l_{nt} \times l_{nt-1}$</td>
<td>3.479</td>
<td></td>
</tr>
<tr>
<td>$l_{nt} \times l_{nt-1} \times \text{male dummy}_{nt}$</td>
<td>2.357</td>
<td></td>
</tr>
<tr>
<td>$l_{nt} \times l_{nt-2}$</td>
<td>-2.575</td>
<td></td>
</tr>
<tr>
<td>$l_{nt} \times l_{nt-2} \times \text{male dummy}_{nt}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard deviation 42553 (12376)
### Table 6 Decomposition of the Gender Earnings Gap

*(Median Women Earnings over Median Men Earnings (%))*

<table>
<thead>
<tr>
<th>Source</th>
<th>Professional</th>
<th>Nonprofessional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw</td>
<td>87</td>
<td>76</td>
</tr>
<tr>
<td>Predicted</td>
<td>92</td>
<td>81</td>
</tr>
<tr>
<td>Human Capital</td>
<td>60</td>
<td>56</td>
</tr>
<tr>
<td>Beliefs</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>fixed effect</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Other</td>
<td>10</td>
<td>4</td>
</tr>
</tbody>
</table>

### Table 7 Decomposition of Change in the Gender Earnings Gap


<table>
<thead>
<tr>
<th>Source</th>
<th>Professional</th>
<th>Nonprofessional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw</td>
<td>30</td>
<td>25</td>
</tr>
<tr>
<td>Predicted</td>
<td>28</td>
<td>22</td>
</tr>
<tr>
<td>Human Capital</td>
<td>67</td>
<td>65</td>
</tr>
<tr>
<td>Beliefs</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>other</td>
<td>25</td>
<td>29</td>
</tr>
</tbody>
</table>

### Table 8 Decomposition of Change in Human Capital Source of Gender Earnings Gap


<table>
<thead>
<tr>
<th>Source</th>
<th>Professional</th>
<th>Nonprofessional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switching Cost</td>
<td>38</td>
<td>32</td>
</tr>
<tr>
<td>Private information</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>Demographic</td>
<td>28</td>
<td>34</td>
</tr>
<tr>
<td>Home Production shock</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Production shock</td>
<td>18</td>
<td>11</td>
</tr>
</tbody>
</table>