Essays in Labor Economics

by

Xiaomin Fu

Department of Economics
Duke University

Date: ____________________
Approved:

____
Peter Arcidiacono, Co-supervisor

____
Arnaud Maurel, Co-supervisor

____
V. Joseph Hotz

____
Marjorie B. McElroy

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2017

Abstract

Essays in Labor Economics

by

Xiaomin Fu

Department of Economics
Duke University

Date: ______________________
Approved:

____
Peter Arcidiacono, Co-supervisor

____
Arnaud Maurel, Co-supervisor

____
V. Joseph Hotz

____
Marjorie B. McElroy

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2017

Abstract

This dissertation presents two essays in labor economics. In the first essay, I study employer learning in a labor market with dynamic statistical discrimination on the basis of time-varying worker characteristics such as marital status. In the second essay, I explore the relationship between workplace flexibility and worker and occupation characteristics. These essays provide insights into the information frictions in the labor market and the cost of providing job amenities.

Dedication

To my family and friends, for supporting me.

Contents

List of Tables

Acknowledgements

I am extremely grateful to the co-chairs of my committee, Peter Arcidiacono and Arnaud Maurel, for their constant support and guidance as well as the rest of my committee members, V. Joseph Hotz and Marjorie McElroy, for contributing their insights. I thank participants in the Labor Economics lunch seminar at Duke as well as discussants in the 2015 Southern Economic Association Annual Meeting for helpful comments. Also, I would like to thank Claudia Goldin, Alexandra Killewald and Audrey Light for sharing their computer codes and data with me. Finally, I thank DuPRI, Duke SSRI, Duke Graduate School and Economics Department for providing funding throughout my graduate studies.

1

Introduction

In this dissertation, I focus on two important issues in the labor market — i.e. information frictions and the cost of providing job amenities. I study their differential impacts on various groups of workers and the implications for wage determination and inequality.

In Chapter 2, I develop an employer learning model with dynamic statistical discrimination on the basis of time-varying worker characteristics. I generalize the traditional employer learning model, which only accommodates time-invariant worker characteristics such as race. I design an empirical test for the generalized model against the alternative assumption of complete information in the labor market. Moreover, I establish the identification of the treatment effect of time-varying worker characteristics on worker productivity in a panel data setting under the assumptions of the generalized employer learning model, and propose treatment effect estimators that are robust to the information structure of the labor market. I show that individual fixed effects fail to identify the treatment effect in the presence of incomplete information and employer learning. I consider marital status as an empirical example of time-varying worker characteristics and implement the empirical test using a sample of NLSY79. The results support the hypothesis of employer learning and dynamic statistical discrimination on the basis of marital status. I obtain treatment effect estimates that are larger than the individual fixed effects estimate, suggesting that the latter is downward biased in the setting of this paper.

In Chapter 3, I explore the relationship between workplace flexibility and worker and occupation characteristics. Using an ACS sample of college-educated full-time workers, I find that women, women with children, married individuals and advanced degree holders are likely to work in flexible occupations. I explain the varying cost of providing flexibility across occupations with differences in required skill levels and work context. I apply PCA to O*NET occupation characteristics data and obtain indices of motor, social and cognitive skills as well as white-collar, pink-collar and blue-collar work context. My results suggest that occupations requiring high motor skills or low cognitive skills are less flexible. Holding skill levels constant, occupations conducted in blue-collar and white-collar work settings are less flexible. These results highlight the importance of both the human capital theory and personnel economics in understanding the differences in workplace flexibility across occupations.

2

Employer Learning and Dynamic Statistical Discrimination: Theory and Application to Marriage

2.1 Introduction

In the traditional model of employer learning (Farber and Gibbons 1996; Altonji and Pierret 2001, hereinafter AP), employers do not perfectly observe the productivity of a new worker and therefore have an incentive to statistically discriminate using easily observable worker characteristics (such as race and education) as signals of productivity when setting wages for new workers.[1] As employers gradually learn about productivity by observing the worker's performance over time, wages become more reflective of true productivity and less dependent on worker characteristics initially used to predict productivity. Therefore, as AP first point out, the hypothesis of employer learning with statistical discrimination (EL-SD) can be empirically tested by regressing log wages on easily observable worker characteristics and correlates of

[1] In this paper, I reserve the term "signals of productivity" for worker characteristics used by employers to infer worker productivity in an event of statistical discrimination. To avoid confusion, I do not use the term in this paper to refer to performance measures which also provide information about worker productivity.

productivity that are not directly observed by employers. The traditional EL-SD model implies that the coefficients on easily observable worker characteristics should fall, and the coefficients on hard-to-observe correlates of productivity should rise.

One important yet often overlooked assumption in this classical EL-SD model is that easily observable worker characteristics are time-invariant. Under this assumption, all information contained in these signals of productivity is extracted in the first period, and therefore statistical discrimination is a static event. As employers learn about the worker's productivity over time, the initial effect of statistical discrimination fades away, which is the essence of the empirical test for the classical employer learning model. This model has been widely used in the empirical employer learning literature to study statistical discrimination on the basis of race and education.[2]

In this paper, I relax this assumption and develop a more flexible EL-SD model where easily observable worker characteristics transition over time. This model allows for dynamic selection, where the transition pattern of worker characteristics—which depends on worker productivity—varies over time. In this case, worker characteristics in every period potentially contain new information about productivity. I assume that employers observe the whole history of worker characteristics and use it to statistically discriminate throughout a worker's career until employer learning is completed, taking into account the time-varying correlation between worker characteristics and productivity. In this setting, statistical discrimination is a dynamic and continual event which extends beyond the first period. This is the first paper that considers employer learning with dynamic statistical discrimination on the basis of time-varying worker characteristics.

Allowing for time-varying worker characteristics in the EL-SD model is interesting for three reasons. First, it paves the way for empirical studies of statistical

[2] See, for example, Lange (2007), Arcidiacono, Bayer, and Hizmo (2010), and Light and McGee (2015).

discrimination on the basis of time-varying worker characteristics such as marital status, parenthood, location and neighborhood. Among these, I focus on marital status in the rest of this paper.

Second, it creates new challenges of testing for employer learning with dynamic statistical discrimination. When worker characteristics are time-varying, the traditional test for EL-SD no longer delivers clear theory implications, since the coefficients on time-varying worker characteristics are complicated by the change in the correlation between worker characteristics and productivity over time. And the pattern of this change in correlation is determined by the transition process of worker characteristics, which is outside the employer learning model. Therefore, it is unclear whether the EL-SD model with time-varying signals has testable implications and what the proper empirical test should be.

Third, it opens up the discussion of treatment effect identification in a panel data setting with incomplete information. In the marriage wage premium literature, individual fixed effects regressions are commonly used to estimate treatment effects of marriage on log wages.[3] However, these studies are silent about the information structure of the labor market, and it is unclear whether the information setting actually affects treatment effect identification. Given that the employer learning literature has provided ample evidence for incomplete information in the labor market, it is especially important to know whether individual fixed effects identify the treatment effect in an incomplete information setting.

Along these lines, I make two separate theoretical contributions, first to the employer learning literature and second to the treatment effect identification literature. The first contribution is that I derive testable implications of the EL-SD model with time-varying signals, based on which I provide an empirical test for the model against

[3] Examples of such papers include Rodgers and Stratton (2009) and Killewald and Gough (2013) among others.

the alternative hypothesis of complete information in the labor market. Just as the traditional EL-SD model, the key features and implications of this model are that certain productive traits of the worker are initially unobserved and unpredictable by employers, and that as new information becomes available over time, employers are able to better predict productivity and therefore the prediction becomes more correlated with ability measures that are not directly observed by employers. Unlike in the traditional model, however, this no longer translates into well-defined trends in coefficients on worker characteristics and ability measures in the log wage regression, since the coefficients are complicated by the fluctuating correlation between time-varying worker characteristics and productivity.

To solve this problem, I regress log wages in all periods on worker characteristics in the *initial* period with time-dependent coefficients, also including ability measures to capture the effect of employer learning. I show that in theory, the trends of coefficients in this regression are the sum of two separate components: the effect of employer learning, and the transition pattern of worker characteristics scaled by the treatment effect parameter. Since the first component holds the key implication of the EL-SD model, I need to separately identify the two components in order to obtain an empirical test for the model. The transition pattern of worker characteristics is simply identified by an auxiliary regression of worker characteristics in all periods on the same characteristics in the first period and the same ability measures. Based on these, I find that the effect of employer learning is identified by solving a system of equations involving the coefficients from these two regressions above and another theory-implied relationship. I thereby obtain an empirical test for employer learning with dynamic statistical discrimination on the basis of time-varying worker characteristics.

The second theoretical contribution is to shed light on the treatment effect identification in a panel data setting under the assumptions of the employer learning model in this paper. I first show that solving the system of equations above yields a treatment effect estimator robust to the information structure of the labor market. Furthermore, I find that identification can also be achieved without involving ability measures. Specifically, I regress log wages only on worker characteristics in the initial period with time-dependent coefficients, not including ability measures to capture the effect of employer learning. In this regression the rate of change in coefficients is solely determined by the transition pattern of worker characteristics scaled by the treatment effect parameter, where the transition pattern can be identified by the auxiliary regression of worker characteristics in all periods on those in the first period with time-dependent coefficients. Therefore, the treatment effect parameter is simply identified by the derivative of coefficients on worker characteristics in the log wage regression divided by that in the auxiliary regression. I prove that the estimators derived from these two identification strategies are in fact mathematically equivalent. Finally, I show that individual fixed effects fail to identify the treatment effect in the EL-SD model, and that they only achieve identification under the alternative assumption of complete information. The reason is that in the employer learning model, the employer's updated belief enters the log wage equation as a time-varying and unobserved variable that is correlated with worker characteristics, which causes biases in individual fixed effects estimates.

In this paper I consider marital status as an empirical example of time-varying signals and implement the empirical test using a sample of white male high school graduates from the 1979 National Longitudinal Survey of Youth (NLSY79). My test results support the EL-SD model with marital status as a time-varying signal of productivity. Moreover, I estimate the treatment effect of marriage under the assumptions of the EL-SD model according to the identification strategies established in this paper. My baseline estimates from both estimators are around 17%, which are statistically significant. These information-structure-robust estimates are considerably larger than the individual fixed effects estimate, which is around 5% in the same sample. This suggests that the individual fixed effects estimate is downward biased in the setting of this paper. One possible explanation is that workers who prioritize work over personal life are more productive but less likely to be married on average. Therefore, as employers gradually learn about this hard-to-observe productive trait, their updated belief of this trait will be negatively correlated with the worker's marital status, causing a downward bias in the individual fixed effects estimate.

The rest of this paper is organized as follows. In Section 1.2, I present the employer learning model with time-varying worker characteristics, derive its testable implications and establish the identification of the treatment effect parameter. Based on these, I propose an empirical test for the employer learning model and design treatment effect estimators that are robust to the information structure. I also present several alternative empirical strategies and show that they do not serve the intended purposes. Section 2.3 describes my sample of NLSY79 and formulates empirical specifications. Section 2.4 presents baseline empirical test results and treatment effect estimates. Section 2.5 conducts robustness checks and Section 1.6 concludes.

2.2 Theory

2.2.1 The Model

In this subsection I present an employer learning model with time-varying worker characteristics and derive the log wage equation. I directly generalize the model in AP and follow their notations closely.

Denote worker i's log productivity by y_{it}:[4]

$$y_{it} = rs_{it} + \alpha_1 q_i + \Lambda z_i + \eta_i + H(t), \tag{2.1}$$

[4] In this paper, changes in the worker's productivity over time are observed by the employer. This is a major difference from Kahn and Lange (2014) where innovations in worker productivity are not perfectly observed by the employer and where worker productivity is a moving target which the employer continuously struggles to learn about.

where t denotes labor market experience. In (2.1), s_{it} represents worker characteristics observed by both the employer and the econometrician, which I allow to be time-varying. This key generalization of the classical model enables me to study worker characteristics such as marital status. I allow the transition process of s_{it} to be arbitrary, which leaves room for dynamic selection. Parameter r is by definition the causal effect of worker characteristics on log productivity. Since I consider marital status as an empirical example of time-varying worker characteristics, parameter r can be interpreted as the productivity enhancement effect of marriage, or the treatment effect of marriage. I assume that parameter r is known to the employer. q_i denotes variables observed by the employer but not observed or used by the econometrician. An example of such variables is interview performance at the time of hiring. z_i consists of variables not directly observed by the employer but observed and used by the econometrician. The most commonly used z variable in the employer learning literature is the AFQT score. η_i is an index of other determinants of productivity that are not directly observed by the employer and not observed or used by the econometrician. $H(t)$ is the experience profile of productivity, which is assumed to be observed by the employer and independent of s_{it}, q_i, z_i and η_i. I suppress the i subscript in the analysis below.

Not directly observing z and η, the employer forms expectations of z and η using all available and relevant information up to each period. At the time when worker i enters the labor market (i.e. when $t = 0$), the employer observes the worker's initial marital status s_0 and interview performance q. With these two pieces of information, the employer forms conditional expectations $\mathrm{E}(z|s_0, q)$ and $\mathrm{E}(\eta|s_0, q)$, which I assume to be linear in s_0 and q. Therefore, z and η can be expressed as

$$z = \mathrm{E}(z|s_0, q) + v_0 = \gamma_1 q + \gamma_2 s_0 + v_0, \quad (2.2)$$

$$\eta = \mathrm{E}(\eta|s_0, q) + e_0 = \alpha_2 s_0 + e_0, \quad (2.3)$$

where the error terms v_0 and e_0 have mean 0 and are uncorrelated with q and s_0 by definition.[5] Substituting z and η in (2.1) gives

$$y_t = rs_t + (\Lambda\gamma_2 + \alpha_2)s_0 + (\alpha_1 + \Lambda\gamma_1)q + H(t) + (\Lambda v_0 + e_0). \tag{2.4}$$

Each period the employer observes a new draw of marital status s_t for each worker. In the presence of dynamic selection where individuals of various productivity levels transition into marriage at different points in time, each new draw of s_t contains new information about productivity components z and η, and the whole history of marital status is relevant when the employer forms conditional expectations of worker productivity. Denote the marital status history up to period t as $S_t = \{s_0, s_1, \cdots, s_t\}$, which I assume to be perfectly observed by the employer in period t. Moreover, each period the employer also observes a noisy indicator of the worker's productivity $\xi_t = y_t + \epsilon_t$, where ϵ_t is white noise reflecting transitory shocks to the worker's performance. Since $\Lambda v_0 + e_0$ is the only factor unknown to the employer in y_t, observing ξ_t is equivalent to observing $d_t = \Lambda v_0 + e_0 + \epsilon_t$. Denote the worker's performance history as $D_t = \{d_1, d_2, \cdots, d_t\}$, which is also assumed to be observed by the employer in period t. Upon receiving new information in each period, the employer updates the belief about the worker's productivity, or equivalently, about the initial assessment error $\Lambda v_0 + e_0$. Denote the remaining error in the employer's belief as $\mu_t = \Lambda v_0 + e_0 - \mathrm{E}(\Lambda v_0 + e_0 | S_t, q, D_t)$. By definition, μ_t is uncorrelated with S_t, q, and D_t. I assume in addition that μ_t is independent of S_t, q, and D_t. I maintain the public learning assumption which states that all employers in the labor market have the same information set.[6]

[5] Variable q is excluded from Equation (2.3) simply because one can always normalize η and α_1 such that η is mean independent of q.

[6] Several papers test for the public learning assumption and find mixed evidence. See, for example, Bauer and Haisken-DeNew (2001), Galindo-Rueda (2003), Schönberg (2007), Zhang (2007), Pinkston (2009), and Kahn (2013). Using a sample of NLSY79, Schönberg (2007) and Zhang (2007) cannot reject the public learning assumption for high school graduates.

In a competitive labor market, equilibrium wage is given by

$$W_t = \mathrm{E}(Y_t|S_t, q, D_t)\exp^{\zeta_t}, \tag{2.5}$$

where Y_t is the level of productivity $\exp^{y_t}$, $\mathrm{E}(Y_t|S_t, q, D_t)$ is expected productivity given S_t, q and D_t, and $\exp^{\zeta_t}$ represents factors outside the model that are unrelated to S_t, z and q. Substituting (2.4) into (2.5) and taking logs, I obtain the data generating process (DGP) for log wages

$$w_t = rs_t + (\Lambda\gamma_2 + \alpha_2)s_0 + H^*(t) + (\alpha_1 + \Lambda\gamma_1)q + \mathrm{E}(\Lambda v_0 + e_0|S_t, q, D_t) + \zeta_t, \tag{2.6}$$

where $w_t = \log(W_t)$ and $H^*(t) = H(t) + \log(\mathrm{E}(\exp^{\mu_t}))$. Notice that if s_t is constant over time, the log wage process reduces to the one in AP, which makes perfect sense given that my model is otherwise identical to theirs.

