

Multidimensional Learning, Job Mobility, and Earnings Dynamics*

Joseph Han[†]

University of Wisconsin-Madison

April 12, 2016

Abstract

Various types of uncertainties can coexist and jointly affect young workers' decisions. This paper newly introduces the possibility of multidimensional learning about worker ability and job match quality to a model of work decisions under search frictions. In this setup, worker ability can affect job change and individual employment through an information channel. This mechanism produces a unique prediction which is testable if there exists a measure of unknown ability. I estimate the life-cycle structural model, which also allows flexible general and job-specific skill accumulation, by indirect inference using a sample from the NLSY79 data. From simulation results on life-cycle earnings dynamics, I find that the contribution of job shopping to average earnings growth is much higher than previous estimates; also, individual heterogeneity in earnings growth is mostly explained by the process of resolving uncertainties.

Keywords: Uncertainties, Ability, job match quality, job change, unemployment, individual heterogeneity, earnings dynamics, indirect inference.

JEL Classification Numbers: J24, J31, J62.

*I am deeply grateful to Chris Taber, John Kennan, and Rasmus Lentz for their guidance and support. I also would like to thank Naoki Aizawa, Richard Blundell, Mark Colas, Chao Fu, Christa Gibbs, Jesse Gregory, Andrea Guglielmo, Junjie Guo, Kevin Hutchinson, Akio Ino, Byunghoon Kang, Andrew Kidd, Tae Hoon Kim, Jeremy Lise, Chenyan Lu, Ryne Marksteiner, Ronni Pavan, Eishro Takeda, Kegon Teng Kok Tan, Sergio Urzua, Jim Walker, and participants in the Empirical Micro Lunch seminar at the University of Wisconsin-Madison. All remaining errors are mine.

[†]Ph.D. Candidate, University of Wisconsin-Madison, e-mail: joseph.han@wisc.edu.

1 Introduction

Uncertainties characterize young workers' career. Most young workers begin their career without much knowledge on their job. It is also rare that the young workers who have just completed their education fully understand their own strength and weakness in the world of work. Nevertheless, they need to make their work decisions each moment, which will affect their lifetime earnings and utility.

This paper tries to understand how uncertainties shape work decisions of young workers. In particular, the main focus here is the joint effect of uncertainties which can be different from their separate effects. This paper extends previous learnings models in the labor market by integrating two different types of uncertainty: worker ability¹, which is only partially known to both workers and employers at labor market entry; and job match quality², which is also unknown at job entry. In addition, the model has several more important features: employment-related shocks such as exogenous job destruction and recall offer arrival³, job search costs, general and job-specific skills which may accumulate differently by ability, consumption/saving choice by risk-averse workers.

Although the main contribution of this paper is empirical, some theoretical predictions are new and noteworthy. First, the generalized learning mechanism predicts that job mobility and movement into and out of employment are heterogeneous by unknown ability – less able workers are predicted to change jobs more frequently and work less than other workers. This is basically a story of *misperceived* productivity⁴ where workers cannot directly distinguish

¹Employer learning (e.g., Farber and Gibbons, 1996; Altonji and Pierret, 2001; Lange, 2007) and learning about ability (e.g., James, 2011; Papageorgiou, 2014) literature suggest worker ability may not be fully known at labor market entry. This paper can be read in connection with employer learning when asymmetric information does not meaningfully affect workers' decisions. It is also interesting to see that this multi-dimensional learning under information symmetry has a very similar prediction on job mobility with asymmetric information models (e.g., Greenwald, 1986; Schönberg, 2007; Pinkston, 2009)

²In Jovanovic (1979, 1984), where workers gradually learn about the quality of a job match, job hazards are predicted initially increasing then decreasing; also, average wages are predicted to increase in job tenure. These predictions are widely supported by empirical evidence (e.g., Farber, 1999).

³Fujita and Moscarini (2013) show recall to a previous employer is important for understanding unemployment.

⁴Jovanovic (1979, 1984) shows how *perceived* productivity affects career decisions over the life cycle, using a model of learning about job match quality.

between ability and job match quality in their signals: less able workers get worse signals than others; they misinterpret the signals as evidence of a bad job match; as a result, they are more likely to search and move; also, less able workers work *more than they would do under full information* although they still work less than other workers if leisure utility is positive. Second, the differences in job mobility between less able and more able workers are expected to disappear over time. With more observations over the life cycle, workers' belief on their ability eventually reflects their true ability unless the speed of learning is too slow. The initially strong but disappearing negative selection into job mobility, conditional on endogenous job separation, provides a potentially testable implication of this learning story. Third, the positive selection into employment by ability becomes stronger while the initially strong and negative selection by unknown ability disappears over time.

These predictions are indeed from the information side of the story.⁵ Also, the predictions, especially the over-time variation in the difference in job hazard between less and more able workers, separately identify the information story from other possible mechanisms such as job-specific skills which are accumulated differently by ability.⁶ These predictions are, however, only potentially useful for identification because it requires a measure of *unknown* ability, not only of known ability.

This paper takes an additional assumption on the nature of the Armed Forces Qualification Test (AFQT) score in the National Longitudinal Survey of Youth 1979 (NLSY79) data. I assume that the AFQT score carries over some information unused by the agents.

⁵Moscarini (2005) criticizes that Jovanovic (1979, 1984)'s predictions are from the selection side rather than the information side of the learning model. Moscarini (2005) shows, in his general equilibrium setup, that information friction uniquely predicts a Pareto upper tail of the empirical wage distribution.

⁶This is comparable to Nagypál (2007)'s. Learning-by-doing on the job increases the opportunity cost of moving as job tenure increases; learning about job match quality reduces the option value of a match as job tenure increases. The two mechanisms have similar predictions on average job mobility and wage growth. Nonetheless, the size of the match-specific surplus changes in different directions over time (on the job), so this can provide a source of identification. In her general equilibrium setting, she uses a negative productivity shock to change the value of a match and observes which matches are more fragile – old vs. new matches. Instead of the negative productivity shock in a general equilibrium setup, I use preference shocks in a partial equilibrium setup in order to trigger different responses in match continuation. In addition, I add one more dimension in the match surplus, other than job tenure, which is worker ability. Then I compare high and low ability groups over time.

This assumption is similar in spirit to Altonji and Pierret (2001) in that the econometrician has more information on worker ability than the agents.⁷ I modify the Altonji and Pierret (2001)'s assumption in two ways. First, I assume that the agents (both the workers and employers) have another source of information that is not available to the econometrician such as the SAT score or high school GPA. Second, I assume that the AFQT score was not properly understood (or rationally ignored) by the NLSY respondents except for those with some (future) military career.⁸ This is essentially about the nature of the data not related to the model, but this is important for the identification of the model.

With the additional identifying assumption mentioned above, I estimate the structural model of life-cycle labor supply and earnings, using a sample of only white male high school graduates without any further education or military experience from the NLSY79 data. For the estimation method, I use indirect inference, which is characterized by the use of auxiliary model. The auxiliary model in this paper comprises seven regression equations which explain conditional wages for job movers and stayers and job/employment transitions over the life cycle. This is in many ways similar to Altonji et al. (2013)'s auxiliary model although their structural model is totally different from mine. Nonetheless, I follow Sauer and Taber (2013) instead of Keane and Smith (2003) for the method of smoothing in indirect inference, as Keane and Smith (2003) was no longer applicable in this dynamic setting because of heavy dependence on past variables.

The estimation results are in general consistent with previous results; however, I have two new findings on earnings dynamics. First, the simulation results show that rapid average earnings growth over the first 10 years (66 percent) can be attributed to general and job-specific skill accumulation and improved job match quality, at 33 percent, 10 percent, and 24 percent, respectively. While these findings are comparable to the results from other recent models of earnings dynamics (e.g., Altonji et al., 2013), the contribution of job shopping to average earnings growth through improved job match quality is much higher in this paper

⁷Lange (2007) has more discussions on this assumption.

⁸The AFQT test was basically for military services.

where the job shopping process is allowed to be heterogeneous across workers. Recall to the last previous employer also plays a role for preserving good matches and skills specific to the matches in late career.

Second, individual heterogeneity in earnings growth is mostly explained by the process of resolving uncertainties. Learning about ability and subsequent wage changes are obvious reasons. Individual earnings converge to the level associated with true ability with more information on ability. Heterogeneous job mobility by unknown ability is another important reason. More able workers are predicted to stay longer during their early career but move relatively more during their late career. Their average match quality increases slowly in their early career, and fast afterwards, compared to median-ability workers. Other channels of earnings growth are not very different by ability. Both general and job-specific skill accumulation processes are fairly homogeneous across workers.