Equation (2.6) reveals the mechanism through which each wage determinant affects wages. The worker's current marital status s_t affects wages through the causal effect of marriage on the worker's productivity. The initial marital status s_0 is used by the employer to statistically discriminate among new workers and therefore enters the wage equation through its correlation with productive traits z and η which are unobserved to the employer. The variable q affects wages both because it directly affects productivity and because it is used by the employer to infer hard-to-observe productive traits. The conditional expectation $\mathrm{E}(\Lambda v_0 + e_0|S_t, q, D_t)$ captures the effect of employer learning. As new information arrives through realizations of marital status and work performance in each period, the employer is able to better predict the worker's productivity, and this process of employer learning causes wages to converge to true productivity over time. The experience profile $H^*(t)$ and transitory variations ζ_t are also included in the log wage equation.

2.2.2 Testable Implications and Treatment Effect Identification

In this subsection, I find testable implications of the above model and propose an empirical test. I also establish identification of the treatment effect parameter r under the assumptions of the above model and provide treatment effect estimators that are robust to the information structure of the labor market. I assume that s and z are scalars in the analysis below.

Empirical Strategy (A): Regress w_t on s_0 and z

Consider the conditional expectation function

$$\mathrm{E}(w_t|s_0, z, t) = b_{st}s_0 + b_{zt}z + H^*(t). \tag{2.7}$$

Given the true DGP for log wages in Equation (2.6), the omitted variable bias formula implies that

$$b_{st} = (\Lambda\gamma_2 + \alpha_2) + r\Phi_t^{ss} + \Phi_{qs} + \Phi_{st}, \tag{2.8}$$

$$b_{zt} = r\Phi_t^{sz} + \Phi_{qz} + \Phi_{zt}, \tag{2.9}$$

where Φ_t^{ss} and Φ_t^{sz} denote the coefficients of the auxiliary regression of s_t on s_0 and z; Φ_{qs} and Φ_{qz} are coefficients of the regression of $(\alpha_1 + \Lambda\gamma_1)q$ on s_0 and z; and Φ_{st} and Φ_{zt} are coefficients of the regression of $\mathrm{E}(\Lambda v_0 + e_0|S_t, q, D_t)$ on s_0 and z.

Equations (2.8) and (2.9) shed light on how each wage determinant comes into play in the log wage regression coefficients b_{st} and b_{zt}. First, notice that when $t = 0$, $\mathrm{E}(\Lambda v_0 + e_0|S_0, q, D_0) = 0$ because there is no work history in period 0 and that initial assessment errors, by definition, cannot be predicted by what the employer initially observes (i.e. s_0 and q), so Φ_{s0} and Φ_{z0} equal zero. Therefore,

$$b_{s0} = (\Lambda\gamma_2 + \alpha_2) + r + \Phi_{qs},$$

$$b_{z0} = \Phi_{qz}.$$

The expression of b_{s0} shows that the initial coefficient on s_0 picks up not only the treatment effect r but also two additional effects represented by the terms $(\Lambda\gamma_2 + \alpha_2)$ and Φ_{qs}. The first term, which traces back to Equations (2.2) and (2.3), reflects statistical discrimination on the basis of initial marital status s_0. It shows that s_0 correlates with wages partly because the employer uses s_0 to infer the worker's productivity. The second term, Φ_{qs}, reveals that the coefficient b_{s0} also picks up the effect of q, which is observed and used by the employer to determine wages but is unobserved by the econometrician and therefore omitted from the regression. The effect of the omitted variable q is also present in the coefficient b_{z0}, and is the only effect in this initial coefficient on z. Particularly, the causal effect of z on productivity is absent, since variable z is not directly observed by the employer and therefore not factored into wages in period 0, expect through its correlation with s_0.

Moreover, Equations (2.8) and (2.9) reveal that the coefficients on s_0 and z change over time due to marital status transitions and employer learning, which are respectively reflected in the first and third auxiliary regressions mentioned above. In the auxiliary regression of marital status s_t on s_0 and z, the time-dependent coefficients Φ_t^{ss} and Φ_t^{sz} partly reflect the time-varying correlation between s_t and ability z, which characterizes dynamic selection into marriage on the basis of ability over time. The auxiliary regression of the employer's updated belief $\mathrm{E}(\Lambda v_0 + e_0 | S_t, q, D_t)$ on s_0 and z captures the key theoretical implication of employer learning, which forms the essence of the empirical test for the EL-SD model in this paper. This is formally stated in Proposition 1.

Proposition 1. *(Testable Implications) Under the assumptions of the EL-SD model above,*

$$\frac{\partial \Phi_{zt}}{\partial t} \geqslant 0.$$

Proof. See Appendix A.1. $\square$

Proposition 1 states that the key implication of the employer learning model is that in the regression of the employer's updated belief $\text{E}(\Lambda v_0 + e_0 | S_t, q, D_t)$ on s_0 and z, the coefficient Φ_{zt} is nondecreasing in t.

As mentioned earlier, $\Phi_{zt} = 0$ when $t = 0$. As experience t increases, the marital status history (S_t) and work performance history (D_t) accumulate, both of which are potentially informative of the initial assessment error $\Lambda v_0 + e_0$. Therefore, as the information set grows over time, the employer is able to better predict the worker's productivity, and the correlation between the employer's conditional expectation of $\Lambda v_0 + e_0$ and its true value increases. This increase in correlation drives up the coefficient Φ_{zt} under two conditions. First of all, it is crucial that the ability measure z is not directly observed by the employer so that it contains part of the initial assessment error. This implies that the correlation between the employer's prediction of $\Lambda v_0 + e_0$ and ability z also increases as the prediction converges to the truth. Secondly, it is also important that the other regressor s_0 is not correlated with the employer's prediction (see Appendix A.1). Otherwise, the coefficient Φ_{zt} would also be affected by the correlation between s_0 and the prediction, in which case the theory implication for Φ_{zt} would be unclear. The intuition behind this zero correlation is that the initial marital status s_0 is observed and used by the employer to infer the worker's productivity in the beginning, and therefore s_0 is not correlated with the initial assessment error or the employer's updated belief of it in any later periods, as long as s_0 always remains in the employer's information set.

Eventually, when employer learning is completed, $\text{E}(\Lambda v_0 + e_0 | S_t, q, D_t) = \Lambda v_0 + e_0$, and the experience profile of Φ_{zt} plateaus out. It immediately follows that if the employer could perfectly observe the worker's productivity from the very beginning, the experience profile of Φ_{zt} would be flat throughout. Therefore, Proposition 1 forms the basis for testing the employer learning model in this paper against the alternative hypothesis of complete information in the labor market.[7]

Clearly, testing for the EL-SD model requires identification of the test statistic $\frac{\partial \Phi_{zt}}{\partial t}$. This does not simply follow from the definition of $\frac{\partial \Phi_{zt}}{\partial t}$, since the regression of $\mathrm{E}(\Lambda v_0 + e_0 | S_t, q, D_t)$ is infeasible. Also, Equation (2.9) shows that the experience profile of Φ_{zt} is not the same as that of b_{zt}, since parameter Φ_t^{sz} also changes over time as marital status s_t transitions. This is the fundamental challenge in testing for the employer learning model with time-varying signals in this paper. In the classical model where signal s is a constant, the experience path of Φ_{zt} is the same as that of b_{zt}, which can be directly estimated from the regression of w_t on s_0 and z with time-dependent coefficients. This is no longer true when signal s transitions. Next I present Lemma 2, which is a building block in the identification of the test statistic $\frac{\partial \Phi_{zt}}{\partial t}$ and the treatment effect r.

Lemma 2. *Under the assumptions of the EL-SD model above,*

$$\frac{\partial \Phi_{st}}{\partial t} = -\Phi_{zs} \frac{\partial \Phi_{zt}}{\partial t}, \tag{2.10}$$

where Φ_{zs} is the coefficient of the regression of z on s_0.

Proof. See Appendix A.1. □

Lemma 2 follows from the least squares regression formula. It shows that in the regression of $\mathrm{E}(\Lambda v_0 + e_0 | S_t, q, D_t)$ on s_0 and z, the coefficient on s_0 decreases as the coefficient on z increases if s_0 and z are positively correlated, and vice versa. In other words, the effect of employer learning on the coefficient on z spills over to the

[7] With complete information, the worker's productivity is perfectly observed upon labor market entry, leaving no room for statistical discrimination or employer learning. Therefore, complete information is clearly an alternative to the EL-SD model. However, one might argue that it is not the only alternative. For example, statistical discrimination might be forbidden by law, even though it is a rational behavior in an incomplete information setting. This is important when studying statistical discrimination on the basis of race, but less so in the case of marital status. In this paper I assume that the employer uses marital status as a cheap source of information whenever needed.

coefficient on s_0, precisely because s_0 is correlated with z and is observed and used by the employer to statistically discriminate among new workers. Lemma 2 also shows that the rates of change of the two coefficients are linked by the constant Φ_{zs}, which can be directly estimated from the data.

Although Lemma 2 is also a key theory implication of the EL-SD model, it is not testable in the setting of this paper, because the two derivative parameters $\frac{\partial \Phi_{st}}{\partial t}$ and $\frac{\partial \Phi_{zt}}{\partial t}$ cannot be identified without Equation (2.10). This is made clear in the next proposition, which gives the identification of the two derivative parameters as well as the treatment effect parameter r.

Proposition 3. *(Identification) Under the assumptions of the EL-SD model above, the treatment effect r and the experience profiles of Φ_{st} and Φ_{zt} are identified from the regression of w_t on s_0 and z and its auxiliary regressions.*

Proof. Taking derivatives with respect to t in Equations (2.8) and (2.9) gives

$$\frac{\partial b_{st}}{\partial t} = r\frac{\partial \Phi_t^{ss}}{\partial t} + \frac{\partial \Phi_{st}}{\partial t}, \tag{2.11}$$

$$\frac{\partial b_{zt}}{\partial t} = r\frac{\partial \Phi_t^{sz}}{\partial t} + \frac{\partial \Phi_{zt}}{\partial t}. \tag{2.12}$$

Combining these with Equation (2.10) yields a system of 3 equations with 3 unknowns (i.e. r, $\frac{\partial \Phi_{st}}{\partial t}$ and $\frac{\partial \Phi_{zt}}{\partial t}$), and identification can be achieved by solving this system of equations.

It is easy to show that this linear system has a unique solution if and only if $\frac{\partial \Phi_t^{ss}}{\partial t} + \Phi_{zs}\frac{\partial \Phi_t^{sz}}{\partial t} \neq 0$. Substituting in the OLS regressions formula for these coefficients, one can show that this rank condition is equivalent to $\frac{\partial \mathrm{cov}(s_t, s_0)}{\partial t} \neq 0$ and $|\mathrm{corr}(z, s_0)| \neq 1$. The first inequality is satisfied as long as s_t transitions, and the second simply requires the dispersion of z conditioning on s_0, both of which are satisfied in the EL-SD model above. □

This is the main theoretical result of this paper. It shows that the parameters r, $\frac{\partial \Phi_{st}}{\partial t}$ and $\frac{\partial \Phi_{zt}}{\partial t}$ are identified because they are functions of several nuisance parameters (i.e. $\frac{\partial b_{st}}{\partial t}$, $\frac{\partial b_{zt}}{\partial t}$, $\frac{\partial \Phi_t^{ss}}{\partial t}$, $\frac{\partial \Phi_t^{sz}}{\partial t}$ and Φ_{zs}) which can be directly estimated from the data.

This result achieves two main purposes. First, the identification of the test statistic $\frac{\partial \Phi_{zt}}{\partial t}$ makes it possible to empirically test for the EL-SD model in this paper where the signal s_t is time-varying. Clearly, the sign of $\frac{\partial \Phi_{st}}{\partial t}$ does not serve as an additional test, because it is determined by Equation (2.10) which is used to identify (and later to estimate) the parameters. In this setting, the system is exactly identified, and therefore by design the estimated sign of $\frac{\partial \Phi_{st}}{\partial t}$ is always consistent with the EL-SD model, which does not lend more credibility to the model. This is a major difference from the classical employer learning model. In the classical model where the signal s is constant, $\frac{\partial \Phi_t^{ss}}{\partial t} = \frac{\partial \Phi_t^{sz}}{\partial t} = 0$, and therefore $\frac{\partial \Phi_{st}}{\partial t}$ and $\frac{\partial \Phi_{zt}}{\partial t}$ are over-identified, which provides an additional test for the classical model. This is no longer true in the generalized model where s_t transitions. For this reason, I will only present the estimates of the test statistic $\frac{\partial \Phi_{zt}}{\partial t}$ in the empirical sections of this paper.

Second, the identification of the treatment effect parameter r facilitates estimation of this key parameter of interest in an incomplete information setting. The novel treatment effect estimator implied by Proposition 3 is robust to the information structure of the labor market,[8] whereas the commonly used individual fixed effects estimator is only valid in the complete information setting as I show later in Section 2.2.3. Also, notice that the identification of r is unique to the generalized EL-SD model where s_t transitions. If s_t is constant, then $\frac{\partial \Phi_t^{ss}}{\partial t} = \frac{\partial \Phi_t^{sz}}{\partial t} = 0$, in which case the parameter r is not identified.[9]

[8] Under the complete information assumption, Equations (2.10) – (2.12) hold with $\frac{\partial \Phi_{st}}{\partial t} = \frac{\partial \Phi_{zt}}{\partial t} = 0$. Therefore, solving this system of equations always yields the identification of r regardless of the labor market information setting.

[9] Lange (2007) also points out that the parameter r is not identified in the classical employer learning model with constant worker characteristics.

It turns out that the ability variable z, although crucial to the empirical test, is not essential to the identification of r. Next I show that the parameter r can also be identified without variable z.

Empirical Strategy (B): Regress w_t on s_0

Consider the conditional expectation function

$$\mathrm{E}(w_t|s_0, t) = b_{s_0 t} s_0 + H^*(t).$$

Given the true log wage process in Equation (2.6), the omitted variable bias formula implies that

$$b_{s_0 t} = (\Lambda\gamma_2 + \alpha_2) + r\Phi_t^{ss_0} + \Phi_{qs_0} + \Phi_{s_0 t},$$

where $\Phi_t^{ss_0}$, Φ_{qs_0} and $\Phi_{s_0 t}$ denote the coefficients of the auxiliary regressions of s_t, $(\alpha_1 + \Lambda\gamma_1)q$ and $\mathrm{E}(\Lambda v_0 + e_0|S_t, q, D_t)$ on s_0, respectively.

As mentioned earlier, the employer's belief $\mathrm{E}(\Lambda v_0 + e_0|S_t, q, D_t)$ is not correlated with initial marital status s_0 (see Appendix A.1), and therefore $\Phi_{s_0 t} = 0$. This is crucial for identifying r, as I show in the next proposition.

Proposition 4. *(Treatment Effect Identification) Under the assumptions of the EL-SD model above, the treatment effect r is identified from the regression of w_t on s_0 and its auxiliary regressions.*

Proof. Given that $\Phi_{s_0 t} = 0$,

$$\frac{\partial b_{s_0 t}}{\partial t} = r\frac{\partial \Phi_t^{ss_0}}{\partial t}. \tag{2.13}$$

Since s_t transitions, coefficient $\Phi_t^{ss_0}$ changes over time, and $\frac{\partial \Phi_t^{ss_0}}{\partial t} \neq 0$. Therefore, r is identified. □

This proposition shows that parameter r is identified by the ratio of two nuisance parameters estimable from the data. Since this identification strategy does not involve variable z, it reveals that z is not essential to the identification of r. In fact, it can be proved that the two estimators of r implied by Proposition 3 and Proposition 4 are mathematically equivalent in theory.

Corollary 5. *The estimators of r implied by Proposition 3 and Proposition 4 are mathematically equivalent.*

Proof. See Appendix A.2. □

This reveals that variable z does not play an active role in the identification of r even in Proposition 3. On the other hand, variable z is indeed crucial to testing for employer learning, since the regressions without z in Proposition 4 do not provide a test. The reason is that the condition $\Phi_{s_0 t} = 0$ also holds under the alternative assumption of complete information.[10]

Table 2.1 summarizes theory implications and identification results under the EL-SD model and the alternative complete information assumption respectively.

2.2.3 Alternative Empirical Strategies

In the previous subsection, I present empirical strategies that achieve the purposes of empirical testing and treatment effect identification. In this subsection, I list several alternative empirical strategies and show that they do not achieve the same purposes in the setting of this paper.

Alternative Empirical Tests for EL-SD

Regress w_t on s_t and z AP show that in the traditional employer learning model where signal s is constant over time, the model can be empirically tested by regress-

[10] To see this, recall that complete information can be considered as the limiting case of employer learning as t increases.

ing log wages on signal s and ability z with time-dependent coefficients. Since I directly generalize the traditional model by allowing signal s to vary over time, it is tempting to assume that the generalized model can still be tested by the same log wage regression simply with s_t replacing s. However, this is not the case. Following the analysis above, one can easily verify that the regression of w_t on s_t and z does not deliver meaningful theory implications. The main reason is that as individuals of various abilities select into marriage over time, marital status s_t is likely to be correlated with initial assessment errors, and therefore $\text{cov}[s_t, \text{E}(\Lambda v_0 + e_0|S_t, q, D_t)] \neq 0$. Since this covariance being zero is crucial for theory implications to come through, the regression of w_t on s_t and z does not constitute a proper empirical test for EL-SD.

Regress w_t on s and z in a Selected Subsample with Constant s Another plausible strategy is to select a subsample where marital status does not transition and then directly apply AP's empirical test with that subsample. However, this approach does not serve the intended purpose either. The fundamental reason is that the sample selection criteria defined by the researcher do not alter the true DGP in the population. Even though signal s is restricted to be a constant in the subsample, it still transitions in the population, and the latter is what matters when the employer forms expectations and sets wages. Therefore, selecting a subsample with constant s does not automatically entail the applicability of AP's empirical strategy which is designed for time-invariant signals. I prove in Appendix A.3 that the theory implications in AP do not hold in a selected subsample with constant s given the generalized EL-SD model where s_t transitions in the population. The main reason is that the key condition $\text{cov}[s_0, \text{E}(\Lambda v_0 + e_0|S_t, q, D_t)] = 0$ does not hold in the selected subsample, since the sample selection rule is defined based on the marital status transition pattern, which is correlated with the employer's initial assessment errors.

Alternative Treatment Effect Identification Strategies

Individual Fixed Effects Regression of w_t on s_t The individual fixed effects estimator is widely used to estimate treatment effects with panel data. However, I show that it is sensitive to the information assumption in the context of this paper.

In a labor market with complete information, the worker's productivity is perfectly observed by the employer upon labor market entry, and the log wage process is simply given by

$$w_t = y_t + \zeta_t = rs_t + \alpha_1 q + \Lambda z + \eta + H(t) + \zeta_t.$$

Therefore, r is identified in the regression of w_t on s_t when individual fixed effects are included to control for time-invariant individual characteristics (i.e. q, z and η).

However, under the assumptions of the EL-SD model in this paper, the log wage process is given by Equation (2.6), where $\mathrm{E}(\Lambda v_0 + e_0|S_t, q, D_t)$ is a time-varying unobserved heterogeneity term correlated with s_t.[11] Therefore, individual fixed effects regressions produce biased estimates of the treatment effect, and the direction and size of the bias depend on the sign and magnitude of $\mathrm{cov}[\tilde{s}_t, \tilde{\mathrm{E}}(\Lambda v_0 + e_0|S_t, q, D_t)]$, where tildes indicate individual-demeaned variables.[12] This covariance term is expected to be non-zero due to selection into marriage over time which implies that the transition of marital status provides information about initial assessment errors.

Therefore, I conclude that the individual fixed effects regression only identifies the treatment effect in a complete information setting. It cannot be used to estimate the treatment effect under the assumptions of the employer learning model in this paper.

[11] Formally, $\mathrm{cov}[s_t, \mathrm{E}(\Lambda v_0 + e_0|S_t, q, D_t)] = \mathrm{cov}(s_t, \Lambda v_0 + e_0) \neq 0$. The equality in Step 1 is obtained similarly to $\mathrm{cov}[s_0, \mathrm{E}(\Lambda v_0 + e_0|S_t, q, D_t)] = \mathrm{cov}(s_0, \Lambda v_0 + e_0)$, which is proved in Appendix A.1. The inequality in Step 2 follows from the assumption that s_t contains new information not initially available and therefore is potentially correlated with initial assessment errors.

[12] $\mathrm{cov}[\tilde{s}_t, \tilde{\mathrm{E}}(\Lambda v_0 + e_0|S_t, q, D_t)] = \mathrm{cov}[s_t, \mathrm{E}(\Lambda v_0 + e_0|S_t, q, D_t)] - \mathrm{cov}[\bar{s}, \bar{\mathrm{E}}(\Lambda v_0 + e_0|S_t, q, D_t)] = \mathrm{cov}(s_t, \Lambda v_0 + e_0) - \mathrm{cov}[\bar{s}, \bar{\mathrm{E}}(\Lambda v_0 + e_0|S_t, q, D_t)]$, where variables with upper bars are within-individual averages.

OLS Regression of w_t on s_t It is commonly accepted that the OLS estimate of the marriage wage premium is an upward biased estimate of the causal effect of marriage. However, this paper presents a more nuanced picture of the biases in the OLS estimate under the EL-SD model assumptions, which opens up new possibilities.

The log wage process in Equation (2.6) reveals that in the OLS regression of log wage w_t on marital status s_t, the coefficient on s_t as an estimate of r is biased due to omitting the following three terms: the initial marital status s_0, the ability measure q (e.g. interview performance) which is directly observed by the employer but not the econometrician, and the conditional expectation term $\mathrm{E}(\Lambda v_0 + e_0 | S_t, q, D_t)$, which is the employer's updated belief about initial assessment errors.

In the empirical context of this paper, the first bias term is expected to be positive, since initial marital status s_0 is positively correlated with succeeding marital status s_t. The second bias term is also expected to be positive given that ability q is likely to be positively correlated with marital status s_t,[13] since certain traits demonstrated by a productive worker are also desirable in the marriage market. This is essentially the classic selection bias problem where wage determinants unobserved by the econometrician are positively correlated with marriage, causing the OLS estimate to be upward biased in a labor market with complete information.

However, this paper shows that the process of employer learning produces a third bias term, which is captured by the correlation between marital status s_t and initial employer assessment errors.[14] Recall that initial assessment errors are productive traits that can be neither observed nor predicted by the employer at the beginning of a worker's career. One example of such traits is the tendency to put work first

[13] I assume that the parameters associated with s_0 and q in Equation (2.6) are positive, which is likely to be true in the context of this paper.

[14] Note that $\mathrm{cov}[s_t, \mathrm{E}(\Lambda v_0 + e_0 | S_t, q, D_t)] = \mathrm{cov}(s_t, \Lambda v_0 + e_0)$. See Footnote 11. Formally speaking, $\mathrm{cov}(s_t, \Lambda v_0 + e_0) < 0$ implies that at least one of the two initial assessment errors (i.e. v_0 and e_0) is negatively correlated with s_t, assuming that $\Lambda > 0$.

when facing the challenge of work-life balance.[15] Intuitively, this trait contributes positively to productivity at work but correlates negatively with the likelihood of marriage. In this example, the third bias term is negative, and the overall direction of bias of the OLS estimate depends on the relative size of the three bias terms.

2.3 Data and Econometric Specification

The data are drawn from the 1979-2004 waves of NLSY79. To abstract from statistical discrimination on race, gender and education, I restrict the sample to white men who have completed exactly 12 years of education upon labor market entry and do not change education level afterwards.[16] Labor market entry is defined as the year in which an individual reports to have left school for the first time (which I call Period 0, i.e. $t = 0$),[17] and potential experience t is the number of years since labor market entry. I code marital status to be 1 if an individual is currently married, and 0 otherwise (i.e. never married, separated, divorced, or widowed). I then construct variable s_0—i.e. marital status in Period 0—for each individual. For the ability

[15] This trait is understandably hard to observe or predict by the employer, since it might be unknown even to the worker himself at the outset of his career.

[16] Strictly speaking, education is also a time-varying signal. Although the model can accommodate more than one time-varying signal when s_t is a vector, the resulting empirical strategy is data-demanding. I obtain very noisy estimates when I implement the empirical strategy with both education and marital status as time-varying signals. Therefore, I choose to focus on one time-varying signal in my analysis. In general, when there are two time-varying signals, the test in this paper does not apply in a selected subsample where one of the time-varying signals is restricted to be constant. To circumvent this issue, I impose an additional assumption that the worker's entire path of future education is foreseeable by the employer at the time of hiring, and therefore all the information in education is extracted upon labor market entry. This assumption allows me to select a subsample where education is constant and apply the test to marital status, which is the time-varying signal I focus on in this paper.

[17] Light and McGee (2015) argue that this definition often picks up short-term school exits rather than permanent ones. As a result, it starts the career too early and overstates potential experience especially for highly educated workers. They propose a new definition of labor market entry—the start of the first nonenrollment spell lasting at least 12 months. As a robustness check, I repeat my analysis using their definition and find that my results are not qualitatively affected. This is not surprising given that my sample consists of high school graduates instead of highly educated workers.

measure z, I use the AFQT score and a noncognitive ability measure similar to the one in Heckman, Stixrud, and Urzua (2006). Appendix A.4 provides more details about the sample and the construction of key variables.

Table 3.1 contains summary statistics for the main variables in my sample. Noticeably, wages increase with potential experience and are higher for married individuals within each experience category. Combining all experience levels, the raw marriage wage premium in my sample is around 26% (= 6.945 – 6.686). Table 3.1 also shows that individuals who marry within the first five years of labor market experience actually have slightly lower AFQT scores on average compared with those unmarried in the same experience category, implying that marriage is a negative signal of ability in the early stage of career. This is not surprising given that my sample consists of high school graduates who enter the labor market at relatively young ages. As more individuals transition into marriage over time, the relationship between marriage and AFQT is reversed and marriage signals high ability. Moreover, the AFQT gap increases as experience grows, largely due to the decline in the average AFQT score among unmarried individuals. The gap in average noncognitive ability, though much smaller in magnitude, also changes over time. Therefore, Table 3.1 displays dynamic selection into marriage on the basis of ability, which implies that marriage sends different signals of ability in various stages of a worker's career. This motivates the model in this paper where employers use the entire history of marital status to infer the worker's productivity and statistically discriminate over the course of the worker's career (until employer learning is completed).

Next I describe the estimation procedure and econometric specifications. To implement empirical strategy (A), I first jointly estimate three regressions to obtain estimates for the five nuisance parameters (i.e. $\frac{\partial b_{st}}{\partial t}$ and $\frac{\partial b_{zt}}{\partial t}$, $\frac{\partial \Phi_t^{ss}}{\partial t}$ and $\frac{\partial \Phi_t^{sz}}{\partial t}$, and Φ_{zs}) in Equations (2.10) – (2.12), and then calculate estimates for the three parameters of interest (i.e. r, $\frac{\partial \Phi_{st}}{\partial t}$ and $\frac{\partial \Phi_{zt}}{\partial t}$), which are solutions to the linear system of equations (2.10) – (2.12). The three regressions used to estimate nuisance parameters are specified as follows. To estimate nuisance parameters $\frac{\partial b_{st}}{\partial t}$ and $\frac{\partial b_{zt}}{\partial t}$, i.e. the rates of change of coefficients b_{st} and b_{zt} in Equation (2.7), I regress log wage w_t on s_0 and z and linear interactions of potential experience t with s_0 and z, whereby I restrict coefficients b_{st} and b_{zt} to be linear in t. In this baseline linear specification, parameters $\frac{\partial b_{st}}{\partial t}$ and $\frac{\partial b_{zt}}{\partial t}$ are simply the coefficients on the interaction terms.[18] I also include a cubic in potential experience as a proxy for $H^*(t)$ and a set of standard controls—i.e. urban residence, region of residence, and a dummy for part-time versus full-time jobs.[19] Similarly, to estimate nuisance parameters $\frac{\partial \Phi_t^{ss}}{\partial t}$ and $\frac{\partial \Phi_t^{sz}}{\partial t}$, I regress marital status s_t on the same set of regressors. And to estimate the nuisance parameter Φ_{zs} in Equation (2.10), I regress z on s_0 with the same standard controls.

To implement empirical strategy (B), I first estimate nuisance parameters $\frac{\partial b_{s_0 t}}{\partial t}$ and $\frac{\partial \Phi_t^{ss_0}}{\partial t}$ in Equation (2.13) by regressing w_t and s_t respectively on the same set of regressors above excluding z and $z \times t$. Similarly, since I only include the linear interaction of t with s_0 in this baseline specification, the two nuisance parameters are simply the coefficients on the interaction term $s_0 \times t$ in the two regressions. I then calculate the ratio of the two nuisance parameter estimates to obtain the treatment effect estimate. Although in theory the estimators in (A) and (B) are equivalent, they are not exactly the same empirically due to parametric specifications of derivative parameters and adding additional controls.

Several remarks are in order. First, there is little variation in s_0, since the vast majority of high school graduates are not married at the time of labor market entry. A lack of variation in regressors reduces estimation precision. To circumvent this

[18] I relax this linearity restriction later in Section 2.5 and show that my results are not sensitive to linearity.

[19] Excluding these standard controls does not qualitatively affect my results.

issue, I relabel a later period as Period 0 and redefine variable s_0 accordingly. I then drop observations from earlier periods and carry out the analysis using the smaller sample starting from the relabeled Period 0. Therefore, relabeling a later period as Period 0 increases variation in s_0 yet reduces sample size, and the overall effect on estimation precision is ambiguous. In order to find the best balance, I sequentially relabel each period as Period 0 and repeat the empirical analysis. I find that standard errors decrease up to Period 4 and increase afterwards. Therefore, I relabel Period 4 as Period 0 in my baseline results in Section 2.4, and show results from relabeling other periods as a robustness check in Section 2.5. In theory, if the model is free from any form of misspecification, period relabeling should not affect identification of the treatment effect parameter or the qualitative conclusion from the empirical test (see Appendix A.5 for a proof), and therefore all period relabeling rules should produce similar treatment effect estimates. In Section 2.5, I show the extent to which this is true in my estimates and the reasons why this might be compromised.

Second, in addition to dropping early periods as a result of period relabeling, I also drop late periods which have fewer observations due to attrition in NLSY79. In my baseline specification, I drop observations with more than 20 years of potential experience. I also present results from dropping other late periods as a robustness check.

Third, I separately use AFQT and the noncognitive ability measure as variable z and obtain two sets of results for (A). I do not include the two measures of z at the same time because the empirical strategy in this paper does not easily generalize to the case where z is a vector. Moreover, there is no obvious benefit from including more than one z variable at the same time for my purposes. In fact, treating z as a scalar is innocuous, since I can redefine any element of z as part of the unobserved ability η.[20] This further implies that different empirical measures of z identify the

[20] Recall that η is an index of determinants of productivity that are not directly observed by the

same treatment effect parameter r, since the definition of z only affects the interpretation of variable η and parameter Λ, while other parts of the model, especially those involving parameter r, are not affected. Empirically, I find that AFQT and noncognitive ability as separate measures of z produce similar estimates of r as I show below.

2.4 Baseline Results

The baseline results are presented in Table 2.3. In Columns (A) and (B), I report results from empirical strategies (A) and (B) described above. I also estimate individual fixed effects and OLS regressions of the log wage on marital status with the same additional controls mentioned above, and report coefficients on marital status in the last two columns. For results to be comparable across columns, I use the same subsample throughout this table where I keep observations with 4 to 20 years of potential experience.

In Column (A), using AFQT as a measure of z, I estimate that marriage increases log wage by 0.1672, which is both economically and statistically significant. Using noncognitive ability as variable z, I obtain a similar-sized treatment effect estimate of 0.1578, consistent with the notion that different empirical measures of z identify the same treatment effect parameter. In both cases, I obtain positive and significant estimates of the test statistic $\frac{\partial \Phi_{zt}}{\partial t}$, which rejects the complete information assumption in favor of the employer learning model, as stated in Proposition 1.

In Column (B), I present the treatment effect estimate from empirical strategy (B). I obtain a positive and significant estimate of 0.1685, similar to those in Column (A). This is expected given that both empirical strategies identify the same treatment effect parameter. In the last two columns, I report individual fixed effects and

employer and not observed *or used* by the econometrician. Therefore, if an element of z is not used by the econometrician, it automatically becomes part of η.

OLS estimates of the coefficient on marital status in the log wage regression. With individual fixed effects, I estimate that marriage is associated with an increase in the log wage of 0.0461, which is statistically significant yet economically moderate. In the OLS regression, I estimate that the marriage gap in the log wage is 0.1660, both statistically and economically significant.

Comparing across columns, we see that my treatment effect estimates in Column (A) and (B) are larger than the individual fixed effects estimate and yet similar to the OLS estimate. This suggests that the individual fixed effects estimate is downward biased and that the biases in the OLS estimate happen to add up to zero in the setting of this paper. This confirms my intuition in Section 2.2.3 that marital status s_t is negatively correlated with initial employer assessment errors. As I mentioned earlier, this makes sense given that workers who prioritize work over personal life are more likely to be productive and yet less likely to be married. The tendency to put work first is a valid example of initial assessment errors since it can not be observed or predicted by the employer at the beginning of a worker's career.

In summary, the baseline empirical results support the employer learning model in this paper where marital status is a time-varying signal of worker productivity. I estimate the treatment effect of marriage to be around 17%, which reveals that the individual fixed effects estimate is downward biased in this setting. I show that this can arise from employer learning when productive traits initially unobserved and unpredictable by the employer (such as the tendency to put work first) are negatively correlated with the likelihood of marriage in later periods.

2.5 Robustness Checks

In the baseline results above, I relabel Period 4 as Period 0 and keep observations with 4 to 20 years of potential experience. I also impose linearity when estimating the nuisance parameters in Equations (2.10) – (2.13). In this section, I conduct robustness checks along these dimensions. Meanwhile, I provide further evidence that the results are not affected by either the measurement of z in empirical strategy (A) or the exclusion of z as in (B).

2.5.1 Alternative Period Relabeling and Sample Periods

To see how period relabeling affects results, I sequentially relabel each period as Period 0 and cut off the corresponding left tail of potential experience. For each left-censored sample, I also cut off a progressively longer right tail of potential experience. This creates a grid of subsamples possibly censored on both sides. I repeat the entire empirical analysis above for each subsample to see whether results are affected by period relabeling and censoring on either side. The results are presented in Table 2.5 through Table 2.8 (see Table 2.4 for sample sizes).