The rest of this paper is structured as follows: Section 2 presents the model. Section 3 explains the data with several observed empirical patterns. Section 4 illustrates the identification and the estimation of the model. Section 5 performs several counterfactual experiments, and Section 6 concludes.

2 The Model

2.1 Overview

I present a model where individuals make work-related decisions and choose consumption so as to maximize lifetime utility, looking forward up to 60 years. In this model economy, worker ability and job match quality are at least partially uncertain for both workers and employers, and these two unknown objects must be inferred from observed signals on the unknown objects. The labor market is also marked with search frictions.

Each individual begins career after completion of their study.⁹ Each period of their

⁹I assume away educational choice; then, I restrict my sample to high school graduates who never attained

working life begins with *perceived* productivity shock. Workers (and all employers) observe a new signal on their productivity in the previous period – for example, a formal evaluation on job performance – and updates their beliefs accordingly. Then, all individuals can choose whether or not to do job search which incurs search costs. After job search, shocks directly related to employment are realized: jobs are exogenously destroyed; unemployed workers get a recall offer from the most recent employer; and job seekers get a new job offer. After observing all realized shocks, individuals choose the best career option available. After then, people who choose to work sign on a new wage contract for the period. Finally, individuals choose the level of consumption and savings. This process is repeated for $T(=40)$ years, and people live $R(= 20)$ years after fully retired.

Some other features of the model are noteworthy: workers are either risk-neutral or risk-averse, and partial insurance provided by the government such as welfare and unemployment insurance can play a role on career decisions; labor productivity is log-linear in ability and job match quality that are constant over time; also, there exist two types of skill accumulation (learning-by-doing), general and job-specific¹⁰, and each skill accumulation can be heterogeneous according to (learning) ability.¹¹

In the following description of the model, I omit the individual subscript, i , for expository convenience. I use the subscript only where individual heterogeneity is the main topic.

additional post-secondary education. Incorporating educational choice is very interesting but left for future research.

¹⁰The distinction between general and job-specific skills is the same with Becker (1964)’s. General skills are transferable to other jobs, while job-specific skills are not. This distinction is an extreme case – most skills may be only partially transferable across jobs (e.g., Gathmann and Schönberg, 2010; Sanders and Taber, 2012). This paper, however, focuses on job mobility, so the traditional distinction is good enough for that purpose.

¹¹For example, smart people learn how to do things faster than others. The speed of skill accumulation of each type can be differently affected by ability. I assume that ability and “learning ability” are the same, but I allow the productivity effects of ability to be different in each case (including zero effect).

2.2 Ability and Prior Information

New entrants to the labor market are the same but different in one aspect, which is their work ability¹², θ_i . This ability is not directly observed and follows a standard normal distribution. A signal on individual ability is available at (high school) graduation, right before labor market entry: A test score, θ_i^{TS} , conveys partial information on individual ability to each individual. This test score is a noisy measure of true ability: $\theta_i^{TS} = \theta_i + \zeta_i, \zeta_i \sim N(0, \sigma_\zeta^2), \sigma_\zeta^2 < \infty$. The measurement errors are independent of true ability. This test score is assumed to be publicly accessible in the labor market.¹³ The distributions of ability and measurement errors are common knowledge.

The econometrician does not observe θ_i^{TS} . Instead, the econometrician observes another ability measure: $\theta_i^{AFQT} = \theta_i + \zeta'_i, \zeta'_i \sim N(0, \sigma_{\zeta'}^2), \sigma_{\zeta'}^2 < \infty$. ζ'_i is independent of true ability and ζ_i . This information is not available in the labor market. This assumption is discussed in more detail in the data section.¹⁴

The individual i 's belief on own ability is updated with the realization of the period 0 signal on ability, θ_i^{TS} . Each individual's belief on ability at labor market entry (period 0) changes according to the (standardized) test score: $E[\theta_i | \theta_i^{TS}] = \theta_i^{TS} / (1 + \sigma_\zeta^2)$. The uncertainty is still the same across individuals: $Var[\theta_i | \theta_i^{TS}] = \sigma_\zeta^2 / (1 + \sigma_\zeta^2)$.

$E[\theta_i | \theta_i^{AFQT, STD}]$ and $Var[\theta_i | \theta_i^{AFQT, STD}]$ also follow similar expressions only with additional scale adjustment. These expressions are useful for simulation and estimation purpose.

¹²Although the model is easily generalizable to a multidimensional ability case, the current version focuses only single-dimensional ability which is general across all jobs. Also, I do not distinguish between innate ability and initial skill because the model focuses on post-schooling career decisions and earnings dynamics.

¹³Asymmetric information between employers and employees or between an incumbent employer and outside employers is also an interesting possibility (e.g., Schönberg, 2007; Pinkston, 2009), but it is very difficult to handle in a structural model of life-cycle earnings dynamics.

¹⁴The NLSY data has the Armed Forces Qualification Test (AFQT) score of almost all respondents (Appendix 1), but other test scores such as the SAT/ACT score or high school GPA are missing for many of them. I assume that the information carried over by the AFQT score was partly unused in the decision-making process of workers (and employers). This identifying assumption is similar to Altonji and Pierret (2001)'s and important for the testability of the model.

2.3 Production Function

Labor is the only input in this economy. A worker's log productivity (x_{jt}) at job j at period t is explained by general work experience (E_t), job tenure (T_{jt}), ability (θ) and job-specific match quality (ϵ_j) at job j :

$$x_{jt} = \gamma_0 + g(E_t, \theta) + s(T_{jt}, \theta) + \gamma_\theta \theta + \gamma_\epsilon \epsilon_j$$

where E_t is experience at the beginning of period t , T_{jt} is job tenure at job j at the beginning of period t , $g()$ is general human capital which is a function of general experience and (learning) ability, $s()$ is job-specific human capital which is a function of job tenure and (learning) ability, θ is ability, ϵ_j is the matching quality at job j .

Although I normalize the distributions of θ and ϵ , I allow the productivity effect of those variables to differ from one. I assume ability and learning ability are the same thing,¹⁵ but I allow the productivity effect in each case ($\gamma_\theta, \gamma_{g,\theta}, \gamma_{s,\theta}$) to be different from each other.

In the current version, I use a quadratic¹⁶ functional form for general and job-specific human capital accumulation (learning-by-doing). I restrict the peak age of productivity to be equal regardless of ability. That is,

$$g(E_t, \theta) = (1 + \gamma_{g,\theta}) (\gamma_1 E_t + \gamma_2 E_t^2)$$

$$s(T_{jt}, \theta) = (1 + \gamma_{s,\theta}) (\gamma_3 T_{jt} + \gamma_4 T_{jt}^2)$$

¹⁵ $\gamma_\theta \theta$ is often called “ability” (or productivity-enhancing effect of ability), while θ in $g(E_t, \theta)$ and $s(T_{jt}, \theta)$ is called “learning ability.”

¹⁶This is a popular but restrictive assumption (Murphy and Welch, 1990). Murphy and Welch (1990) suggest a three-parameter version setup of quartic equation, which is single-peaked.

In this case, the log productivity can be rewritten,

$$\begin{aligned}
x_{jt} &= \gamma_0 + \gamma_1 E_t + \gamma_2 E_t^2 + \gamma_3 T_{jt} + \gamma_4 T_{jt}^2 \\
&+ (\gamma_\theta + \gamma_{g,\theta}(\gamma_1 E_t + \gamma_2 E_t^2) + \gamma_{s,\theta}(\gamma_3 T_{jt} + \gamma_4 T_{jt}^2))\theta + \gamma_\epsilon \epsilon_j \\
&= \underbrace{h(E_t, T_{jt})}_{\text{certain component}} + \underbrace{\gamma_t(E_t, T_{jt})'\Theta}_{\text{uncertain component}}
\end{aligned}$$

The prior information on ability (before graduation) and job match quality (first job), Θ , follows a bivariate standard normal distribution.

$$\Theta = \begin{pmatrix} \theta \\ \epsilon_j \end{pmatrix} \sim N(\mathbf{0}, I).$$

True log productivity (x_{jt}) is not directly observed; however, a signal (y_{jt}) of the true log productivity is observed in each period. The signal at job j at the end of period t is,

$$\begin{aligned}
y_{jt} &= x_{jt} + \eta_t \\
&= h(E_t, T_{jt}) + \gamma_t(E_t, T_{jt})'\Theta + \eta_t
\end{aligned}$$

After the signal is observed, the beliefs on ability and job match quality are updated accordingly.