Table 2.5 reports robustness check results for empirical strategy (A) with AFQT as variable z. Each column represents a different period relabeling and left-censoring rule, while each row group imposes a different right-censoring rule. The table covers the corresponding baseline result as a special case in the column using s_4 as s_0 and the row group restricting potential experience to be no more than 20. First of all, notice that estimates of the test statistic $\frac{\partial \Phi_{zt}}{\partial t}$ are positive and significant in almost all subsamples, providing further evidence for employer learning. And the way these estimates vary across subsamples is consistent with the intuition that the rate of employer learning is higher in earlier periods, which is reassuring.

Secondly, comparing across columns within each row in Table 2.5 shows that relabeling Period 4 as Period 0 indeed produces the smallest standard error in the treatment effect estimate, which justifies this particular relabeling rule as the baseline. However, the point estimate of r is not quite stable across columns. Particularly, the estimate drops considerably when I switch from Period 4 to Period 5. This is concerning given that relabeling Period 5 produces reasonably small standard errors just as when I relabel Period 4. I find that this large gap is caused by an exceptionally high rate of wage growth among those who are single in Period 4 and married in Period 5. The large gap disappears once I drop those individuals (see Table 2.6), who make up roughly 9% of my sample.[21] I acknowledge that this does not resolve the fundamental problem of model misspecification, and yet I do not attempt to explore additional model features given the limitation of my sample size and statistical power. Apart from this large gap in estimates from relabeling Period 4 and Period 5, estimates across all columns display a general decreasing trend. This suggests that the effect of marriage might be smaller for workers who marry later in their career.[22]

Thirdly, comparing across rows within each column in Table 2.5 reveals that the point estimate of r tends to decrease as I progressively cut off a longer right tail of potential experience. This could be due to nonrandom attrition in the NLSY79 sample, given that the sample size drops sharply in late periods (see Table 3.1). It could also be the result of omitting the duration of marriage in my analysis.[23] If each additional year of marriage has a positive effect on wages, then my estimate of r is a combination of the initial effect of getting married and the continual effect of staying married. This could explain why my estimate of r is larger when later periods are included, given that the average duration of marriage is larger in later periods. Despite this caveat, the estimates are similar when cutting off the right tail close to Period 20, which shows that my baseline results are not qualitatively

[21] This in turn generates a gap in estimates from relabeling Period 5 and Period 6 in Table 2.6, which can be eliminated by further dropping those who are single in Period 5 and married in Period 6. I also find the same to be true for all tables below with a gap of this nature.

[22] I worked on relaxing the constant r assumption by allowing it to vary over time in a linear fashion. I implemented the resulting empirical strategy but obtained very noisy estimates. Therefore I maintain the constant r assumption in this paper due to limited sample size and statistical power.

[23] I worked on including the duration of marriage as an additional time-varying signal s_t. Although the theory in this paper can be generalized to accommodate more than one time-varying signal, the resulting empirical strategy is data-demanding. I obtained very noisy estimates when I included both the marriage dummy and the duration of marriage as time-varying signals. Therefore I only include the marriage dummy as the time-varying signal in this paper.

affected by slight changes in the right-censoring rule.

Table 2.7 reports robustness check results for empirical strategy (A) with noncognitive ability as variable z, and Table 2.8 presents results for strategy (B). These tables show similar patterns as Table 2.5 and confirm the conclusions above. Moreover, the point estimates of r in these tables are similar in size to those in Table 2.5, which is reassuring.

In summary, the results in this subsection provide further evidence for employer learning. The treatment effect estimates are found to be sensitive to alternative sample periods. Nevertheless, most of these estimates are larger than the individual fixed effects estimate, confirming that individual fixed effects estimates are downward biased in the setting of this paper.

2.5.2 Controlling for Marital Status History

In my analysis so far I only include marital status in one period on the right-hand side of my regressions. As a robustness check, I implement empirical strategy (B) controlling for the entire marital status history up to the relabeled Period 0 for each subsample above.

For example, when I relabel Period 4 as Period 0, I drop observations before Period 4, and regress log wage w_t and marital status s_t respectively on marital status in Period 1 through Period 4 and the interactions of these past marital status variables with potential experience t.[24] In theory, the treatment effect is identified by the ratio of coefficients on any pair of corresponding interaction terms in the two regressions. To increase estimation precision, I restrict all four ratios to be the same when I jointly estimate the two regressions to obtain one single treatment effect estimate for each subsample.

[24] I also include a cubic in potential experience t and the same set of standard controls mentioned in Section 2.3.

The results are presented in Table 2.9. The point estimates are very similar to those in Table 2.8, especially for the baseline subsample with 4 to 20 years of potential experience. The standard errors in Table 2.9 are slightly smaller than those in Table 2.8, which shows that controlling for the marital status history marginally increases estimation precision. However, this improvement in precision comes with a cost of robustness. For example, when I include marital status in Period 1 through Period 4, I need an additional assumption that conditional expectations $E(z|s_1, s_2, s_3, s_4)$ and $E(\eta|s_1, s_2, s_3, s_4)$ are linear in the four s variables, which basically rules out any interactions among these variables. Without this assumption, I would have to include all possible interactions among these variables in the regressions, which is cumbersome.

In summary, I show in this subsection that controlling for marital status history does not largely affect the treatment effect estimates, and therefore I conclude that my baseline results are robust in this aspect.

2.5.3 Nonlinear Specifications

In my analysis so far I impose linearity when estimating the nuisance parameters (i.e. the derivatives of b_{st}, b_{zt}, Φ_t^{ss}, Φ_t^{sz}, b_{s_0t} and $\Phi_t^{ss_0}$ with respect to t) in Equations (2.10) – (2.13). In theory, the shapes of experience profiles of these parameters depend on the speed of employer learning over time as well as the marital status transition pattern and therefore are likely to be nonlinear. However, it is possible that both the identification of the treatment effect parameter and the empirical test of the employer learning model are preserved under linear approximations.[25] Next I examine the robustness of the results above with respect to the linearity assumption

[25] If the model is completely free from misspecifications, linear approximations do not affect the identification of the treatment effect parameter. This follows from the fact that if the ratio of two nonlinear functions is a constant, then their comparable linear approximations maintain the same ratio. The test for EL-SD depends solely on the *sign* of parameter $\frac{\partial \Phi_{zt}}{\partial t}$, which is likely to be preserved in a linear approximation.

by comparing them with results from nonlinear specifications presented below in Table 2.10 through Table 2.14.

Table 2.10 through Table 2.12 present results for each empirical strategy using logarithmic specifications. In these specifications, instead of linear interactions of potential experience t with s_0 and z, I use interactions of $\log(t+1)$ with s_0 and z. In this case, the derivative parameters listed above are equal to coefficients on the interaction terms divided by $t+1$. Clearly, this does not complicate empirical strategy (B), since the denominators on both sides of Equation (2.13) are canceled out. In empirical strategy (A), however, the logarithmic specification implies that the derivative parameters $\frac{\partial \Phi_{st}}{\partial t}$ and $\frac{\partial \Phi_{zt}}{\partial t}$ as solutions to Equations (2.10) – (2.12) are no longer constants, but rather some constants divided by $t+1$. Therefore, I treat constants $\frac{\partial \Phi_{st}}{\partial t} \times (t+1)$ and $\frac{\partial \Phi_{zt}}{\partial t} \times (t+1)$ as parameters to be estimated and report their estimates in Table 2.10 and Table 2.11, together with treatment effect estimates.

Additionally, I implement empirical strategy (B) using quadratic and cubic specifications and report the results in Table 2.13 and Table 2.14. In the quadratic specification, I include both interaction terms $s_0 \times t$ and $s_0 \times t^2$ in the regressions of w_t and s_t when estimating the two nuisance parameters in Equation 2.13. In this case, the treatment effect parameter r is identified by the ratio of coefficients on $s_0 \times t$ and also on $s_0 \times t^2$ in the two regressions. If the model is correctly specified, estimates of these two ratios should be similar in size, since they are estimates of the same parameter. However, this flexible quadratic specification produces very noisy estimates in my sample. Therefore I restrict the two ratios to be the same when I jointly estimate the two regressions to obtain only one treatment effect estimate for each subsample. I implement the cubic specification likewise.

Table 2.10 through Table 2.14 show that nonlinear specifications produce similar results as linear specifications, especially when restricting to 4 through 20 years of potential experience as in the baseline. Therefore, I conclude that the main results and conclusions in this paper are robust to linearity specifications.

2.6 Conclusions

In this article I develop the employer learning model with dynamic statistical discrimination on the basis of easily observable worker characteristics that are time-varying. In this model, employers observe and use the entire history of worker characteristics to infer productivity and thereby statistically discriminate among young workers. This paper adds to the existing employer learning literature pioneered by Farber and Gibbons (1996) and Altonji and Pierret (2001) where easily observable worker characteristics are time-invariant.

I find that the testable implications of the employer learning model in this paper lie in the regression of log wages on worker characteristics in the *initial* period and ability measures. I obtain the test statistic by solving a system of equations involving coefficients from this log wage regression and an auxiliary regression capturing the transition pattern of worker characteristics, together with a theory-implied relationship. I also show that AP's empirical test, which is designed for time-invariant worker characteristics, cannot be directly applied in the setting of this paper.

In this paper, worker characteristics affect wages not only by conveying information of worker productivity to the employer, but also by directly affecting the worker's productivity. I establish the identification of the treatment effect of worker characteristics on productivity under the assumptions of incomplete information and employer learning, and propose novel treatment effect estimators that are robust to the information structure of the labor market. Moreover, I show that individual fixed effects only identify the treatment effect in the complete information setting.

I empirically test for employer learning and dynamic statistical discrimination on the basis of marital status using a sample of white male high school graduates from NLSY79 and find support for this hypothesis. My baseline estimates of the treatment effect of marriage are around 17%, which are robust to the choice and absence of ability measures as well as the linearity specification. My estimates vary with different relabeling rules and potential experience ranges. This is possibly due to my simplifying assumption that the treatment effect of marriage is the same regardless of the worker's experience level or the duration of marriage. Nevertheless, my estimates are larger than individual fixed effects estimates in general, suggesting that fixed effects estimates are downward biased in this setting. One reasonable explanation is that the worker's tendency to prioritize work over personal life is a hard-to-observe productive trait negatively correlated with marriage over time.

Table 2.1: Model Implications and Treatment Effect Identification under EL-SD and Complete Information

	EL-SD	Complete information
(A) Regress w_t on s_0 and z	$\frac{\partial \Phi_{zt}}{\partial t} \geqslant 0$	$\frac{\partial \Phi_{zt}}{\partial t} = 0$
	$\frac{\partial \Phi_{st}}{\partial t} = -\Phi_{zs} \frac{\partial \Phi_{zt}}{\partial t}$ (2.10)	—
	$\frac{\partial b_{st}}{\partial t} = r \frac{\partial \Phi_t^{ss}}{\partial t} + \frac{\partial \Phi_{st}}{\partial t}$ (2.11)	$\frac{\partial b_{st}}{\partial t} = r \frac{\partial \Phi_t^{ss}}{\partial t}$
	$\frac{\partial b_{zt}}{\partial t} = r \frac{\partial \Phi_t^{sz}}{\partial t} + \frac{\partial \Phi_{zt}}{\partial t}$ (2.12)	$\frac{\partial b_{zt}}{\partial t} = r \frac{\partial \Phi_t^{sz}}{\partial t}$
	r, $\frac{\partial \Phi_{st}}{\partial t}$ and $\frac{\partial \Phi_{zt}}{\partial t}$ identified	r identified
(B) Regress w_t on s_0	$\frac{\partial \Phi_{s_0 t}}{\partial t} = 0$	$\frac{\partial \Phi_{s_0 t}}{\partial t} = 0$
	$\frac{\partial b_{s_0 t}}{\partial t} = r \frac{\partial \Phi_t^{ss_0}}{\partial t}$ (2.13)	$\frac{\partial b_{s_0 t}}{\partial t} = r \frac{\partial \Phi_t^{ss_0}}{\partial t}$ (2.13)
	r identified	r identified

Table 2.2: NLSY79 Summary Statistics by Marital Status

	Total		Unmarried		Married	
Observations; Individuals	11,408	1,101	5,307	961	6,101	795
Experience 1-5	2,907	943	2,190	826	717	347
Experience 6-9	2,829	970	1,362	601	1,467	567
Experience 10-14	2,737	754	943	348	1,794	533
Experience 15-20	1,780	612	506	231	1,274	459
Experience 21-33	1,155	492	306	160	849	364
Log of real wage	6.824	(0.454)	6.686	(0.447)	6.945	(0.425)
Experience 1-5	6.604	(0.407)	6.560	(0.401)	6.738	(0.398)
Experience 6-9	6.774	(0.422)	6.706	(0.442)	6.838	(0.393)
Experience 10-14	6.895	(0.425)	6.770	(0.442)	6.960	(0.401)
Experience 15-20	6.971	(0.443)	6.842	(0.469)	7.022	(0.422)
Experience 21-33	7.110	(0.444)	6.982	(0.470)	7.156	(0.426)
AFQT	0.261	(0.765)	0.221	(0.799)	0.297	(0.732)
Experience 1-5	0.245	(0.768)	0.246	(0.783)	0.239	(0.720)
Experience 6-9	0.267	(0.759)	0.248	(0.787)	0.285	(0.731)
Experience 10-14	0.268	(0.776)	0.176	(0.829)	0.316	(0.741)
Experience 15-20	0.254	(0.765)	0.158	(0.836)	0.293	(0.732)
Experience 21-33	0.287	(0.746)	0.163	(0.798)	0.332	(0.722)
Noncognitive ability	0.017	(0.747)	-0.026	(0.751)	0.055	(0.741)
Experience 1-5	-0.002	(0.734)	-0.008	(0.738)	0.018	(0.720)
Experience 6-9	0.031	(0.736)	-0.022	(0.735)	0.079	(0.735)
Experience 10-14	0.027	(0.750)	-0.062	(0.769)	0.075	(0.736)
Experience 15-20	0.013	(0.766)	-0.034	(0.786)	0.031	(0.758)
Experience 21-33	0.015	(0.768)	-0.048	(0.799)	0.038	(0.756)
Part-time job (%)	3.29		5.50		1.36	
Urban residence (%)	66.91		72.24		62.27	
Region (%)						
Northeast	19.84		21.05		18.78	
North Central	39.28		38.18		40.24	
South	27.43		25.57		29.04	
West	13.46		15.21		11.93	

Notes: Individual-by-year observations are drawn from NLSY79 for survey years 1979-2004. Experience is measured as the number of years since an individual left school for the first time. Standard deviations are in parentheses.

Table 2.3: Treatment Effect Estimates and Empirical Tests of EL-SD

	(A) Regress w_t on s_4 and z		(B) Regress w_t on s_4	Individual fixed effects	OLS
r	0.1672	0.1578	0.1685	0.0461	0.1660
	(0.0413)	(0.0421)	(0.0420)	(0.0134)	(0.0188)
$\frac{\partial \Phi_{zt}}{\partial t} \times 10^2$	0.2344	0.2835			
	(0.1269)	(0.1291)			
Measure of z	AFQT	Noncognitive ability			

Notes: I relabel Period 4 as Period 0 and keep observations with 4 to 20 years of potential experience. The resulting sample contains 8719 person-year observations from 1069 individuals. Detailed estimation procedures for Column (A) and (B) are described in Section 2.3. In the last two columns, I estimate individual fixed effects and OLS regressions of log wage on marital status, controlling for a cubic in experience, urban residence, region of residence, and a dummy for part-time versus full-time jobs. I report coefficients on marital status in these two columns together with White/Huber standard errors which control for correlation at the individual level.

Table 2.4: Sample Sizes for Censored Subsamples

	$t \geqslant 1$	2	3	4	5	6
$t \leqslant$ max						
	11,398	10,979	10,461	9,872	9,206	8,496
	1,096	1,093	1,088	1,075	1,064	1,038
26						
	11,258	10,839	10,321	9,732	9,066	8,356
	1,096	1,093	1,088	1,075	1,064	1,038
24						
	11,013	10,594	10,077	9,488	8,822	8,112
	1,095	1,092	1,087	1,074	1,063	1,037
22						
	10,674	10,255	9,738	9,149	8,483	7,774
	1,094	1,091	1,086	1,073	1,062	1,036
20						
	10,243	9,824	9,308	8,719	8,053	7,344
	1,090	1,087	1,082	1,069	1,057	1,030
18						
	9,775	9,356	8,840	8,251	7,585	6,877
	1,088	1,085	1,080	1,067	1,055	1,028
16						
	9,204	8,785	8,270	7,681	7,016	6,306
	1,086	1,083	1,078	1,063	1,050	1,022

Notes: Each row or column imposes a different censoring rule on potential experience t. Within each cell, the top number counts person-year observations and the bottom number counts unique individuals.

Table 2.5: Estimates from Empirical Strategy (A), Linear Specification, AFQT as Variable z

	s_1	s_2	s_3	s_4	s_5	s_6
$t \leqslant$ max						
r	0.4120	0.2193	0.1855	0.2407	0.1427	0.1059
	(0.0896)	(0.0588)	(0.0454)	(0.0395)	(0.0427)	(0.0403)
$\frac{\partial \Phi_{zt}}{\partial t} \times 100$	0.2832	0.3309	0.2997	0.2623	0.2885	0.2422
	(0.0843)	(0.0830)	(0.0857)	(0.0911)	(0.0969)	(0.1045)
26						
	0.3442	0.1687	0.1629	0.2123	0.1327	0.0872
	(0.0843)	(0.0578)	(0.0444)	(0.0392)	(0.0421)	(0.0401)
	0.3042	0.3454	0.2986	0.2587	0.2790	0.2268
	(0.0845)	(0.0851)	(0.0885)	(0.0942)	(0.1007)	(0.1090)
24						
	0.2771	0.1450	0.1458	0.1888	0.1218	0.0717
	(0.0801)	(0.0573)	(0.0442)	(0.0394)	(0.0419)	(0.0402)
	0.3024	0.3297	0.2724	0.2249	0.2375	0.1738
	(0.0884)	(0.0907)	(0.0946)	(0.1010)	(0.1087)	(0.1183)
22						
	0.2221	0.1213	0.1277	0.1639	0.0793	0.0334
	(0.0793)	(0.0576)	(0.0442)	(0.0399)	(0.0427)	(0.0415)
	0.3380	0.3532	0.2869	0.2356	0.2543	0.1786
	(0.0947)	(0.0982)	(0.1029)	(0.1105)	(0.1199)	(0.1316)
20						
	0.1689	0.1066	0.1163	0.1672	0.0822	0.0356
	(0.0781)	(0.0574)	(0.0455)	(0.0413)	(0.0435)	(0.0428)
	0.3800	0.3793	0.3008	0.2344	0.2595	0.1701
	(0.1057)	(0.1109)	(0.1172)	(0.1269)	(0.1392)	(0.1546)
18						
	0.1756	0.0961	0.1371	0.1823	0.0927	0.0462
	(0.0785)	(0.0582)	(0.0462)	(0.0424)	(0.0439)	(0.0439)
	0.5466	0.5729	0.4839	0.4383	0.5115	0.4595
	(0.1186)	(0.1254)	(0.1335)	(0.1462)	(0.1622)	(0.1822)
16						
	0.2097	0.0929	0.1503	0.1953	0.0914	0.0585
	(0.0792)	(0.0577)	(0.0470)	(0.0441)	(0.0451)	(0.0458)
	0.6024	0.6628	0.5568	0.5179	0.6389	0.5944
	(0.1349)	(0.1433)	(0.1545)	(0.1714)	(0.1929)	(0.2205)

Notes: In each column, I relabel a different period as Period 0 and drop observations prior to that period. I then use marital status in that period as the empirical counterpart of s_0. In each row group, I impose a different right-censoring rule on potential experience t and drop later periods. Sample sizes are reported in Table 2.4. For each subsample, I implement empirical strategy (A) using AFQT as variable z. The estimation procedure is described in Section 2.3. Standard errors are in parentheses.

Table 2.6: Estimates from Empirical Strategy (A), Linear Specification, AFQT as Variable z, Excluding Individuals Transitioning into Marriage from Period 4 to Period 5

	s_1	s_2	s_3	s_4	s_5	s_6
$t \leqslant$ max						
r	0.4882	0.2237	0.1775	0.2043	0.2294	0.1543
	(0.0966)	(0.0593)	(0.0454)	(0.0377)	(0.0456)	(0.0420)
$\frac{\partial \Phi_{zt}}{\partial t} \times 100$	0.2984	0.3802	0.3539	0.3100	0.3077	0.2816
	(0.0906)	(0.0868)	(0.0894)	(0.0947)	(0.1021)	(0.1093)
26						
	0.3801	0.1710	0.1492	0.1826	0.2052	0.1235
	(0.0891)	(0.0582)	(0.0443)	(0.0374)	(0.0452)	(0.0420)
	0.3328	0.3949	0.3549	0.3030	0.2997	0.2663
	(0.0891)	(0.0888)	(0.0922)	(0.0979)	(0.1057)	(0.1138)
24						
	0.2637	0.1277	0.1210	0.1628	0.1841	0.0982
	(0.0822)	(0.0573)	(0.0439)	(0.0376)	(0.0450)	(0.0421)
	0.3294	0.3704	0.3154	0.2460	0.2339	0.1879
	(0.0920)	(0.0945)	(0.0985)	(0.1049)	(0.1137)	(0.1233)
22						
	0.2033	0.1007	0.0979	0.1334	0.1476	0.0672
	(0.0816)	(0.0578)	(0.0442)	(0.0383)	(0.0458)	(0.0433)
	0.3641	0.3933	0.3309	0.2539	0.2424	0.1853
	(0.0984)	(0.1023)	(0.1072)	(0.1148)	(0.1251)	(0.1370)
20						
	0.1478	0.0866	0.0907	0.1340	0.1569	0.0752
	(0.0808)	(0.0580)	(0.0456)	(0.0397)	(0.0470)	(0.0446)
	0.4014	0.4147	0.3386	0.2460	0.2275	0.1580
	(0.1098)	(0.1154)	(0.1220)	(0.1318)	(0.1453)	(0.1607)
18						
	0.1492	0.0728	0.1119	0.1477	0.1712	0.0895
	(0.0814)	(0.0588)	(0.0462)	(0.0406)	(0.0482)	(0.0460)
	0.5748	0.6177	0.5298	0.4559	0.4691	0.4408
	(0.1231)	(0.1306)	(0.1389)	(0.1517)	(0.1696)	(0.1898)
16						
	0.1897	0.0731	0.1227	0.1553	0.1864	0.1167
	(0.0813)	(0.0581)	(0.0470)	(0.0423)	(0.0500)	(0.0478)
	0.6353	0.7166	0.6186	0.5485	0.5771	0.5649
	(0.1400)	(0.1493)	(0.1607)	(0.1778)	(0.2023)	(0.2301)

Notes: See notes in Table 2.5. I repeat the exact analysis in Table 2.5 dropping individuals who are single in Period 4 and married in Period 5.

Table 2.7: Estimates from Empirical Strategy (A), Linear Specification, Noncognitive Ability as Variable z

	s_1	s_2	s_3	s_4	s_5	s_6
$t \leqslant$ max						
r	0.4256	0.2587	0.1992	0.2272	0.1393	0.1016
	(0.0907)	(0.0588)	(0.0458)	(0.0403)	(0.0433)	(0.0410)
$\frac{\partial \Phi_{zt}}{\partial t} \times 100$	0.0932	0.1101	0.1036	0.1209	0.0486	-0.0094
	(0.0804)	(0.0810)	(0.0847)	(0.0902)	(0.0958)	(0.1038)
26						
	0.3572	0.2057	0.1790	0.2028	0.1303	0.0820
	(0.0854)	(0.0578)	(0.0448)	(0.0399)	(0.0427)	(0.0408)
	0.1199	0.1380	0.1302	0.1505	0.0788	0.0205
	(0.0829)	(0.0847)	(0.0892)	(0.0951)	(0.1016)	(0.1106)
24						
	0.2910	0.1783	0.1607	0.1790	0.1173	0.0647
	(0.0808)	(0.0573)	(0.0445)	(0.0401)	(0.0425)	(0.0409)
	0.1511	0.1665	0.1578	0.1824	0.1075	0.0484
	(0.0874)	(0.0907)	(0.0959)	(0.1026)	(0.1103)	(0.1208)
22						
	0.2371	0.1490	0.1384	0.1538	0.0728	0.0252
	(0.0798)	(0.0577)	(0.0446)	(0.0407)	(0.0434)	(0.0422)
	0.2426	0.2618	0.2592	0.2955	0.2235	0.1709
	(0.0946)	(0.0991)	(0.1054)	(0.1133)	(0.1228)	(0.1358)
20						
	0.1843	0.1322	0.1257	0.1578	0.0775	0.0294
	(0.0785)	(0.0575)	(0.0460)	(0.0421)	(0.0442)	(0.0435)
	0.2385	0.2504	0.2451	0.2835	0.1989	0.1374
	(0.1054)	(0.1113)	(0.1192)	(0.1291)	(0.1412)	(0.1578)
18						
	0.1893	0.1225	0.1464	0.1744	0.0908	0.0454
	(0.0790)	(0.0583)	(0.0467)	(0.0432)	(0.0446)	(0.0445)
	0.3083	0.3345	0.3307	0.3886	0.3021	0.2588
	(0.1179)	(0.1252)	(0.1350)	(0.1477)	(0.1630)	(0.1840)
16						
	0.2202	0.1162	0.1609	0.1914	0.0916	0.0608
	(0.0798)	(0.0578)	(0.0475)	(0.0448)	(0.0457)	(0.0464)
	0.3235	0.3806	0.3680	0.4407	0.3511	0.3143
	(0.1355)	(0.1438)	(0.1563)	(0.1729)	(0.1932)	(0.2215)

Notes: See notes in Table 2.5. I repeat the exact analysis in Table 2.5 using noncognitive ability as variable z.

Table 2.8: Treatment Effect of Marriage Estimates from Empirical Strategy (B), Linear Specification

	s_1	s_2	s_3	s_4	s_5	s_6
$t \leqslant$ max						
	0.4504	0.2732	0.2113	0.2421	0.1434	0.0978
	(0.0897)	(0.0586)	(0.0457)	(0.0402)	(0.0435)	(0.0413)
26						
	0.3768	0.2176	0.1869	0.2139	0.1339	0.0779
	(0.0847)	(0.0576)	(0.0448)	(0.0399)	(0.0429)	(0.0412)
24						
	0.3059	0.1912	0.1678	0.1888	0.1224	0.0626
	(0.0803)	(0.0571)	(0.0445)	(0.0401)	(0.0427)	(0.0412)
22						
	0.2528	0.1627	0.1471	0.1636	0.0794	0.0244
	(0.0794)	(0.0575)	(0.0446)	(0.0407)	(0.0435)	(0.0425)
20						
	0.1975	0.1448	0.1352	0.1685	0.0853	0.029
	(0.0781)	(0.0573)	(0.0459)	(0.0420)	(0.0443)	(0.0438)
18						
	0.2009	0.1369	0.158	0.1864	0.0985	0.0456
	(0.0788)	(0.0580)	(0.0466)	(0.0431)	(0.0446)	(0.0448)
16						
	0.2289	0.1292	0.1702	0.1997	0.0972	0.0578
	(0.0796)	(0.0575)	(0.0474)	(0.0448)	(0.0458)	(0.0467)

Notes: In each column, I relabel a different period as Period 0 and drop observations prior to that period. I then use marital status in that period as the empirical counterpart of s_0. In each row, I impose a different right-censoring rule on potential experience t and drop later periods. Sample sizes are reported in Table 2.4. For each subsample, I implement empirical strategy (B). The estimation procedure is described in Section 2.3. Standard errors are in parentheses.

Table 2.9: Treatment Effect of Marriage Estimates from Empirical Strategy (B), Linear Specification, Controlling for Marital Status History

	s_1	s_1, s_2	$s_1, \cdots, s_3$	$s_1, \cdots, s_4$	$s_1, \cdots, s_5$	$s_1, \cdots, s_6$
$t \leqslant \infty$	0.4504	0.2647	0.2053	0.2382	0.1752	0.1138
	(0.0897)	(0.0585)	(0.0457)	(0.0400)	(0.0430)	(0.0405)
26	0.3768	0.2138	0.186	0.2131	0.1596	0.0901
	(0.0847)	(0.0578)	(0.0449)	(0.0399)	(0.0427)	(0.0406)
24	0.3059	0.1912	0.1692	0.1897	0.1422	0.0709
	(0.0803)	(0.0573)	(0.0447)	(0.0402)	(0.0426)	(0.0408)
22	0.2528	0.1625	0.1488	0.1646	0.0982	0.0343
	(0.0794)	(0.0575)	(0.0447)	(0.0407)	(0.0434)	(0.0421)
20	0.1975	0.146	0.138	0.1715	0.1005	0.038
	(0.0781)	(0.0572)	(0.0459)	(0.0420)	(0.0442)	(0.0434)
18	0.2009	0.1375	0.162	0.1888	0.1096	0.0503
	(0.0788)	(0.0578)	(0.0465)	(0.0431)	(0.0447)	(0.0446)
16	0.2289	0.1258	0.173	0.201	0.1035	0.0591
	(0.0796)	(0.0572)	(0.0473)	(0.0447)	(0.0459)	(0.0465)

Notes: In each column, I relabel a different period as Period 0 and drop observations prior to that period. In each row, I impose a different right-censoring rule on potential experience t and drop later periods. Sample sizes are slightly smaller than those reported in Table 2.4 due to (less than 30) incomplete cases in the history of marital status. For each subsample, I implement empirical strategy (B) controlling for the entire marital status history up to the relabeled Period 0. The estimation procedure is described in Section 2.5.2. Standard errors are in parentheses.

Table 2.10: Estimates from Empirical Strategy (A), Log Specification, AFQT as Variable z

	s_1	s_2	s_3	s_4	s_5	s_6
$t \leqslant \max$						
r	0.3135	0.1669	0.1594	0.2137	0.1236	0.0879
	(0.0720)	(0.0495)	(0.0395)	(0.0357)	(0.0381)	(0.0369)
$\frac{\partial \Phi_{zt}}{\partial t} \times 10(t+1)$	0.3559	0.4202	0.3878	0.3649	0.4371	0.3885
	(0.0807)	(0.0893)	(0.1007)	(0.1150)	(0.1306)	(0.1494)
26						
	0.2755	0.1362	0.1445	0.1941	0.1167	0.0743
	(0.0705)	(0.0494)	(0.0393)	(0.0358)	(0.0380)	(0.0370)
	0.3633	0.4252	0.3826	0.3582	0.4240	0.3690
	(0.0807)	(0.0903)	(0.1021)	(0.1168)	(0.1332)	(0.1529)
24						
	0.2370	0.1208	0.1329	0.1773	0.1087	0.0628
	(0.0693)	(0.0497)	(0.0395)	(0.0362)	(0.0382)	(0.0373)
	0.3544	0.4063	0.3543	0.3222	0.3778	0.3086
	(0.0821)	(0.0929)	(0.1055)	(0.1212)	(0.1392)	(0.1607)
22						
	0.2036	0.1041	0.1196	0.1589	0.0775	0.0337
	(0.0697)	(0.0504)	(0.0400)	(0.0371)	(0.0392)	(0.0387)
	0.3686	0.4155	0.3596	0.3274	0.3918	0.3151
	(0.0846)	(0.0966)	(0.1104)	(0.1276)	(0.1478)	(0.1723)
20						
	0.1711	0.0932	0.1104	0.1602	0.0794	0.0356
	(0.0700)	(0.0510)	(0.0414)	(0.0386)	(0.0403)	(0.0402)
	0.3816	0.4228	0.3617	0.3216	0.3953	0.3115
	(0.0889)	(0.1028)	(0.1187)	(0.1386)	(0.1625)	(0.1918)
18						
	0.1746	0.0851	0.1231	0.1696	0.0865	0.0433
	(0.0713)	(0.0520)	(0.0424)	(0.0399)	(0.0412)	(0.0416)
	0.4607	0.5331	0.4838	0.4795	0.6157	0.5956
	(0.0935)	(0.1093)	(0.1275)	(0.1508)	(0.1790)	(0.2142)
16						
	0.1950	0.0817	0.1302	0.1768	0.0842	0.0519
	(0.0730)	(0.0525)	(0.0435)	(0.0418)	(0.0426)	(0.0437)
	0.4702	0.5642	0.5135	0.5232	0.7062	0.7084
	(0.0992)	(0.1172)	(0.1385)	(0.1663)	(0.2006)	(0.2446)

Notes: See notes in Table 2.5. I repeat the analysis in Table 2.5 using logarithmic specifications described in Section 2.5.3.