2.4 Multidimensional Learning about Two Unknown Objects

All individuals in the labor market are Bayesian-rational. After observing a new signal on productivity, each individual (and all employers) updates beliefs accordingly.

Beliefs are updated following a Bayesian-updating rule (Gaussian distribution, deterministic variance) as in Jovanovic (1984). Two differences are noteworthy. First, each individual now has to update beliefs on two (or possibly more) objects from only one signal. Put differently, the signals for each object are perfectly correlated. The correlation between the signals for each unknown object is a key differentiating characteristics of multi-dimensional

learning from multiple single-dimensional learning.¹⁷ Although the unknown objects are independent of each other, beliefs are correlated because signals are correlated. Also, job mobility is affected by unknown ability because the belief on job-match quality is affected by unknown ability through this correlation in signals.¹⁸ Second, the weights on the hidden objects are allowed to change over time.

When a new signal y_{jt} (at the end of period t) is a weighted sum of unknown objects and a pure noise, the belief-updating process can be formally expressed as follows (certain components are irrelevant so omitted). I first consider a case of a worker who has stayed at his first job without any spell of nonemployment ($1 \leq T_{jt} = E_t = t - 1$) and then consider a more general case with job changes and nonemployment spells ($1 \leq T_{jt} \leq E_t \leq t - 1$). Beliefs are updated at the beginning of period $t + 1$.

$$\text{All signals up to the beginning of period } t: y_{j\tau} = \gamma'_\tau \Theta + \eta_\tau, \tau = 1, \dots, T_{jt}(= E_t = t - 1)$$

where γ'_τ ($1 \times k$ vector) is the weights on unknown objects in the signal observed in period τ , η_τ is a pure Gaussian noise. $\eta_\tau \sim N(0, \sigma_\eta^2), 1 \leq \forall \tau \leq T_{jt}$.

¹⁷“Perfect” correlation is not the main thing. Positively correlated signals would bring qualitatively the same result. Note the difference from Sanders (2014), where signals for each object are independent. Multiple single-dimensional learning would not have the same result with multi-dimensional (and correlated) learning such as one presented here.

¹⁸This kind of negative selection into mobility can appear in other contexts – for example, Arcidiacono (2004) analyze college-major choice using a similar but much simpler setup. I do not know any paper that notices on negative selection into job mobility by ability. Interestingly, Neal (1999) and Pavan (2011)’s two-stage search structure looks similar to mine, but their results come from selection with a two-stage restriction (career change must involve job change) rather than information, and the implications are also different from here.

Prior beliefs: $\Theta \sim N(\mathbf{0}, I)$

$$\begin{aligned} \text{Beliefs (labor market entry): } \Sigma_1 &= \begin{pmatrix} \sigma_\zeta^2/(1 + \sigma_\zeta^2) & 0 \\ 0 & 0 \end{pmatrix}^{-1} \\ \Theta_1 &= \begin{pmatrix} \theta^{TS}/(1 + \sigma_\zeta^2) \\ 0 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \text{Beliefs (after } t - 1 \text{ signals): } \Sigma_t &= (\Sigma_1^{-1} + (\Gamma'\Gamma)/\sigma_\eta^2)^{-1} \\ \Theta_t &= \Sigma_t (\Sigma_1^{-1}\Theta_1 + (\Gamma'y)/\sigma_\eta^2) \end{aligned}$$

where $\Gamma = (\gamma'_1 \gamma'_2 \dots \gamma'_{t-1})'$ (a $(t - 1) \times k$ matrix of deterministic weights), $y = (y_1 \ y_2 \dots y_{t-1})'$,
 $((t - 1) \times 1$ vector)

If one more signal, y_{jt} , is realized,

$$\begin{aligned} \text{Beliefs (after } t \text{ signals): } \Sigma_{t+1} &= (\Sigma_t^{-1} + (\gamma_t\gamma'_t)/\sigma_\eta^2)^{-1} \\ \Theta_{t+1} &= \Sigma_{t+1} (\Sigma_t^{-1}\Theta_t + \gamma_t y_t / \sigma_\eta^2) \end{aligned} \tag{1}$$

This expression is in fact the same with one from Bayesian linear regression. Intuitively, we can interpret signals and deterministic weights as “data” which are given to a Bayesian-rational agent and the unknown objects as “parameters” which the agent wants to find.¹⁹

Also, the equation (1) shows that the beliefs follow a first-order Markov process. Θ_t and Σ_t (and Σ_1) are the sufficient statistics for all t signals observed so far. The belief updating process for each object, however, is not a first-order Markov process.

The following propositions are useful for describing the dynamic programming problem of each individual. How beliefs will change after a new signal is uncertain because the new signal is uncertain. The distribution of the beliefs at the beginning of current period is important for forming expectations to make decisions for current period.

¹⁹Since there are two (or more) unknown productivity factors – ability and job match quality – each individual takes another unknown component into account when he updates his belief on an object. Using matrix notation, we can easily see that this is in fact an idea of Bayesian linear regression. Appendix 2 explains more.

Proposition 1: (the distribution of the period t signal, at the beginning of period t , if $T_{jt} = E_t = t - 1$)

$$y_t | \Theta_t, \Sigma_t \sim N(\gamma'_t \Theta_t, \gamma'_t \Sigma_t \gamma_t + \sigma_\eta^2)$$

Proposition 2: (the distribution of the period $t + 1$ mean beliefs, at the beginning of period t , if $T_{jt} = E_t = t - 1$)

$$\Theta_{t+1} | \Theta_t, \Sigma_t \sim N(\Theta_t, (\Sigma_{t+1} \gamma_t / \sigma_\eta^2)(\gamma'_t \Sigma_t \gamma_t + \sigma_\eta^2)(\gamma'_t \Sigma_{t+1} / \sigma_\eta^2))$$

Although we have considered only the case of $T_{jt} = E_t = t - 1$, it is easy to relax this restriction. First, if experience is different from time in the labor market ($E_t < t - 1$), only experience (E_t) matters for information updating. No productivity signal is observed during non-employment periods, and the productivity weights (γ_t) also does not depend on t once E_t is controlled. Second, even if job tenure is different from experience ($T_{jt} < E_t$) with possibly many job changes, the above expressions are still applicable, with slight modification: we need to add the uncertainty of the belief on ability at current job entry as an additional state variable (See Appendix 2). Proposition 1 and 2 are still useful in that way.

Multidimensional learning has interesting implications for job mobility and wage growth. First, unlike general or specific skill accumulation which explains wage growth *either* across all jobs *or* within a job, this generalized learning process explains both types of wage growth.²⁰ In this setting, a portion of acquired information within a job, which was entirely job-specific in Jovanovic (1979), can be transferred to other jobs. This corresponds to some sorts of public evaluation on worker ability – for example, recommendation letters.

Second, there are two different kinds of job experimentations, which are related to experience and job tenure, respectively. The first kind of job experimentation is learning about ability. Workers need precise information about their ability to choose a right career path,

²⁰Of course, wage growth in multi-dimensional learning occurs because of selection – by ability, across jobs, and by job match quality, within a job.

but to attain this information they have to try many jobs first since they in most cases get the most information by sampling a new job.²¹ This generates job hazard decreasing in work experience. Workers move more frequently when their experience is lower, and they become stabilized with more work experience.²² The second kind is learning about job match, which is shown in Jovanovic (1979). In this case, workers move more when their job tenure is lower, and job hazard is decreasing in job tenure.

Third, ability can be negatively associated with job mobility under multidimensional learning structure. Initially similar individuals (conditional on test scores) receive different signals on their productivity over their career. Their beliefs on ability and job match quality eventually converge to the true values; however, there exists systematic biases in their beliefs, toward their initial beliefs, until the uncertainty about ability is completely resolved. Because of a biased belief on job match quality, which is negatively correlated with unknown ability, job moves can be sub-optimal – a good match can be destroyed because of underestimated match quality (or overestimated ability), whereas a bad match can survive for a long time. In other words, the job shopping process becomes heterogeneous, unlike Jovanovic (1984), because of unknown ability and multidimensional learning. This implication of negative selection into (job) mobility is partly noticed by the previous literature in another context such as college-major choice (e.g., Arcidiacono, 2004). This paper extends the implication to a more generalized situation where the negative selection into job mobility gradually disappears as people eventually learn about their ability. This is new and important for the identification. With additional source of information on ability such as the AFQT score, this implication is directly testable in the data.