Table 2.11: Estimates from Empirical Strategy (A), Log Specification, Noncognitive Ability as Variable z

	s_1	s_2	s_3	s_4	s_5	s_6
$t \leqslant \max$						
r	0.3247	0.1952	0.1702	0.2028	0.1206	0.0843
	(0.0729)	(0.0497)	(0.0400)	(0.0364)	(0.0387)	(0.0375)
$\frac{\partial \Phi_{zt}}{\partial t} \times 10(t+1)$	0.1277	0.1663	0.1717	0.2222	0.1216	0.0455
	(0.0808)	(0.0900)	(0.1018)	(0.1157)	(0.1307)	(0.1499)
26						
	0.2865	0.1629	0.1568	0.1858	0.1143	0.0701
	(0.0713)	(0.0496)	(0.0397)	(0.0364)	(0.0386)	(0.0376)
	0.1469	0.1891	0.1939	0.2502	0.1538	0.0830
	(0.0819)	(0.0919)	(0.1043)	(0.1188)	(0.1349)	(0.1555)
24						
	0.2485	0.1453	0.1444	0.1690	0.1049	0.0572
	(0.0700)	(0.0499)	(0.0399)	(0.0369)	(0.0388)	(0.0379)
	0.1668	0.2097	0.2163	0.2801	0.1850	0.1194
	(0.0837)	(0.0949)	(0.1082)	(0.1238)	(0.1414)	(0.1640)
22						
	0.2154	0.1252	0.1282	0.1504	0.0724	0.0274
	(0.0702)	(0.0506)	(0.0404)	(0.0377)	(0.0398)	(0.0393)
	0.2198	0.2737	0.2939	0.3779	0.2977	0.2518
	(0.0868)	(0.0992)	(0.1138)	(0.1310)	(0.1509)	(0.1768)
20						
	0.1830	0.1128	0.1183	0.1524	0.0757	0.0311
	(0.0705)	(0.0512)	(0.0418)	(0.0392)	(0.0409)	(0.0409)
	0.2088	0.2564	0.2730	0.3576	0.2657	0.2132
	(0.0913)	(0.1054)	(0.1220)	(0.1416)	(0.1648)	(0.1952)
18						
	0.1856	0.1048	0.1309	0.1632	0.0852	0.0434
	(0.0719)	(0.0523)	(0.0429)	(0.0406)	(0.0418)	(0.0422)
	0.2370	0.3015	0.3251	0.4340	0.3497	0.3279
	(0.0963)	(0.1121)	(0.1309)	(0.1536)	(0.1807)	(0.2165)
16						
	0.2042	0.0989	0.1386	0.1733	0.0845	0.0546
	(0.0736)	(0.0529)	(0.0440)	(0.0424)	(0.0432)	(0.0443)
	0.2309	0.3176	0.3387	0.4635	0.3812	0.3734
	(0.1032)	(0.1208)	(0.1426)	(0.1696)	(0.2023)	(0.2464)

Notes: See notes in Table 2.7. I repeat the analysis in Table 2.7 using logarithmic specifications described in Section 2.5.3.

Table 2.12: Treatment Effect of Marriage Estimates from Empirical Strategy (B), Log Specification

	s_1	s_2	s_3	s_4	s_5	s_6
$t \leqslant$ max						
	0.3408	0.2067	0.1798	0.2149	0.1246	0.0807
	(0.0723)	(0.0496)	(0.0400)	(0.0363)	(0.0389)	(0.0378)
26						
	0.2993	0.1727	0.1638	0.1954	0.1179	0.0664
	(0.0709)	(0.0496)	(0.0397)	(0.0364)	(0.0387)	(0.0379)
24						
	0.2585	0.1556	0.1509	0.1776	0.1097	0.0551
	(0.0697)	(0.0497)	(0.0399)	(0.0369)	(0.0389)	(0.0382)
22						
	0.2256	0.1358	0.1358	0.159	0.0783	0.0263
	(0.0701)	(0.0505)	(0.0404)	(0.0377)	(0.0399)	(0.0396)
20						
	0.1914	0.1226	0.1262	0.1614	0.0823	0.0303
	(0.0703)	(0.0511)	(0.0418)	(0.0392)	(0.0410)	(0.0411)
18						
	0.1928	0.1152	0.1399	0.1729	0.0915	0.0429
	(0.0718)	(0.0522)	(0.0428)	(0.0406)	(0.0419)	(0.0425)
16						
	0.2092	0.1081	0.1459	0.1803	0.0892	0.0516
	(0.0735)	(0.0527)	(0.0439)	(0.0424)	(0.0433)	(0.0446)

Notes: See notes in Table 2.8. I repeat the analysis in Table 2.8 using logarithmic specifications described in Section 2.5.3.

Table 2.13: Treatment Effect of Marriage Estimates from Empirical Strategy (B), Quadratic Specification

	s_1	s_2	s_3	s_4	s_5	s_6
$t \leqslant$ max						
	0.261	0.1579	0.1585	0.1934	0.1017	0.0561
	(0.0698)	(0.0490)	(0.0391)	(0.0358)	(0.0374)	(0.0366)
26						
	0.2412	0.1429	0.1517	0.183	0.0971	0.0488
	(0.0693)	(0.0490)	(0.0391)	(0.0359)	(0.0373)	(0.0367)
24						
	0.2253	0.1313	0.1447	0.1724	0.0913	0.0424
	(0.0692)	(0.0492)	(0.0395)	(0.0364)	(0.0376)	(0.0372)
22						
	0.2066	0.1166	0.1363	0.162	0.0788	0.0299
	(0.0696)	(0.0497)	(0.0400)	(0.0373)	(0.0384)	(0.0383)
20						
	0.1921	0.1058	0.1295	0.1616	0.0797	0.0333
	(0.0702)	(0.0503)	(0.0410)	(0.0387)	(0.0395)	(0.0397)
18						
	0.1915	0.0946	0.1292	0.1641	0.0806	0.0374
	(0.0717)	(0.0510)	(0.0419)	(0.0400)	(0.0405)	(0.0410)
16						
	0.1951	0.0838	0.1207	0.1589	0.0725	0.0338
	(0.0735)	(0.0518)	(0.0429)	(0.0417)	(0.0417)	(0.0428)

Notes: See notes in Table 2.8. I repeat the analysis in Table 2.8 using quadratic specifications described in Section 2.5.3.

Table 2.14: Treatment Effect of Marriage Estimates from Empirical Strategy (B), Cubic Specification

	s_1	s_2	s_3	s_4	s_5	s_6
$t \leqslant$ max						
	0.2685	0.1546	0.1535	0.1926	0.0991	0.06
	(0.0685)	(0.0471)	(0.0380)	(0.0351)	(0.0363)	(0.0356)
26						
	0.2515	0.1325	0.1438	0.1793	0.0972	0.0535
	(0.0683)	(0.0472)	(0.0379)	(0.0352)	(0.0363)	(0.0357)
24						
	0.232	0.12	0.1319	0.1642	0.0934	0.0481
	(0.0681)	(0.0475)	(0.0382)	(0.0357)	(0.0365)	(0.0361)
22						
	0.2106	0.1068	0.1178	0.1474	0.0721	0.027
	(0.0687)	(0.0483)	(0.0388)	(0.0366)	(0.0375)	(0.0374)
20						
	0.1864	0.0984	0.1067	0.1456	0.0725	0.0276
	(0.0694)	(0.0491)	(0.0401)	(0.0380)	(0.0386)	(0.0389)
18						
	0.1874	0.0929	0.1147	0.1521	0.0778	0.0362
	(0.0710)	(0.0503)	(0.0412)	(0.0394)	(0.0397)	(0.0404)
16						
	0.2012	0.0883	0.1175	0.1577	0.0753	0.0378
	(0.0730)	(0.0514)	(0.0425)	(0.0413)	(0.0414)	(0.0427)

Notes: See notes in Table 2.8. I repeat the analysis in Table 2.8 using cubic specifications described in Section 2.5.3.

3

Understanding Workplace Flexibility: Worker and Occupation Characteristics

3.1 Introduction

Workplace flexibility is a desirable job amenity costly for employers to provide. Recent economic research on workplace flexibility highlights two important themes. First, different types of workers favor job flexibility to various degrees. Wiswall and Zafar (2016) show that women on average have a higher willingness to pay (WTP) for jobs with flexible work hours relative to men. Edwards (2014) finds that a woman's WTP for work schedule flexibility grows as her number of children increases or if she has an infant. Flabbi and Moro (2012) show that college-educated women value flexibility more than women with only a high school degree. These findings imply that differential preferences for workplace flexibility can explain a fraction of the gender wage gap.

Second, the cost of providing workplace flexibility differs by occupation. Drawing on insights from personnel economics, Goldin (2014) points out that providing flexibility is especially costly in occupations where an employee does not have a perfect substitute. She observes that if workers are perfect substitutes for each other, there will not be a wage premium for working long hours or particular hours, and earnings would be linear with respect to hours (i.e. the hourly wage is constant). On the other hand, if workers cannot easily substitute for each other, there will be a productivity loss and wage penalty from working low hours and conversely a wage premium for working long hours, in which case the earnings curve would be convex (i.e. the hourly wage increases with hours). Therefore, she estimates for each occupation the elasticity of annual earnings with respect to weekly hours worked, which quantifies the convexity of the earnings curve. The estimated elasticity serves as a measure of the cost of providing workplace flexibility in a particular occupation, which clears the way for further research on the varying cost of flexibility across occupations.

As Blau and Kahn (2016) point out, a high wage penalty for temporal flexibility in certain occupations could also be explained in the context of a human capital model, attributable to the particular importance of human capital accumulation in these occupations. Therefore, further research is needed to incorporate both the human capital theory and the personnel economics perspective in explaining the differential wage penalty for flexibility across occupations.

Along these lines, this study makes two main contributions to the above literature. First, I document the occupational choice patterns of various types of workers with respect to workplace flexibility. My findings provide further evidence that the tendency of choosing flexible occupations varies by worker characteristics. Using an ACS sample of college graduates working full-time, I find that women — especially those with children — tend to work in flexible occupations, and that married individuals and advanced degree holders regardless of gender are more likely to be found in flexible occupations. These findings confirm and complement findings in the previous literature.[1]

[1] I should note that I do not estimate the deep preference parameter or the willingness to pay for

Second, I explore the relationship between the cost of providing flexibility and occupation characteristics, which sheds light on the hindrance to workplace flexibility in certain occupations. I measure the cost of flexibility in each occupation using the elasticity of annual earnings with respect to weekly hours worked, following Goldin (2014). I use detailed occupation characteristics information in the O*NET database to quantify the required skill levels and the typical work context in each occupation. Including required skill levels as occupation characteristics allows me to incorporate the human capital theory in explaining the differential cost of flexibility across occupations.

To reduce the dimensionality of the O*NET data, I use Principal Component Analysis (PCA) to extract the top 3 principal components of skills, which I interpret as motor, social and cognitive skills respectively. I apply the same method to work context and obtain the top 3 principal components which I interpret as white-collar, pink-collar and blue-collar work settings respectively.

My results show that occupations requiring high motor skills as well as those performed in a blue-collar work setting are less flexible, and so are occupations in a white-collar office environment when controlling for skill levels. On the contrary, occupations involving high cognitive skills tend to be flexible.[2] These results highlight the explanatory power of the human capital theory in addition to personnel economics in understanding the varying cost of flexibility across occupations.

The rest of this chapter is organized as follows. Section 3.2 describes the ACS sample used in this study and estimates the elasticity of earnings with respect to weekly hours worked for each occupation. Section 3.3 documents the occupational choices of various types of workers regarding workplace flexibility. Section 3.4 intro-

---

workplace flexibility in this study. I only intend to provide descriptive results about various types of workers' occupational choice patterns regarding workplace flexibility.

[2] To write concisely, I sometimes refer to an occupation with a low cost of providing flexibility simply as a flexible occupation.

duces the O*NET data and performs PCA on two sets of occupation characteristics — i.e. required skill levels and the work context. Section 3.5 explains the variation in occupation flexibility using indices of occupation characteristics obtained from PCA. Section 3.6 concludes and provides directions for future research.

3.2 Occupational Flexibility and Nonlinear Pay

3.2.1 ACS Data and Sample Selection

I use data from the 2009-2011 American Community Survey (ACS)[3]. I restrict the sample to individuals between 25 and 64 years of age with a college degree and are working full-time full-year in a civilian job. I define full-time full-year workers as those working at least 35 hours a week and 40 weeks a year. I drop observations with less than $4,200 annual earnings, where annual earnings is defined as wages or salary income plus self-employment income in the past 12 months. The final sample contains 1,023,677 observations.

I convert the occupation codes in the 2010 and 2011 ACS samples to the 2002 Census Occupation Codes used in the 2009 ACS sample. After dropping observations with military occupations, missing occupations, and the unemployed, I have 465 unique occupations identified by three-digit Census occupation codes.

3.2.2 Elasticity of Earnings by Occupation

In this subsection I estimate the elasticity of earnings with respect to weekly hours worked for each occupation using the ACS sample described above.

I estimate the following log earnings equation:

$$\ln(Earnings_i) = \beta_0 + \beta_1 \ln(Hours_i) + \beta_2' \ln(Hours_i) \times OCCP_i + \beta_3' OCCP_i + \beta_4' \Phi_i + \epsilon_i, \tag{3.1}$$

[3] I combine three 1-year Public Use Microdata Samples (PUMS) obtained from the United States Census Bureau website.

where $OCCP_i$ denotes three-digit occupation dummies,[4] and Φ_i denotes other controls such as gender, race, age as a quartic, log weeks worked, education levels, ACS year, and interactions between gender and occupation dummies.

In (3.1), β_1 represents the elasticity of earnings in the baseline occupation, and adding β_1 to each element of β_2 gives the elasticity in each of the other occupations. A larger elasticity means that workers in that occupation face a greater earnings loss for working fewer hours, which translates to less flexibility in the occupation. Throughout the rest of this study, I use the estimated elasticity of earnings as a measure of occupation flexibility.

3.3 Occupational Flexibility and Worker Characteristics

In this section, I conduct descriptive analysis of occupation flexibility and worker characteristics such as gender, marital status, number of children and education levels. This sheds light on the categories of workers who are more likely to choose flexible occupations.

Throughout this section, I use a slightly smaller sample obtained by dropping elasticity outliers in the original sample. I define elasticity outliers as those falling outside the 5th and 95th percentile in the empirical distribution of the estimated elasticity. This is to prevent the outliers from distorting the regression analysis later in this section and also to acknowledge that the elasticity outliers are less precisely estimated in the first place. The number of observations is 1,018,090.

Table 3.1 presents the summary statistics of elasticity by gender. I calculate the standardized elasticity to have mean 0 and standard deviation 1. I also present the summary statistics of weekly hours worked by gender to gain more insights from this sample.

Table 3.1 shows that among full-time workers with a college degree, women are

[4] I leave out the post-secondary teachers occupation as the baseline occupation.

more likely to work in flexible occupations characterized by a smaller elasticity of earnings relative to men. Also, women on average work fewer hours per week than men in this sample, which is not surprising.

Next I perform regression analysis of elasticity of earnings separately for each gender controlling for a rich set of worker characteristics such as marital status, motherhood, ages of children, education levels, etc. The results are presented in Table 3.2. We see that married individuals regardless of gender tend to work in more flexible occupations, and so do women with children.

Also, it first appears that the presence of an older child (aged 6-17) affects women's occupation choice more than that of a younger one (aged less than 6). However, since women's ages are controlled for in these regressions, the child's age variables also capture women's different ages at first birth, which complicates the interpretation. Therefore, one should not read these results purely as the effect of child's age on women's occupational choice.

Moreover, advanced degree holders regardless of gender tend to work in more flexible occupations compared to those with only a college degree. One possible explanation is that advanced degree holders value flexibility more strongly, which is plausible given the finding of Flabbi and Moro (2012) that college-educated women value flexibility more than women with only a high school degree. Another possible explanation is that specialized occupations employing advanced degree holders tend to be more flexible. This hypothesis is supported by my findings presented later in Section 3.5.

Lastly, I present results from regressions of weekly hours worked on worker characteristics by gender in Table 3.3, which confirms the intuition that married women with children work less than their unmarried or childless female counterparts, whereas married men work more than unmarried men. Also, even among full-time workers, advance degree holders work significantly more hours per week than individuals with a college degree only.

3.4 O*NET Occupational Characteristics

In this section, I first describe the O*NET dataset which contains detailed information on occupational characteristics, and then perform principal component analysis (PCA) on O*NET data to summarize the information.

*3.4.1 O*NET Data*

The Occupational Information Network (O*NET) database[5] provides rich occupational information and serves as the successor to the Dictionary of Occupational Titles (DOT).[6] The latest version at the time of writing is O*NET 20.3,[7] which adopts the 2010 O*NET-SOC taxonomy containing 974 occupations.[8] Each occupation is characterized by hundreds of descriptors in several broad domains, describing the daily activities of the job and the requirements of the typical worker. The database is continually updated by surveying each occupation's worker population as well as occupation experts.

I select two sets of occupation characteristics most relevant to my research objectives: skills, and work context. The skills data file contains level and importance ratings of 35 skills for each of the 954 occupations,[9] and the work context data provides relevance ratings of 57 aspects of the job. The complete list of skills and work context items can be found in Table B.1 and Table B.2 in Appendix B.1.

---

[5] More information about the dataset can be found at https://www.onetcenter.org/overview.html.

[6] Both DOT and O*NET data are frequently used in the task literature. See Acemoglu and Autor (2011).

[7] The latest database releases are available at https://www.onetcenter.org/developers.html.

[8] The complete list of occupations can be found at https://www.onetcenter.org/taxonomy/2010/data_coll.html.

[9] Not all 974 occupations in the 2010 O*NET-SOC taxonomy are listed in the skills and work context data files.

3.4.2 Principal Component Analysis

In this subsection, I perform principal component analysis (PCA) separately on skills and work context data.[10] I convert each set of occupational characteristics into three principal components, which contain 78% (= 44% + 28% + 6%) of the variation in the skills data and 64% (= 44% + 12% + 8%) of the variation in the work context data. Although keeping more principal components would retain more variations in the original data, it increases the difficulty in making interpretations. The factor loadings from the skills and work context data are shown in Table B.1 and Table B.2 in Appendix B.1.