²¹This point is clearly presented in an extreme case where $\sigma_\eta^2 = 0$ with time-invariant weights on Θ . In that case, staying at a job would not add any more knowledge on ability. One can learn only by moving to another job. If $\sigma_\eta^2 > 0$, people learn considerably in the initial phase of new employment, but they get less and less information as job tenure increases.

²²The weights on hidden objects are allowed to be time-varying and choice-dependent, which makes this learning process a dynamic experimentation similar to Miller (1984) or Antonovics and Golan (2012). Miller (1984) and Antonovics and Golan (2012) model occupational choices as altering both an occupational match and its degree of uncertainty. This paper incorporates both features by allowing the weights on hidden objects in signals to change depending on the timing of job mobility choices.

Fourth, ability can be increasingly positively associated with employment in this setting. The positive selection into employment by ability is predicted by other components than information such as positive leisure utility (or unemployment benefits) and search frictions – the probability of labor earnings going below a cutoff level is low for high ability workers. Multidimensional learning, in contrast, predicts a small but negative selection into employment by ability. Less able workers overestimate their ability due to the aforementioned bias; they are more likely to search and work because of misperceived ability level. As time passes, the former remains but the latter disappears – this brings an increasingly positive association between ability and employment.

Multidimensional learning, especially regarding the third point, offers an interesting behavioral interpretation to the coexistence of sub-optimally frequent job moves of low skilled workers (Gladden and Taber, 2009) and match-enhancing job moves of young workers (Topel and Ward, 1992). Gladden and Taber (2009) find low skilled workers change their job too frequently under lifetime income maximization hypothesis. At the same time, young workers benefit from job mobility because they can find a better match on average (Topel and Ward, 1992). Both kinds of job moves are well explained under multidimensional learning.

2.5 Search Technology and Partial Insurance

Each individual has to go through costly job search to get a new job offer in this frictional economy. Job search costs are the same on and off the job, c in utility unit. After job search, a new job offer arrives with a probability λ_R . The new job match (ϵ') is randomly drawn from a distribution of job match quality ($\epsilon' \sim N(0, \sigma_\epsilon^2)$), and there exists no additional information on the quality of the new match at job offer, other than the population distribution which is common knowledge.²³

Endogenous job search is important here, unlike Jovanovic (1984), because a strong negative information shock can lower perceived ability and make the worker directly go out

²³Allowing initial screening as in Jovanovic (1984) is not difficult but will not change any important result.

of labor force – this can be permanent if there are no other kinds of shocks. In contrast, homogeneous workers in Jovanovic (1984) always want to find a new job. They may choose to be unemployed, but it is only to search better.

Whereas being unemployed does not provide any advantage for searching for a new job in this setting, an unemployed status does have an advantage for searching for an old job. A recall offer from the most recent employer arrives with a probability of λ_R only for the unemployed who ever worked. This possibility creates an option value of being unemployed as in Fujita and Moscarini (2013).²⁴ Recall possibility is also potentially more important in this multidimensional learning setting because it could be the only way get back to work for some workers. This also matters more with job-specific human capital accumulation.²⁵

There also exist involuntary job separations. δ proportion out of currently working population undergoes an exogenous job destruction shock such as plant closing. Unemployment insurance partially insures against this shock. People who become involuntarily unemployed are eligible for unemployment benefits, which are 50 percent of previous earnings capped at \$400 (weekly) and \$10,000 (annually).²⁶ In addition, nonemployed workers including unemployed workers get welfare. I have a very simplified version of welfare: the benefits are \$5,000²⁷ (not means-tested); every individual who is not working is eligible for the benefits; and the take-up rate is assumed 100 percent. This number is in fact not important – I have another parameter, b , the leisure utility, which essentially does the same role – but the benefits guarantee individual income is above zero.

²⁴Fujita and Moscarini (2013) report 20 percent of the workers who were permanently separated from their previous employer eventually return back to the employer. Following Fujita and Moscarini (2013), I assume that the recall offer is only from the most recent employer. That is, if the worker takes a new job offer, it is not possible to go back to the previous employer.

²⁵The possibility of recall in a dynamic setting creates an incentive to stay unemployed, especially for workers with a good previous match or a high level of job-specific skill. If job-specific skill accumulation is faster for more able workers, they in this case have a stronger incentive to stay unemployed when their job has been destroyed by an exogenous shock, compared to other workers.

²⁶The cap of unemployment benefits is different from state to state, but the average is about \$400 per week. Also, the benefits are provided up to 26 weeks in the U. S. The income replacement rate is approximately 50 percent.

²⁷This number is borrowed from French and Jones (2011) – they find the average welfare benefits for households without a member over 65 is about 3,500 in 1998 dollars. Their number is converted into 2010 dollars using the CPI index.

2.6 Preference

Individual preference is time-separable, and the time discount factor, $\beta (< 1)$, is constant over time. Also, consumption and leisure are additively separable. The periodic utility from consumption follows a Constant Relative Risk Aversion (CRRA) function.

The periodic utility function is

$$U(c_t, s_t, d_t) = \frac{c_t^{1-\rho} - 1}{1-\rho} + b1[d_t = 0] + \xi_t(s_t) + \kappa\xi_t(d_t)$$

where s_t is an indicator of job search at period t , d_t is a work/job choice at period t ($d_t = 0$ if not working, $d_t = 1$ if working at the same job, $d_t = 2$ if working at a new job), $\xi_t(s_t)$ is a preference shock associated with job search, $\xi_t(d_t)$ is a preference shock to work decisions.

The relative risk aversion, ρ , is greater than or equal to 0. If $\rho = 0$, the CRRA function becomes a linear utility function. If $\rho = 1$, the utility function becomes a logarithm function. b denotes utility from leisure.

Each individual experiences sequential and transient preference shocks across job search (s_t) and work (d_t) alternatives ($\xi_t(s_t), \xi_t(d_t)$), which all follow the standard Type I Extreme Value (Gumbel) distribution: $\xi_t(s_t), \xi_t(d_t) \sim i.i.d. \text{ Gumbel}(0,1)$.²⁸ Allowing a sequential structure is equivalent to generalized nested logit choice structure (Kennan, 2015). Preference shocks are often for a technical purpose, to make choice probabilities non-trivial ones, but they play economically meaningful roles here: the relative value of old and new matches over leisure utility change exogenously.

2.7 Wage Determination

The equilibrium wages are set at the *ex ante* expected productivity level given all realized signals, experience and job tenure. That is, $w_{jt}^* = E[x_{jt} | \Theta_t, \Sigma_t, E_t, T_{jt}]$. As in Jovanovic (1979) or Jovanovic (1984), I assume firms have zero profits on average – that is, workers

²⁸CDF: $F(x) = e^{-e^{-x}}$.

fully internalize all benefits and costs into their decisions. This “competitive” wage scheme is equivalent to a special case of generalized Nash-bargaining with all surplus given to workers when job posting is costless (Moscarini, 2005).²⁹ A wage contract is signed right after job acceptance, and a new productivity signal, y_{jt} , is observed at the end of the period.

The econometrician observes a noisy measure of the latent earnings. The logarithm of the observed earnings (w_{jt}) is a sum of the logarithm of latent earnings (w_{jt}^*) and measurement error: $w_{jt} = w_{jt}^* + \nu_t$, where $\nu_t \sim i.i.d.N(0, \sigma_\nu^2)$.

2.8 Dynamic Programming Problem of Each Individual

Given the wage equation and learning process, the worker maximizes his lifetime utility, looking forward up to 60 years. The work decisions are only for the first $T(=40)$ years.

$$\begin{aligned} \max_{\{(s_\tau, d_\tau)_{\tau=t}^T, (c_t, A_{t+1})_{\tau=t}^{T+R}\}} & \sum_{\tau=t}^T \beta^{\tau-t} E[V_\tau(\Theta_\tau, \Sigma_\tau, A_\tau, E_\tau, T_{j\tau}, d_{\tau-1}^W) | \Theta_t, \Sigma_t, A_t, E_t, T_{jt}, d_{t-1}^W] \\ & + \sum_{\tau=T+1}^{T+R} \beta^{\tau-t} E[u(c_\tau) | \Theta_t, \Sigma_t, A_t, E_t, T_{jt}, d_{t-1}^W], 1 \leq \forall t \leq T \end{aligned}$$

The initial conditions other than the beliefs at period 1 are set all zeros: $A_1 = 0, E_1 = 0, T_{1,1} = 0, d_0^W = 0$.

In each period, the first choice is job search (right after a new productivity signal on the last period’s productivity). The value function at the beginning of period t is a $E\max$ function of choice specific values:

$$\begin{aligned} V_t(\Theta_t, \Sigma_t, A_t, E_t, T_{jt}, d_{t-1}^W) &= E\max \{V_t^S(\Theta_t, \Sigma_t, A_t, E_t, T_{jt}, d_{t-1}^W) + \xi_t^S - c, \\ &V_t^{NS}(\Theta_t, \Sigma_t, A_t, E_t, T_{jt}, d_{t-1}^W) + \xi_t^{NS}\} \end{aligned}$$

²⁹The existence of match-specific surplus under search frictions leaves a difficult issue of how to divide the surplus between a worker and employers. Following Jovanovic (1984), I choose to give all surplus to workers all the time. By doing this, the entire decision problem becomes a simpler one because there is no active role of employers. This cannot be fully satisfactory, so this setup should be considered as a starting point. Note that *ex ante* efficient separation is still achieved since workers fully internalize all benefits and costs into their decisions.

where V_t^S is the choice-specific value of job search, V_t^{NS} is the choice-specific value of no job search, $\xi_t^{S,NS}$ are transient preference shocks (known at search choice), c is the sum of job search costs, Θ_t is posterior beliefs (mean), Σ_t is posterior beliefs (covariance matrix), A_t is assets, E_t is work experience at the beginning of period t , T_{jt} is job tenure at job j at the beginning of period t , and d_{t-1}^W is employment status at the beginning of period t (0: not working, 1: working).

The sets of possible shocks and available options change depending on current employment status and job search. First, the option of a new job is available only after a (successful) job search. A success of job search can be stochastic with a probability of λ .³⁰ Second, only the currently employed ($d_{t-1}^W = 1$) experience an exogenous job destruction with a probability of δ . Third, only the currently non-employed ($d_{t-1}^W = 0$) can have a recall offer from their most recent employer with a probability of λ_R .

$$V_t^S(\Theta_t, \Sigma_t, A_t, E_t, T_{jt}, d_{t-1}^W = 1) = \lambda[(1 - \delta)\text{Emax}\{V_t^{Old}, V_t^{New}, V_t^U\} + \delta\text{Emax}\{V_t^{New}, V_t^{UB}\}] \\ + (1 - \lambda)[(1 - \delta)\text{Emax}\{V_t^{Old}, V_t^U\} + \delta V_t^{UB}]$$

$$V_t^{NS}(\Theta_t, \Sigma_t, A_t, E_t, T_{jt}, d_{t-1}^W = 1) = (1 - \delta)\text{Emax}\{V_t^{Old}, V_t^U\} + \delta V_t^{UB}$$

$$V_t^S(\Theta_t, \Sigma_t, A_t, E_t, T_{jt}, d_{t-1}^W = 0) = \lambda[\lambda_R\text{Emax}\{V_t^{Old}, V_t^{New}, V_t^U\} + (1 - \lambda_R)\text{Emax}\{V_t^{New}, V_t^U\}] \\ + (1 - \lambda)[\lambda_R\text{Emax}\{V_t^{Old}, V_t^U\} + (1 - \lambda_R)V_t^U]$$

$$V_t^{NS}(\Theta_t, \Sigma_t, A_t, E_t, T_{jt}, d_{t-1}^W = 0) = \lambda_R\text{Emax}\{V_t^{Old}, V_t^U\} + (1 - \lambda_R)V_t^U$$

where V_t^{New} is the choice-specific value of job change, V_t^{Old} is the choice-specific value of working with no job change, V_t^U is the choice-specific value of not working, V_t^{UB} is the

³⁰In a discrete-time setting with annual frequency, it must be rare for anyone not to receive any offer. This is one reason why I use job search costs.

choice-specific value of not working with unemployment benefits, λ is the probability of job offer arrival, δ is the probability of job destruction, and λ_R is the probability of recall offer arrival.

Individuals adjust consumption level after their employment status is fixed. Each work-choice specific value is given as follows:

$$\begin{aligned}
V_t^{New} &= \text{Max}_{A_{t+1}} \{u((1+r)A_t + W_{j't}^* - A_{t+1}) \\
&\quad + \beta E_{\Theta_{t+1}|\Theta_t, \Sigma_t, move}[V_{t+1}(\Theta_{t+1}, \Sigma_{t+1}, A_{t+1}, E_t + 1, 0, d_t^W = 1)]\} + \kappa \xi^{New} \\
V_t^{Old} &= \text{Max}_{A_{t+1}} \{u((1+r)A_t + W_{jt}^* - A_{t+1}) \\
&\quad + \beta E_{\Theta_{t+1}|\Theta_t, \Sigma_t, stay}[V_{t+1}(\Theta_{t+1}, \Sigma_{t+1}, A_{t+1}, E_t + 1, T_{jt} + 1, d_t^W = 1)]\} + \kappa \xi^{Old} \\
V_t^U &= \text{Max}_{A_{t+1}} \{u((1+r)A_t + 5,000 - A_{t+1}) + b \\
&\quad + \beta V_{t+1}(\Theta_t, \Sigma_t, A_{t+1}, E_t, T_{jt}, d_t^W = 0)\} + \kappa \xi^U \\
V_t^{UB} &= \text{Max}_{A_{t+1}} \{u((1+r)A_t + 5,000 + UI - A_{t+1}) + b \\
&\quad + \beta V_{t+1}(\Theta_t, \Sigma_t, A_{t+1}, E_t, T_{jt}, d_t^W = 0)\} + \kappa \xi^{UB}
\end{aligned}$$

where $W_{j't}^*$ is (latent) labor earnings at a new job j , W_{jt}^* is (latent) labor earnings at current job j , b is leisure utility, UI is unemployment benefits (only for exogenous job separations), r is real interest rate, ξ 's are transient preference shocks (realized right before work choice), and κ is a scale parameter.

The transient preference shocks ($\xi_t^{S,NS}, \xi_t^{New,Old,U,UB}$) are sequential and follow i.i.d. Type I extreme value distribution. Kennan (2015) shows this sequential shock structure has the same choice probabilities with those under the Nested Logit structure but this structure is more general in terms of the scale parameter (κ). Preference shocks pick up unexplained variation in choices and, more importantly, make the choice probabilities non-zero. This guarantees there always exists some people who change jobs (or employment status) volun-

tarily.

2.9 Solution Concepts

Although all state variables are obviously relevant for work decisions, I focus on the unobserved individual beliefs $(\theta_t, \epsilon_{jt})$. Other variables such as the uncertainty of the beliefs (Σ_t) , experience, job tenure, and assets are fixed. I here consider the case of $\rho = 0$ (linear utility) and the case of $\rho = 1$ (log utility) without savings. This section does not provide analytical solutions; it only provides at best a reasonable sketch of the solutions, which are numerical approximated.

First of all, job change choice can be characterized by a cutoff strategy in perceived job match quality ($\epsilon_{jt} < \epsilon_{jt}^*$). The cutoff value (ϵ_{jt}^*) exists within a reasonable range of parameters. The choice-specific value functions for staying and moving ($V^{Old}(\theta_t, \epsilon_{jt}; \dots)$, $V^{New}(\theta_t, 0; \dots)$) are multiplicatively (or additively) separable over perceived ability and job match quality. After crossing out a common component, $\exp((1 - \rho)\theta_t)$, only one of the value functions changes – monotonically increasing in perceived job match quality. The cutoff value exists unless parameter values are extreme. This feature of a cutoff strategy is shared across many other optimal stopping problems (e.g., Jovanovic, 1979; Neal, 1999).

Then, employment status choice is also characterized by a cutoff strategy using both perceived values of ability and job match quality. This strategy depends on the previously-found cutoff value of job change. If the current belief on job match quality is below the cutoff level ($\epsilon_{jt} < \epsilon_{jt}^*$), there exists a cutoff value ($\theta_t < \theta_t^*(\epsilon_{jt}^*)$) below which the person does not take up a new offer, regardless of the perceived job match quality. If the current perceived job match quality exceeds the cutoff level ($\epsilon_{jt} > \epsilon_{jt}^*$), there exist cutoff values as a function of perceived job match quality ($\theta_t < \theta_t^*(\epsilon_{jt})$), which must be decreasing in ϵ_{jt} . Interestingly, this is conceptually similar to Neal (1999)'s solution to a two-stage search problem although the problems are totally different.