Generally, principal components obtained from PCA do not have clear economic meanings. Nevertheless, I intend to provide some basic intuitions for the principal components by analyzing their factor loadings on different skills and work context items. From Table B.1, we see that the first principal component of skills (Skill 1 hereafter) places large positive loadings on various motor skills. Therefore, I interpret Skill 1 as a motor skills index. By the same token, I consider Skill 2 as social skills and Skill 3 as cognitive skills. Next, I provide intuitions for the work context principal components based on their factor loadings in Table B.2. I find that the three principal components roughly correspond to white-collar, pink-collar, and blue-collar work settings respectively.

3.5 Occupational Flexibility and O*NET Characteristics

In this section, I explore the relationship between occupation flexibility and occupation characteristics. Specifically, I regress the standardized elasticity of earnings on standardized indices of skills and work context obtained in Section 3.4.2. I use the occupation crosswalk provided by the U.S. Census Bureau[11] to match ACS oc-

[10] PCA is also used by Yamaguchi (2012) to summarize information in the DOT.

[11] The occupation crosswalk can be found at https://www.census.gov/people/io/methodology/.

cupations with those in O*NET. One ACS occupation can have multiple matches in O*NET, since O*NET occupations are sometimes cross-referenced by industry. In these cases, I duplicate the ACS occupation to match with several O*NET occupations instead of collapsing those O*NET occupations into one. The number of occupations with a match is 912.

The regression results are presented in Table 3.4. The first regression only includes indices of skills as regressors. The results show that occupations that require more social skills are less flexible (i.e. have larger elasticity), so are those involving more motor skills, though with a smaller magnitude. Yet the reverse is true for cognitive skills. More cognitive skills in an occupation translates into more flexibility.

However, the interpretation changes when work context indices are also included as regressors. Comparing Column (3) to Column (1), we see that the coefficient on motor skills increases dramatically, while that on social skills is almost down to zero. The cognitive skills coefficient is largely unaffected. To explain these changes from the Omitted Variable Bias (OVB) perspective, I present results from the auxiliary regressions of each work context index on the full set of skills indices in Table 3.5.

Table 3.5 reveals that, not surprisingly, motor skills are heavily involved in a blue-collar work setting, and less so in a white-collar or pink-collar setting. Given that the direct effects of work context indices on elasticity are all positive (see Column (3) in Table 3.4), omitting white-collar and pink-collar indices causes a downward biased in the motor skills coefficient in Column (1), while omitting the blue-collar index causes an upward bias. It turns out that the size of the downward bias is larger than that of the upward bias, which results in an overall downward bias in the estimated coefficient on motor skills in Column (1). The biases in the coefficients on social skills and cognitive skills can be analyzed in a similar fashion. It is worth noting that the effect of social skills on flexibility can be mostly explained away by the positive correlation between social skills and a white-collar work setting.

Finally, I summarize the main takeaways from Table 3.4. Occupations requiring high motor skills performed in a blue-collar work setting are less flexible, and so are occupations in a white-collar office environment when skill levels are held constant. However, all else equal, an occupation involving high cognitive skills tends to be more flexible. This provides a possible explanation for the finding in Section 3.3 that advanced degree holders are more likely to work in flexible occupations.

3.6 Conclusions

In this study, I make two contributions to the economics literature on workplace flexibility. First, I provide further evidence that different types of workers exhibit varying occupational choice behaviors in terms of workplace flexibility. In the ACS sample of college-educated full-time workers, I document that these groups of individuals are more likely to be found in flexible occupations relative to their comparison groups: women, women with children, married individuals, and advanced degree holders. These findings are largely consistent with those in the previous literature and provide possible explanations for the gender wage gap and motherhood wage penalty.

Second, I use the detailed occupation characteristics information in O*NET to explain the differential cost of providing flexibility across occupations. I focus on two sets of occupation characteristics — i.e. the required skill levels and the typical work context. By including required skill levels as occupation characteristics, I incorporate the human capital theory in explaining the varying cost of flexibility across occupations. Through PCA, I reduce the dimension of each set of characteristics down to 3. Based on the factor loadings, I interpret the skills indices as motor, social and cognitive skills, and the work context indices as white-collar, pink-collar and blue-collar work settings. My results show that occupations requiring high motor skills are less flexible, yet the reverse is true for cognitive skills. Holding skill levels constant, occupations performed in blue-collar and white-collar work settings are less flexible.

These results suggest the important roles of both the human capital theory and personnel economics in understanding the varying cost of flexibility across occupations. In future research, it would be useful to construct a human capital model with a skill formation process which also incorporates the differences in flexibility across occupations.

Table 3.1: Summary Statistics of Elasticity and Hours Worked by Gender

	Female	Male
Elasticity (raw)	0.532	0.593
	(0.370)	(0.345)
Elasticity (standardized)	-0.114	0.059
	(1.038)	(0.969)
Weekly Hours Worked	43.54	46.46
	(7.42)	(9.00)
N	471,706	546,384

Notes: The data are drawn from the 2009-2011 waves of ACS. I restrict the sample to college graduates between 25 and 64 years of age working full-time full-year in a civilian job. I estimate the elasticity of annual earnings with respect to weekly hours worked for each occupation. I then drop observations with extreme values of elasticity. Standard deviations are in parentheses.

Table 3.2: Regressions of Elasticity of Earnings on Worker Characteristics

	Female	Female	Male
Married	-0.0802***	-0.0823***	-0.0351***
	(0.0033)	(0.0033)	(0.0031)
Mom	-0.1019***		
	(0.0037)		
Child Age < 6 Only		-0.0405***	
		(0.0058)	
Child Age 6-17 Only		-0.1167***	
		(0.0042)	
Children Age < 6 & 6-17		-0.1633***	
		(0.0072)	
Education (default = BA)			
Masters	-0.4568***	-0.4578***	-0.1792***
	(0.0032)	(0.0032)	(0.0030)
Professional	-0.4198***	-0.4217***	-0.4478***
	(0.0081)	(0.0081)	(0.0067)
Doctorate	-0.5141***	-0.5166***	-0.5859***
	(0.0066)	(0.0067)	(0.0055)
R-squared	0.0547	0.0553	0.0344
N	471,706	471,706	546,384

Notes: All regressions also control for age as a quartic, race and ACS years. Robust standard errors are in parentheses.

Table 3.3: Regressions of Weekly Hours Worked on Worker Characteristics

	Female	Female	Male
Married	-0.1192***	-0.0933***	0.8991***
	(0.0242)	(0.0244)	(0.0291)
Mom	-1.1735***		
	(0.0262)		
Child Age < 6 Only		-1.4398***	
		(0.0373)	
Child Age 6-17 Only		-1.0159***	
		(0.0305)	
Children Age < 6 & 6-17		-1.3635***	
		(0.0478)	
Education (default = BA)			
Masters	0.7005***	0.7049***	0.4591***
	(0.0232)	(0.0232)	(0.0279)
Professional	4.0319***	4.0409***	4.2277***
	(0.0623)	(0.0623)	(0.0537)
Doctorate	3.1946***	3.2065***	2.0236***
	(0.0709)	(0.0709)	(0.0608)
R-squared	0.0308	0.0310	0.0306
N	471,706	471,706	546,384

Notes: All regressions control for age as a quartic, race and ACS years. Robust standard errors are in parentheses.

Table 3.4: Regressions of Elasticity of Earnings on O*NET Occupation Characteristics Indices

	(1)	(2)	(3)
Skill 1: Motor Skills	0.0685*		0.2700***
	(0.0400)		(0.0614)
Skill 2: Social Skills	0.1001***		0.0256
	(0.0274)		(0.0372)
Skill 3: Cognitive Skills	-0.0759***		-0.0709***
	(0.0211)		(0.0233)
Work Context 1: White-Collar		0.0595	0.2525***
		(0.0365)	(0.0609)
Work Context 2: Pink-Collar		-0.0371	0.0697
		(0.0465)	(0.0623)
Work Context 3: Blue-Collar		0.1543***	0.1140***
		(0.0237)	(0.0241)
R-squared	0.0206	0.0288	0.0570
N	912	912	912

Notes: The dependant variable in all regressions is standardized elasticity of annual earnings with respect to weekly hours worked. The regressors are standardized principal components of skills and work context data from O*NET. Robust standard errors are in parentheses.

Table 3.5: Auxiliary Regressions of Work Context on Skills Indices

	Work Context 1: White-Collar	Work Context 2: Pink-Collar	Work Context 3: Blue-Collar
Skill 1: Motor Skills	-0.7431***	-0.4130***	0.1303***
	(0.0170)	(0.0282)	(0.0310)
Skill 2: Social Skills	0.3806***	-0.3188***	0.0058
	(0.0208)	(0.0266)	(0.0266)
Skill 3: Cognitive Skills	0.0569***	-0.1996***	-0.0479
	(0.0162)	(0.0223)	(0.0310)
R-squared	0.6909	0.3174	0.0192
N	912	912	912

Notes: Robust standard errors are in parentheses.

Appendix A

Appendix to Chapter 2

A.1 Proof of Proposition 1 and Lemma 2

First, I show that $\text{cov}[s_0, \text{E}(\Lambda v_0 + e_0|S_t, q, D_t)] = 0$, and $\text{cov}[z, \text{E}(\Lambda v_0 + e_0|S_t, q, D_t)] = \text{cov}[v_0, \text{E}(\Lambda v_0 + e_0|S_t, q, D_t)]$.

$$
\begin{aligned}
& \text{cov}[s_0, \text{E}(\Lambda v_0 + e_0|S_t, q, D_t)] \\
= \; & \text{E}[s_0 \text{E}(\Lambda v_0 + e_0|S_t, q, D_t)] - \text{E}(s_0) \times \text{E}[\text{E}(\Lambda v_0 + e_0|S_t, q, D_t)], \\
= \; & \text{E}[\text{E}[s_0(\Lambda v_0 + e_0)|S_t, q, D_t]] - \text{E}(s_0) \times \text{E}[\text{E}(\Lambda v_0 + e_0|S_t, q, D_t)], \text{ since } s_0 \subseteq S_t \\
= \; & \text{E}[s_0(\Lambda v_0 + e_0)] - \text{E}(s_0) \times \text{E}(\Lambda v_0 + e_0), \text{ by Law of Iterated Expectations} \\
= \; & \text{cov}(s_0, \Lambda v_0 + e_0) \\
= \; & 0.
\end{aligned}
$$

Similarly, $\text{cov}[q, \text{E}(\Lambda v_0 + e_0|S_t, q, D_t)] = 0$, and therefore

$$
\begin{aligned}
& \text{cov}[z, \text{E}(\Lambda v_0 + e_0|S_t, q, D_t)] \\
= \; & \text{cov}[\gamma_1 q + \gamma_2 s_0 + v_0, \text{E}(\Lambda v_0 + e_0|S_t, q, D_t)] \\
= \; & \text{cov}[v_0, \text{E}(\Lambda v_0 + e_0|S_t, q, D_t)].
\end{aligned}
$$

Using these and the OLS regression formula, one can express Φ_{st} and Φ_{zt} as

$$\begin{aligned}
\Phi_{st} &= \frac{\text{var}(z)\text{cov}[s_0, \text{E}(\Lambda v_0 + e_0|S_t, q, D_t)] - \text{cov}(s_0, z)\text{cov}[z, \text{E}(\Lambda v_0 + e_0|S_t, q, D_t)]}{\text{var}(s_0)\text{var}(z) - [\text{cov}(s_0, z)]^2} \\
&= -\frac{\text{cov}(s_0, z)\text{cov}[v_0, \text{E}(\Lambda v_0 + e_0|S_t, q, D_t)]}{\text{var}(s_0)\text{var}(z) - [\text{cov}(s_0, z)]^2}, \\
\Phi_{zt} &= \frac{\text{var}(s_0)\text{cov}[z, \text{E}(\Lambda v_0 + e_0|S_t, q, D_t)] - \text{cov}(s_0, z)\text{cov}[s_0, \text{E}(\Lambda v_0 + e_0|S_t, q, D_t)]}{\text{var}(s_0)\text{var}(z) - [\text{cov}(s_0, z)]^2} \\
&= \frac{\text{var}(s_0)\text{cov}[v_0, \text{E}(\Lambda v_0 + e_0|S_t, q, D_t)]}{\text{var}(s_0)\text{var}(z) - [\text{cov}(s_0, z)]^2}.
\end{aligned}$$

As t increases and more information becomes available to the employer, $\text{E}(\Lambda v_0 + e_0|S_t, q, D_t)$ converges from 0 to $\Lambda v_0 + e_0$, and $\text{cov}[v_0, \text{E}(\Lambda v_0 + e_0|S_t, q, D_t)]$ converges monotonically from 0 to $\text{cov}(v_0, \Lambda v_0 + e_0)$. Therefore, Φ_{zt} increases with t, which proves Proposition 1.

Since

$$\frac{\Phi_{st}}{\Phi_{zt}} = -\frac{\text{cov}(s_0, z)}{\text{var}(s_0)} = -\Phi_{zs},$$

Lemma 2 follows.

A.2 Proof of Corollary 5

In this appendix, I prove that the estimator of r implied by Proposition 3 is mathematically equivalent to the one implied by Proposition 4.

In Proposition 3, solving Equation (2.10) – (2.12) yields

$$r = \frac{\dot{b}_{st} + \Phi_{zs}\dot{b}_{zt}}{\dot{\Phi}_t^{ss} + \Phi_{zs}\dot{\Phi}_t^{sz}},$$

where dotted variables are derivatives with respect to t.

Using the OLS formula and denoting $V_s = \text{Var}(s_0)$, $V_z = \text{Var}(z)$ and $COV_{sz} =$

$\text{cov}(s_0, z)$, I obtain

$$
\begin{aligned}
r &= \frac{V_z \dot{\text{cov}}(w_t, s_0) - COV_{sz}\dot{\text{cov}}(w_t, z) + \frac{COV_{sz}}{V_s}[V_s \dot{\text{cov}}(w_t, z) - COV_{sz}\dot{\text{cov}}(w_t, s_0)]}{V_z \dot{\text{cov}}(s_t, s_0) - COV_{sz}\dot{\text{cov}}(s_t, z) + \frac{COV_{sz}}{V_s}[V_s \dot{\text{cov}}(s_t, z) - COV_{sz}\dot{\text{cov}}(s_t, s_0)]} \\
&= \frac{\dot{\text{cov}}(w_t, s_0)}{\dot{\text{cov}}(s_t, s_0)},
\end{aligned}
$$

which is equivalent to the one derived from Equation (2.13) in Proposition 4.

A.3 Theory Implications in a Selected Subsample with Constant s

In this appendix I derive theory implications of the generalized EL-SD model in a selected subsample with constant s. I define CS_t as the event that the realizations of s do not change up to period t.[1] Since the data generating process in the population is not affected by sample selection rules, the log wage equation (2.6) still holds. However, when regressing w_t on s_0 and z in the selected subsample instead of the whole sample, the conditional expectation function (CEF) in Equation (2.7) becomes

$$
\text{E}(w_t | s_0, z, t, CS_t) = b_{st}^{CS} s_0 + b_{zt}^{CS} z + H^*(t),
$$

where the sample selection rule CS_t is added to the list of conditions. The rest of the analysis closely mirrors that in Section 2.2.2, with the CS_t condition added to the CEFs for all the auxiliary regressions. Notice that when conditioning on CS_t, the first auxiliary regression yields $\Phi_t^{ss} = 1$ and $\Phi_t^{sz} = 0$. This seems to reduce the theory implications in Equations (2.8) and (2.9) to those in AP, which would prove the applicability of AP's empirical strategy in a selected sample with constant s. However, this is not the case, because Proposition 1 no longer holds when conditioning on CS_t. Recall that one of the key building blocks in the proof of Proposition 1 is that $\text{cov}(s_0, \text{E}(\Lambda v_0 + e_0 | S_t, q, D_t)) = 0$. Yet when conditioning on

[1] Technically, $CS_t = \{\omega \in \Omega | s_0(\omega) = s_1(\omega) = \cdots = s_t(\omega)\}$.

CS_t,

$$\begin{aligned}
& \mathrm{cov}[s_0, \mathrm{E}(\Lambda v_0 + e_0|S_t, q, D_t)|CS_t] \\
= \; & \mathrm{E}[s_0 \mathrm{E}(\Lambda v_0 + e_0|S_t, q, D_t)|CS_t] - \mathrm{E}(s_0|CS_t) \cdot \mathrm{E}[\mathrm{E}(\Lambda v_0 + e_0|S_t, q, D_t)|CS_t] \\
= \; & \mathrm{E}\{\mathrm{E}[s_0(\Lambda v_0 + e_0)|S_t, q, D_t]|CS_t\} - \mathrm{E}(s_0|CS_t) \cdot \mathrm{E}(\Lambda v_0 + e_0|CS_t) \\
= \; & \mathrm{E}[s_0(\Lambda v_0 + e_0)|CS_t\} - \mathrm{E}(s_0|CS_t) \cdot \mathrm{E}(\Lambda v_0 + e_0|CS_t), \text{ since } CS_t \subseteq S_t \\
= \; & \mathrm{cov}(s_0, \Lambda v_0 + e_0|CS_t),
\end{aligned}$$

which is not zero in general, since the condition CS_t might contain information about the initial assessment error $\Lambda v_0 + e_0$. Therefore, the theory implications in AP do not hold in a selected subsample with constant s given the generalized EL-SD model where s_t transitions in the population.

A.4 The NLSY79 Data

The data are drawn from the 1979-2004 waves of NLSY79. I restrict the sample to non-Hispanic white males who have completed exactly 12 years of education upon labor market entry and do not change education afterwards. I define labor market entry as the year in which an individual reports to have left school for the first time, and define potential experience as the number of years since labor market entry. I only keep observations after the individual has left school for the first time—i.e. observations with positive potential experience.

I keep all valid and non-missing marital status records (i.e. never married, married, separated, divorced, and widowed), and code marital status to be 1 if an individual is married and 0 otherwise. I drop individuals with inconsistent records, i.e. individuals who report to be married at some point but then report to be never married in later years. I construct the Period 1 marital status variable s_1 for each individual using marital status records in the first year after an individual reports to have left school for the first time. When such records are missing, I impute them as much as possible using information about the years in which marriage spells began or ended. And I similarly construct marital status variables in other periods for each individual.

Wage is defined as the hourly rate of pay at the most recent job, which is obtained from the Current Population Survey (CPS) section of the NLSY79. Observations holding military jobs, jobs at home or jobs without pay are excluded from the analysis. Observations with missing wages or extreme wages (less than \$1 or more than \$100) are also dropped. I standardize the AFQT score to have mean zero and standard deviation one for each age at the time of the test. I define the noncognitive ability measure as the average of the Rotter Locus of Control Scale and the Rosenberg Self-Esteem Scale. I standardize both scales by age before taking the average. Individuals with missing ability measures are dropped, as well as observations with missing values in the urban residence and region of residence variables. The final sample contains 11,408 person-year observations from 1101 individuals. Table A.1 shows in detail the sample selection process and sequential changes in the sample size.

A.5 Proof of Invariance under Period Relabeling

Without loss of generality, I prove that both the treatment effect identification strategy and empirical test for EL-SD described in Section 2.2.2 still hold when replacing s_0 with s_4. To facilitate the proof, I first rewrite the EL-SD model in terms of s_4.

Recall that the worker's log productivity is specified in Equation (2.1) as[2]

$$y_t = rs_t + \alpha_1 q + \Lambda z + \eta + H(t),$$

where z and η are not directly observed by the employer. When setting wages in each

[2] Subscript i is suppressed.

period, the employer forms expectations of worker productivity (i.e. expectations of z and η) conditioning on all available and relevant information up to that period. However, regardless of how the employer actually forms expectations, I can always express z and η in any form of conditional expectations plus corresponding residuals. To facilitate the intended proof, I express z and η as

$$\begin{aligned} z &= \mathrm{E}(z|s_4) + v_4 = \gamma_0 + \gamma s_4 + v_4, \\ \eta &= \mathrm{E}(\eta|s_4) + e_4 = \alpha_0 + \alpha s_4 + e_4. \end{aligned}$$

By definition, the error terms v_4 and e_4 have mean zero and are uncorrelated with s_4. I emphasize that by expressing z and η in the above manner, I am not imposing any assumptions about how the employer forms expectations of z and η in each period. Especially, I am not assuming that the employer only uses information contained in s_4 when forming expectations in any period. These expressions of z and η simply define residuals v_4 and e_4, and are completely free from any assumptions of the employer's expectation formation. Moreover, these expressions are also free from any functional form assumptions, since $\mathrm{E}(z|s_4)$ and $\mathrm{E}(\eta|s_4)$ are always linear when s_4 is a dummy variable. In fact, one can easily show that $\gamma_0 = \mathrm{E}(z|s_4 = 0)$, and $\gamma = \mathrm{E}(z|s_4 = 1) - \mathrm{E}(z|s_4 = 0)$, and similarly for α_0 and α.

Substituting z and η into the log productivity specification yields

$$y_t = rs_t + (\Lambda\gamma_0 + \alpha_0) + (\Lambda\gamma + \alpha)s_4 + \alpha_1 q + H(t) + (\Lambda v_4 + e_4).$$

Each period the employer observes a noisy indicator of the worker's productivity $\xi_t = y_t + \epsilon_t$, where ϵ_t is transitory variation in the worker's performance. When $t \geqslant 4$, since the employer directly observes s_4, the only factor in y_t that is unknown to the employer is $\Lambda v_4 + e_4$, and therefore observing ξ_t is equivalent to observing $d_t = \Lambda v_4 + e_4 + \epsilon_t$. Denote the history of d as $D_t = \{d_1, d_2, \cdots, d_t\}$ and the history of s as $S_t = \{s_0, s_1, \cdots, s_t\}$. Define $\mu_t = \Lambda v_4 + e_4 - \mathrm{E}(\Lambda v_4 + e_4|S_t, q, D_t)$. By definition,

μ_t is uncorrelated with S_t, q and D_t. I assume in addition that μ_t is independent of S_t, q and D_t.

Substituting y_t into the equilibrium wage equations

$$W_t = \mathrm{E}(Y_t|S_t, q, D_t)\exp^{\zeta_t}$$

gives the log wage process

$$w_t = rs_t + (\Lambda\gamma_0 + \alpha_0) + (\Lambda\gamma + \alpha)s_4 + \alpha_1 q + H^*(t) + \mathrm{E}(\Lambda v_4 + e_4|S_t, q, D_t) + \zeta_t,$$

where $H^*(t) = H(t) + \log(\mathrm{E}(\exp^{\mu_t}))$.

The rest of the analysis resembles that in Section 2.2.2, with s_0 replaced by s_4 in all regressions. The two key building blocks, $\mathrm{cov}[s_4, \mathrm{E}(\Lambda v_4 + e_4|S_t, q, D_t)] = 0$ and $\mathrm{cov}[z, \mathrm{E}(\Lambda v_4 + e_4|S_t, q, D_t)] = \mathrm{cov}[v_4, \mathrm{E}(\Lambda v_4 + e_4|S_t, q, D_t)]$, can be proved similarly to those in Appendix A.1.

Table A.1: Sample Size by Sample Selection Criteria

	Individuals	Observations
Full Sample (Year 1979-2004, NLSY79)	12,686	266,406
Remaining Sample Size		
Keep non-Hispanic white males	3,790	79,590
Drop if marital status was missing	3,790	58,481
Drop individuals with inconsistent marital status records	3,771	58,149
Drop individuals who never left school or do not declare when they first left school	3,553	54,892
Keep observations after the individual has left school for the first time	3,546	47,385
Keep observations with exactly 12 years of education	2,059	22,629
Keep individuals who never obtained more years of education after first leaving school	1,361	17,981
Drop if wage was missing	1,257	15,111
Drop if not working in civilian jobs for pay, or wages less than $1 or more than $100	1,228	12,612
Drop if AFQT was missing	1,139	11,931
Drop if noncognitive ability measure was missing	1,105	11,651
Drop if urban residence was missing	1,101	11,423
Drop if region of residence was missing	1,101	11,408
Final Sample	1,101	11,408

Appendix B

Appendix to Chapter 3

B.1 PCA Factor Loadings

Table B.1 and Table B.2 show the PCA factor loadings from the O*NET skills and work context data, both sorted by the factor loadings of the first principal component.

Table B.1: PCA Factor Loadings: O*NET Skills Data

Skills	Loading 1	Loading 2	Loading 3
Equipment Maintenance	0.229	-0.319	-0.018
Repairing	0.224	-0.321	-0.035
Operation and Control	0.205	-0.293	-0.050
Troubleshooting	0.168	-0.347	-0.001
Equipment Selection	0.142	-0.303	-0.004
Operation Monitoring	0.119	-0.282	-0.030
Quality Control Analysis	0.082	-0.307	-0.016
Installation	0.071	-0.162	-0.008
Technology Design	-0.070	-0.186	0.007
Coordination	-0.102	-0.044	-0.132
Time Management	-0.118	-0.050	-0.097
Programming	-0.122	-0.131	0.087
Management of Material Resources	-0.125	-0.142	-0.396
Service Orientation	-0.125	0.024	-0.082
Monitoring	-0.135	-0.072	-0.048
Management of Personnel Resources	-0.148	-0.089	-0.195
Negotiation	-0.154	-0.002	-0.202
Complex Problem Solving	-0.156	-0.092	0.029
Social Perceptiveness	-0.156	0.012	-0.041
Critical Thinking	-0.161	-0.051	0.029
Management of Financial Resources	-0.162	-0.117	-0.488
Judgment and Decision Making	-0.167	-0.071	0.002
Persuasion	-0.167	-0.017	-0.155
Active Listening	-0.168	-0.007	0.047
Instructing	-0.171	-0.079	0.017
Mathematics	-0.173	-0.157	0.003
Learning Strategies	-0.188	-0.072	-0.002
Speaking	-0.189	-0.008	0.035
Active Learning	-0.201	-0.079	0.046
Systems Analysis	-0.207	-0.130	-0.054
Operations Analysis	-0.209	-0.165	0.095
Systems Evaluation	-0.219	-0.140	-0.075
Reading Comprehension	-0.219	-0.058	0.128
Writing	-0.223	-0.033	0.083
Science	-0.233	-0.256	0.635

Table B.2: PCA Factor Loadings: O*NET Work Context Data

Work Context	Loading 1	Loading 2	Loading 3
Electronic Mail	0.243	0.335	-0.033
Spend Time Sitting	0.195	0.143	0.149
Indoors, Environmentally Controlled	0.176	0.005	-0.173
Telephone	0.113	0.266	-0.097
Letters and Memos	0.113	0.228	-0.082
Public Speaking	0.080	0.131	-0.020
Deal With External Customers	0.074	0.164	-0.220
Structured versus Unstructured Work	0.054	0.100	-0.021
Freedom to Make Decisions	0.032	0.105	-0.025
Contact With Others	0.031	0.069	-0.140
Level of Competition	0.026	0.070	0.009
Frequency of Conflict Situations	0.022	0.115	-0.156
Coordinate or Lead Others	0.017	0.132	-0.133
Face-to-Face Discussions	0.013	0.072	-0.063
Exposed to Disease or Infections	0.011	0.019	-0.505
Work With Work Group or Team	0.010	0.073	-0.127
Importance of Being Exact or Accurate	0.010	-0.001	-0.069
Degree of Automation	0.006	-0.020	0.015
Importance of Repeating Same Tasks	0.005	-0.027	-0.096
Deal With Unpleasant or Angry People	0.005	0.043	-0.206
Impact of Decisions on Co-workers or Company Results	0.004	0.123	-0.116
Duration of Typical Work Week	-0.003	0.093	0.036
Frequency of Decision Making	-0.005	0.107	-0.155
Deal With Physically Aggressive People	-0.007	0.052	-0.162
Time Pressure	-0.007	0.028	-0.039
Exposed to Radiation	-0.019	0.016	-0.164
Work Schedules	-0.025	0.027	0.040
Responsibility for Outcomes and Results	-0.030	0.121	-0.117
Physical Proximity	-0.034	-0.030	-0.273
Consequence of Error	-0.064	0.105	-0.183
Spend Time Making Repetitive Motions	-0.076	-0.197	-0.011
Spend Time Climbing Ladders, Scaffolds, or Poles	-0.098	0.073	0.050
Spend Time Keeping or Regaining Balance	-0.099	0.008	-0.029
Outdoors, Under Cover	-0.109	0.218	0.089
Exposed to Whole Body Vibration	-0.109	0.067	0.060
In an Enclosed Vehicle or Equipment	-0.114	0.363	0.134
Spend Time Kneeling, Crouching, Stooping, or Crawling	-0.125	-0.011	-0.035

Wear Specialized Protective or Safety Equipment such as Breathing Apparatus, Safety Harness, Full Protection Suits, or Radiation Protection	-0.132	0.062	-0.104
Responsible for Others' Health and Safety	-0.141	0.107	-0.233
Spend Time Walking and Running	-0.141	-0.077	-0.122
In an Open Vehicle or Equipment	-0.143	0.100	0.103
Exposed to High Places	-0.149	0.138	0.070
Pace Determined by Speed of Equipment	-0.157	-0.134	0.026
Cramped Work Space, Awkward Positions	-0.168	0.057	-0.061
Extremely Bright or Inadequate Lighting	-0.168	0.110	0.031
Spend Time Standing	-0.171	-0.157	-0.144
Spend Time Using Your Hands to Handle, Control, or Feel Objects, Tools, or Controls	-0.172	-0.187	-0.064
Sounds, Noise Levels Are Distracting or Uncomfortable	-0.180	0.022	-0.020
Spend Time Bending or Twisting the Body	-0.185	-0.112	-0.104
Indoors, Not Environmentally Controlled	-0.187	0.162	0.152
Exposed to Hazardous Conditions	-0.194	0.068	-0.042
Outdoors, Exposed to Weather	-0.200	0.346	0.155
Exposed to Minor Burns, Cuts, Bites, or Stings	-0.218	-0.018	-0.008
Very Hot or Cold Temperatures	-0.228	0.143	0.123
Exposed to Contaminants	-0.251	-0.014	-0.085
Exposed to Hazardous Equipment	-0.257	0.059	0.074
Wear Common Protective or Safety Equipment such as Safety Shoes, Glasses, Gloves, Hearing Protection, Hard Hats, or Life Jackets	-0.307	0.003	-0.153

Bibliography

Acemoglu, D. and Autor, D. (2011), "Skills, Tasks and Technologies: Implications for Employment and Earnings," *Handbook of labor economics*, 4, 1043–1171.

Altonji, J. G. and Pierret, C. R. (2001), "Employer Learning and Statistical Discrimination," *The Quarterly Journal of Economics*, 116, 313–350.

Arcidiacono, P., Bayer, P., and Hizmo, A. (2010), "Beyond Signaling and Human Capital: Education and the Revelation of Ability," *American Economic Journal: Applied Economics*, 2, 76–104.

Bauer, T. K. and Haisken-DeNew, J. P. (2001), "Employer Learning and the Returns to Schooling," *Labour Economics*, 8, 161–180.

Blau, F. D. and Kahn, L. M. (2016), "The Gender Wage Gap: Extent, Trends, and Explanations," *NBER Working Paper*.

Bronson, M. A. (2015), "Degrees are Forever: Marriage, Educational Investment, and Lifecycle Labor Decisions of Men and Women," *Working Paper*.

Edwards, R. (2014), "Women's Labor Supply – Motherhood and Work Schedule Flexibility," *Working Paper*.

Farber, H. S. and Gibbons, R. (1996), "Learning and Wage Dynamics," *The Quarterly Journal of Economics*, 111, 1007–1047.

Flabbi, L. and Moro, A. (2012), "The Effect of Job Flexibility on Female Labor Market Outcomes: Estimates from a Search and Bargaining Model," *Journal of Econometrics*, 168, 81–95.

Galindo-Rueda, F. (2003), "Employer learning and schooling-related statistical discrimination in Britain," *IZA Discussion Paper*.

Goldin, C. (2014), "A Grand Gender Convergence: Its Last Chapter," *American Economic Review*, 104, 1091–1119.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006), "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24, 411–482.

Kahn, L. B. (2013), "Asymmetric Information between Employers," *American Economic Journal: Applied Economics*, 5, 165–205.

Kahn, L. B. and Lange, F. (2014), "Employer Learning, Productivity, and the Earnings Distribution: Evidence from Performance Measures," *The Review of Economic Studies*, 81, 1575–1613.

Killewald, A. and Gough, M. (2013), "Does Specialization Explain Marriage Penalties and Premiums?" *American Sociological Review*, 78, 477–502.

Lange, F. (2007), "The Speed of Employer Learning," *Journal of Labor Economics*, 25, 1–35.

Light, A. and McGee, A. (2015), "Does Employer Learning Vary by Schooling Attainment? The Answer Depends on How Career Start Dates Are Defined," *Labour Economics*, 32, 57–66.

Pinkston, J. C. (2009), "A Model of Asymmetric Employer Learning with Testable Implications," *The Review of Economic Studies*, 76, 367–394.

Rodgers, W. M. and Stratton, L. S. (2009), "Male Marital Wage Differentials: Training, Personal Characteristics, and Fixed Effects," *Economic Inquiry*, 48, 722–742.

Schönberg, U. (2007), "Testing for Asymmetric Employer Learning," *Journal of Labor Economics*, 25, 651–691.

Wiswall, M. and Zafar, B. (2016), "Preference for the Workplace, Investment in Human Capital, and Gender," *Working Paper*.

Yamaguchi, S. (2012), "Tasks and Heterogeneous Human Capital," *Journal of Labor Economics*, 30, 1–53.

Zhang, Y. (2007), "Employer Learning Under Asymmetric Information: The Role of Job Mobility," *Working Paper*.

Biography

Xiaomin Fu was born in Liaoning, China on March 19, 1989. She attended the University of International Business and Economics in Beijing, where she received her Bachelor's degree in Finance in 2010. She started her graduate studies in the economics department at Duke University in 2010, where she earned a Master's degree in Economics in 2012 and a Ph.D. degree in Economics in 2017. As a graduate student researcher, she specialized in Labor Economics. She will join Amazon as an economist in the summer of 2017.