Finally, job search under various employment-related shocks is also characterized by a

cutoff strategy. The basic principle is that those who want to work at a new job choose to search, whether they are employed or unemployed. The two dimensional cutoffs for job search ($\epsilon_{jt} < \epsilon_{jt}^+(\theta_t), \theta_t < \theta_t^+(\epsilon_{jt})$) are basically similar to the aforementioned cutoffs for job-movers ($\epsilon_{jt} < \epsilon_{jt}^*, \theta_t < \theta_t^*(\epsilon_{jt}^*)$). People whose value is the highest when they are not working would never search unless there are preference shocks – put differently, the positive selection into employment is reinforced by the existence of job search choice. Also, some people whose value is the highest when they stay at their current job might search if their perceived job match quality is only marginally higher than the job change cutoff (ϵ_{jt}^*) and their perceived ability is high – this is an insurance motive against job destruction. This incentive is small, however, and arises only because of the assumption on choice-shock sequence, so it is not very interesting. A more interesting feature of these search cutoffs is how they move with the presence of search costs and employment-related shocks. With search costs, some people who want a new job give up job search if their perceived ability is only marginally higher than the work cutoff $\theta_t^*(\epsilon_{jt})$ or if their perceived job match quality is only marginally higher than the job change cutoff. In general, it is likely that the positive selection into employment would get stronger with more costly job search.

Solving the finite horizon Bellman equations in the previous subsection is easily done by backward recursion. The expected values over the predicted distribution of the next period’s beliefs (or evaluations) are calculated by simple averages over equally-probable support points. This method offers the best approximation given a fixed number of grid points (Kennan, 2006).

2.10 A Note on Deterministic Covariance Matrices (Σ_t)

The beliefs system follows a first-order Markov process, which makes it easy to solve the aforementioned Bellman equations. The size of state space in that case grows linearly not exponentially, so computational burden should not be heavy, in principle.

The current version, however, does not fully take such advantage of this Markov structure.

It is because of the covariance matrices (Σ_t). It is acceptable to (coarsely) discretize posterior mean space (Θ_t) and find the solutions given exact Σ_t ; however, doing the same thing given (coarsely) discretized Σ_t can bring non-negligible approximation biases into the solutions.

Because of this possibility, I calculate exact Σ_t in this version, using full job change history (but not signal history) of each individual. Given each experience level E_t , not only the number of job changes but also the timing of each job change matter. If the weights were time-invariant, only the number of job changes would matter. This is equivalent to think about all job change possibilities for each experience level, and the size of state space grows exponentially in time in the labor market ($= K_t \times \sum_{E_t=1}^t 2^{E_t}$, K_t is fixed or only linearly growing in time) although the speed is not too fast during the first 15 periods.

To keep the state space manageable in this case, I additionally assume that the worker receives an exact signal on their ability after spending $T_0 (= 15)$ years in the labor market. Hence, the worker essentially switches from multidimensional learning to single-dimensional learning about job match quality after T_0 years of labor market experience. The posterior beliefs after this special signal is obtained by finding the limit of the beliefs as the noise in the signal approaches zero. That number is an arbitrary choice, but the number will not matter if the speed of learning is quite fast. I have tested other numbers up to 22, but found no evidence of significant changes in estimates.

3 Data

3.1 The National Longitudinal Study of Youth 1979 (NLSY79)

The data used in this analysis is the National Longitudinal Study of Youth 1979 (NLSY79), Round 1-25. The NLSY79 surveys a nationally representative sample of young men and women, including minority, poor and military oversamples, who were 14-22 years old in 1979. The data was collected annually from 1979 to 1993 and then biannually from 1994.

The NLSY data offers an opportunity to study how ability interacts with career decisions of young people in the labor market. First, almost all respondents took the Armed Services Vocational Aptitude Battery (ASVAB) at the very beginning of the survey.³¹ In addition, several scales of non-cognitive traits near labor market entry are available. For example, Rosenberg’s Self-Esteem Scale and Rotter’s Internal-External Locus of Control Scale in 1980 are available – these scores are recently used as measures of non-cognitive ability in several studies (e.g., Heckman et al., 2006).

Second, the NLSY data provides an event history of each respondent’s education and work decisions. For example, the NLSY79 has kept weekly arrays of all jobs ever held for all respondents, from which job changes can be detected very accurately. Even though someone have missed several rounds of interviews, the respondent was asked about the missing work history at a later interview. More details on the construction of ability measures and education/work history are in Appendix 1.

To avoid any complication from oversampling, gender and racial issues, I focus on the Cross-sectional White Male sample from the NLSY 1979 data (2,236). I remove all records from the respondents whose highest grade completed was less than 12 years at their first entry into the labor market or from those with a G.E.D. (-425) because high school dropouts and G.E.D.’s are reported to show very different behavioral patterns from other educational groups. The respondents whose post-schooling work history cannot be correctly constructed are also dropped from the sample (-24) – this is either because their first graduation year is not specified from their records or because their first graduation year is before 1975. I further drop individuals if they have ever joined the Military Services (-263) throughout the surveys. This is very conventional but has a specific meaning in this paper, which is related to the assumption on the AFQT score. All records from individuals are dropped if any of five

³¹The ASVAB was administered to 11,914 NLS respondents (94 percent) during July through October of 1980 for the purpose of establishing a new national norm of the test (NLSY Attachment 106). ASVAB score are used to determine eligibility and assignment qualifications for specific military jobs for new enlistees, and the AFQT score, the sum of four subsection scores (word knowledge, paragraph comprehension, arithmetic reasoning and numeric operations), is a general measure of trainability and a primary criteria of enlisted eligibility for the Armed Forces (NLSY Attachment 106).

ability measures – AFQT, ASVAB Verbal, ASVAB Math, Rotter’s Internal-External Scale and Rosenberg’s Self-Esteem Scale(1980) – are missing (-145), or if any parent’s education is missing (-61). Finally, I consider weekly earnings less than \$1 or greater than \$10,000 as missing to eliminate potentially influential data points.

The final sample (all education levels) has 1,294 individuals and 27,218 person-year observations after high school graduation (25,257 observations after the first labor market entry). Table 1 has the descriptive statistics of the final sample. Among them, 634 individuals entered the labor market as a high school graduate. Among the initially high school graduates, 131 finish at least some post-secondary education before the last interview.

I use the final sample of all education levels for the following discussion of observed patterns in the data – for the comparison purpose with the previous literature. In the estimation of the structural model, I use only the sample of high school graduates who never attained extra college education (503 individuals) because the model does not have educational decisions. Restricting the sample is also a (crude) way of controlling for occupational heterogeneity.

Table 1: Descriptive Statistics

Variable	Obs.	Mean	SD	Min	Max
Birth Year	27,218	60.53	2.19	57	65
Age at Interview	27,218	31.44	9.39	16	56
Father's Education	27,218	13.04	3.16	0	20
Mother's Education	27,218	12.54	2.23	0	20
Siblings	27,218	2.69	1.69	0	15
AFQT (age-adjusted)	27,218	0.03	0.98	-4.44	1.75
ASVAB Math (age-adjusted)	27,218	0.02	0.99	-3.22	2.77
ASVAB Verbal (age-adjusted)	27,218	0.02	0.98	-2.85	3.02
Rosenberg's Scale (age-adjusted)	27,218	-0.01	1.00	-2.89	2.65
Rotter's Scale (age-adjusted)	27,218	0.01	1.00	-2.28	3.48
College Enrollment (May 1)	27,218	0.14	0.35	0	1
Highest Grade Completed	27,218	13.98	2.24	12	20
H.G.C. at the First Graduation	27,218	13.79	2.13	12	20
H.G.C. at Last Interview	27,218	14.44	2.51	12	20
Potential Experience	25,264	12.95	9.19	0.04	38.5
Work Experience	25,264	11.60	8.69	0	38.1
Working	25,258	0.91	0.28	0	1
New Employer (within 52 weeks)	25,258	0.26	0.44	0	1
Job Turnover (within 52 weeks)	25,258	0.21	0.40	0	1
Employer Tenure	26,050	4.99	5.89	0.02	35.5
Weekly Earnings (Main Job, \$, 2010)	24,793	996.1	881.2	1.17	9,992
Annual Labor Earnings (\$, 2010)	25,560	49,049	49,915	0	326,609
Net Assets (\$, 2010)	15,627	184,052	415,343	-745,165	3,506,059
Married	27,218	0.51	0.50	0	1
Residential Area (Census)	27,218	2.36	1.01	0	4
SMSA	27,218	0.47	0.50	0	1
Urban	27,218	0.73	0.45	0	1
Unemployment Spell	27,218	5.66	18.73	0	1451

3.2 Observed Correlation Patterns from Linear Regressions

In this subsection, I report observed correlation patterns between an ability measure, the Armed Forces Qualification Test (AFQT) score, and various labor market decisions/outcomes. This provides the motivation why we need to introduce individual heterogeneity in job change and employment into the model of career decisions and earnings over the life cycle. Moreover, this intuitively shows how the structural model in the previous section can be identified from the data. Many correlation patterns presented here are already introduced in the literature, especially in the context of employer learning (e.g., Farber and Gibbons, 1996; Altonji and Pierret, 2001; Lange, 2007; Schönberg, 2007; Pinkston, 2009). I add more patterns unnoticed by previous literature and discuss the limitations of previous interpretations.

The Armed Forces Qualification Test (AFQT) score is a well-established measure of ability in the literature. It is a special³² measure of ability which is highly correlated with both the entry level and the growth rate of individual earnings (Table 2, Column 1). The AFQT score and earnings growth over potential experience remains even after education level is controlled for (Table 2, Column 2) as already noticed by the previous studies (e.g., Farber and Gibbons, 1996; Altonji and Pierret, 2001). The AFQT score and earnings growth over potential experience are also positively correlative among high school graduates (Table 2, Column 3).³³ In addition, the positive correlation between the AFQT score and earnings is increasing but concave in potential experience (Table 2, Column 5) as noticed by Lange (2007).³⁴ That is, the AFQT score is associated with both higher earnings level and faster

³²Other ability measures, such as Rotter’s Locus of Control scale and Rosenberg’s Self-Esteem Scale, are strongly positively correlated with earnings level but not significantly correlated with earnings growth. These self-reported measure are obviously measures of *known* ability.

³³Unlike Arcidiacono et al. (2010), I find strong positive correlation between the AFQT score and earnings growth over potential experience among college graduates as well. The differences mainly arise from the empirical definition of “college graduates”. Arcidiacono et al. (2010)’s definition is more conventional – exactly 16 years of education at the final survey. My definition is less conventional – the individuals who entered the labor market after 16 years of education with no extra education. That is, I remove all records from the individuals who worked first and then finished college. This is because (extra) educational decision is highly correlated with ability. Of course, the current way of controlling for such bias is not fully satisfactory.

³⁴Lange (2007) uses this pattern for estimating the speed of employer learning.

earnings growth but the difference in earnings growth disappear over time.

Ability matters for entry level earnings as well as subsequent earnings growth, which is a well-established empirical pattern in the literature. Then, what is the mechanism(s)? Those correlation patterns are well matched with an (employer) learning hypothesis if the AFQT score has some information on unknown ability. For example, Altonji and Pierret (2001) assume that the AFQT score was not available to employers.³⁵ After then, they interpret the positive correlation between the AFQT score and earnings-experience interaction, along with the negative correlation between the AFQT score and education-experience interaction, as evidence of employer learning. This implies that the positive correlation between the AFQT score and earnings growth over experience, conditional on high school graduates, is also from employer learning. This implication (not the Altonji and Pierret (2001)'s original test), however, can be challenged by alternative explanations with similar predictions. For example, if more able workers accumulate general or job-specific skills faster than others, we will observe the same patterns (learning ability and differential skill production). Moreover, career decisions such as job change, movements into and out of nonemployment, can be all affected by ability.

Job change is important for wage growth (Topel and Ward, 1992), and ability may affect earnings through job changes. Table 3 shows that the AFQT score is negatively correlated with job change, and the negative correlation disappears over the life cycle (Column 1). A job change is empirically defined by working at interview date with job tenure less than 52 weeks, conditional on working 52 weeks before. This disappearing negative correlation over the life cycle is true even after education is controlled (Table 3, Column 2-4). This is not likely totally driven by negative selection into unemployment. When we compare job-to-job changes versus all job changes, the correlation between the AFQT score and job-to-job change (vs. all job changes) is increasing positive (Table 4, column 1), but the correlation becomes small and statistically not different from zero conditional on high school graduates (Table 4,

³⁵Lange (2007) has a detailed discussion on this assumption.

Table 2: AFQT Score and Labor Earnings

Dependent Variable: Log Real Weekly Earnings					
	(1)	(2)	(3)	(4)	(5)
	All	All	HSG	CLG	All
AFQT	0.092*** (0.008)	0.039*** (0.009)	0.041*** (0.011)	0.043 (0.025)	-0.007 (0.013)
AFQT x P. Exper. /10	0.035*** (0.005)	0.035*** (0.006)	0.028*** (0.007)	0.067*** (0.017)	0.127*** (0.018)
AFQT x P. Exper. ² /100					-0.029*** (0.005)
Educ.		0.074*** (0.005)			0.075*** (0.005)
Educ. x P. Exper. /10		-0.001 (0.003)			-0.002 (0.003)
Observations	23,319	23,319	10,640	4,960	23,319

From the NLSY79, Round 1-25. Cross-Sectional, White Male Sample.

a. AFQT score is standardized within each birth year group. Mean 0 and SD 1.

b. All specifications control for a cubic in potential experience, year fixed effects, regional fixed effects, parents' education and the number of siblings.

Standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

column 3). Job-to-job change here is defined by a job change without any non-employment spells within 52 weeks.

Table 3: AFQT Score and Job Change

Dependent Variable: Job Change (last 52 weeks) Work				
	(1)	(2)	(3)	(4)
	All	All	HSG	CLG
AFQT	-0.032*** (0.005)	-0.029*** (0.006)	-0.027*** (0.007)	-0.034* (0.015)
AFQT x P. Exper. /10	0.013*** (0.003)	0.011** (0.003)	0.009* (0.004)	0.012 (0.010)
Educ.		-0.004 (0.003)		
Educ. x P. Exper. /10		0.002 (0.002)		
Observations	23,028	23,028	10,332	4,944

From the NLSY79, Round 1-25. Cross-Sectional, White Male Sample.

a. AFQT score is standardized within each birth year group. Mean 0, SD 1.

b. All specifications control for a cubic in potential experience, year fixed effects, regional fixed effects, parents' education, the number of siblings.

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Since job changes includes both job-to-job changes (E-E') and job changes with a gap (E-N-E'), the correlation between the ability measure and (non)employment also needs to be

Table 4: AFQT Score and Job-to-Job Transitions

Dependent Variable: Job-to-Job Job Change (last 52 weeks)				
	(1)	(2)	(3)	(4)
	All	All	HSG	CLG
AFQT	-0.005	0.001	0.014	-0.022
	(0.009)	(0.010)	(0.013)	(0.028)
AFQT x P. Exper. /10	0.025**	0.018	0.010	-0.004
	(0.009)	(0.010)	(0.012)	(0.032)
Educ.		-0.007		
		(0.006)		
Educ. x P. Exper. /10		0.008		
		(0.005)		
Observations	6,675	6,675	3,048	1,411

From the NLSY79, Round 1-25. Cross-Sectional, White Male Sample.

a. AFQT score is standardized within each birth year group. Mean 0, SD 1.

b. All specifications control for a cubic in potential experience, year fixed effects, regional fixed effects, parents' education, the number of siblings.

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

examined. Previous studies report that job loss has long-lasting negative impacts on labor earnings. The positive selection into employment gets clearer over time (Table 5, Column 1-2). Less able workers are less and less likely to work as time in the labor market goes on, especially those among high school graduates (Table 5, Column 3). Among college graduates, such tendency is less clear (Table 5, Column 4). These patterns clearly show that employment status is also affected by ability. Also, this selection is likely to arise at nonemployment to employment (N-to-E) transitions rather than at employment to nonemployment (E-to-N) transitions.³⁶

Job change and employment status decisions are heterogeneous by an ability measure, at least within the sample used here.³⁷ These patterns show that we need a more flexible model of life-cycle career decisions. Allowing individual heterogeneity in career decisions is

³⁶This result is not fully tested yet. A crude analysis using employment status changes between interview dates show that N-E transition is significantly different by ability (especially at later career), while E-N transition is not that different by ability. This result needs to be tested using fixed and shorter time intervals.

³⁷This result needs further robustness checks because these correlation patterns are not fully established in the literature. For example, Schönberg (2007) finds the negative correlation between the AFQT score and job mobility among college graduates, but not among high school graduates. Her definition of education groups are very different from mine: high school graduates includes high school dropouts and G.E.D's; college graduates include those who attained additional education after first labor market entry.

Table 5: AFQT Score and Employment

Dependent Variable: Being employed (interview date)				
	(1)	(2)	(3)	(4)
	All	All	HSG	CLG
AFQT	0.009** (0.003)	0.007* (0.004)	0.013** (0.005)	0.002 (0.009)
AFQT x P. Exper. /10	0.007*** (0.002)	0.006** (0.002)	0.008* (0.003)	0.009 (0.006)
Educ.		0.003 (0.002)		
Educ. x P. Exper. /10		0.000 (0.001)		
Observations	25,257	25,257	11,465	5,324

From the NLSY79, Round 1-25. Cross-Sectional, White Male Sample.

a. AFQT score is standardized within each birth year group. Mean 0, SD 1.

b. All specifications control for a cubic in potential experience, year fixed effects, regional fixed effects, parents' education, the number of siblings.

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

important *per se*; also, it can produce useful insights to understand extremely heterogeneous earnings dynamics even if we cannot model all heterogeneity.

One possible way of modeling these heterogeneous job mobility and employment is to incorporate multi-dimensional learning about ability and job match quality presented in the model section. Information shocks on *perceived* productivity affect career decisions as in Jovanovic (1984). Information shocks on *misperceived* productivity explain the disappearing negative selection into job change.³⁸ The positive selection into employment is mainly explained by positive leisure utility and search frictions – especially, job search costs reinforce the positive selection into employment by raising the huddle at N-to-E transitions. The increasing positive selection, however, is explained by information shocks – less able workers work *more than they would do under full information* because of their incorrectly high belief on ability, and the negative selection into employment due to misperceived ability disappears over time.

³⁸The observed correlation patterns between the AFQT score and job-to-job change are interpreted as evidence of asymmetric employer learning in the literature (e.g., Schönberg, 2007; Pinkston, 2009). This is not coincidence, and the difference between their asymmetric model and my symmetric model is probably smaller than it appears. For example, if I interpret a job-specific match quality, which is a *productivity* inapplicable to outside employers in this paper, as an *information on ability* unavailable to outside employers, a multidimensional learning story is then interpreted, very roughly, as a kind of asymmetric learning story.

In addition, putting all separate pictures together can give us a better idea about how to identify observationally-equivalent mechanisms at one dimension. For example, a skill accumulation story (learning ability) and an information-updating story (learning about ability) predict the same thing on earnings growth – the positive association between ability and earnings growth over experience. The extended versions of the two stories, have different predictions on job mobility: a combination of general and job-specific skill accumulation (or more generally, occupation-specific skill accumulation) predicts that more able workers are less likely to move and this difference in job hazard between more able and less able workers increases over time; multidimensional learning about ability and job match quality also predicts more able workers are less likely to move, but the difference in job hazard decreases over time. In other words, the coefficient on (unknown) ability-experience interaction term in a (voluntary) job hazard equation can provide a source of identification between skill- and information-based mechanisms. This is important for the identification of the structural model presented in the previous section.

3.3 Identifying Assumption: A Measure of Unknown Ability

One last thing that should be mentioned in this data section is the measure of *unknown* ability. The identification argument mentioned above depends on the assumption that the econometrician has a measure of *unknown* ability. In this paper, I make an identifying assumption which is about the nature of the AFQT score. This modifies Altonji and Pierret (2001)’s assumption on the AFQT score in two ways: first, there exists a test score (θ^{TS}), a measure of ability, that is available to all workers and employers but not for the econometrician (for example, the SAT/ACT score or high school GPA); second, the AFQT score is another independent measure of ability that is available to the econometrician but not available to employers and unnoticed by (non-military) workers. The first one is certainly relaxing Altonji and Pierret (2001)’s, but the second one is probably stronger. The AFQT score was in fact sent to the NLSY respondents although not directly to employers. I assume

that non-military workers do not know how to interpret their AFQT score or just ignore it because the score is basically for the Military Services and they have another good measure of ability – I believe this is reasonable in reality. Also, this is why I restrict the sample for only civilians without any military experience throughout all surveys.

4 Identification and Estimation

4.1 Identification

The econometrician observes (log) labor earnings (w_{ijt}) and employment and job change decisions (d_{it}) of each individual over periods $t = 2, \dots, 30$ in a representative sample of the population. In addition, the econometrician observes a measure of ability (θ_i^{AFQT}). Other variables such as job match quality (ϵ_{ij}), individual beliefs on ability and job match quality ($\theta_{it}, \epsilon_{ijt}$), job search (s_{it}) or savings (A_{it}) are not observed.³⁹

As the number of model parameters are not many, it is not very difficult to identify the structural model from the observed data. Nonetheless, not all parameters can be identified, so some are fixed: for example, relative risk-aversion(=0 or 1), time discounting factor (=0.95) and the real interest rates (=0.05)⁴⁰, the (annual) probability of a recall offer conditional on unemployed (=0.2)⁴¹ are fixed; the variances of measurement error in test scores ($\sigma_{\zeta}^2, \sigma_{\zeta'}^2$) are normalized at 1, which is equal to the normalized variance of ability in population.⁴² The number of parameters to be identified after this is only 16.

There can be more than one way to construct the identification argument. Here I start

³⁹The NLSY79 data has some variables on assets, but the variables are not available for all periods, especially in early career. In principle, the risk aversion parameter can be estimated from saving patterns; instead, I fix the parameter value and compare the results across some possible cases.

⁴⁰ $0.95 \times 1/1.05 = 0.9975 \approx 1$.

⁴¹Fujita and Moscarini (2013) report the probability of returning back to a previous employer after a permanent separation is about 20%. This number may be different from the recall offer arrival rate, but it must be very close to the arrival rate provided that returning to the previous employer must be the best option for workers who are currently unemployed.

⁴²This normalization is innocuous because the productivity effect (γ_{θ}) is adjusted corresponding to the population variance of perceived ability.

from the log labor earnings (or log productivity) equation. In the model, $w_{ijt} = (1 + \gamma_{g,\theta}\theta_{it})g(E_{it}) + (1 + \gamma_{s,\theta}\theta_{it})s(T_{ijt}) + \gamma_{\theta}\theta_{it} + \gamma_{\epsilon}\epsilon_{ijt} + \nu_t$. (ν_t is an measurement error.) If job change were a random choice and every individual worked, the $g()$ and $s()$ functions could be separately identified from within-job wage growth of job-stayers ($E_i[w_{ijt} - w_{ijt-1}]$) and the first-year wages of job-movers ($E_i[w_{ij't}, \forall j' \neq j(t-1)]$), following the strategy of Topel (1991) (or Altonji and Shakotko (1987)) to identify $s()$. If job change and individual employment are endogenous choices, the expected change in the unobserved productivity within a job spell will not be zero due to selections.

The structural model describes the choices, and choice probabilities are important for identification. Observed choice probabilities can help us to identify some productivity/learning parameters ($\gamma_{\theta}, \gamma_{\epsilon}, \sigma_{\eta}^2$) along with search technology (δ, c) and preference (b) parameters. Over-time variation in job change probability ($Pr[1(d_{it} = emp, move)|1(d_{i,t-1} = emp)]$), for example, identifies signal-to-noise ratios. Employment-to-employment (E-to-E) transition ($Pr[1(d_{it} = emp)|1(d_{i,t-1} = emp)]$) identifies job destruction probability δ . Nonemployment-to-employment (N-to-E) transition ($Pr[1(d_{it} = emp)|1(d_{i,t-1} = nonemp)]$), along with E-to-E transition, identifies the leisure utility b . Although we do not observe job search choice probabilities, the difference between E-to-N and N-to-E transitions, especially by the AFQT score, reveals information on the search technology (c)

Some parameters have qualitatively the same predictions on both log earnings and choice probabilities, which makes it difficult to identify them separately. For example, both learning about job-specific match quality ($E_i[\epsilon_{ijt} - \epsilon_{ijt-1}]$) and job specific human capital accumulation ($s()$) predict increasing log earnings and decreasing job mobility in time on the job (and in the labor market). The over-time variation in job change probability has useful information on both $s()$ and signal-to-noise ratios in the belief-updating process, but they cannot be separately identified without additional source of identification.

With the additional identifying assumption on the AFQT score, I focus on how the correlation between the AFQT score and job change probability changes over time. The

two mechanisms have different predictions on this correlation as discussed in the previous section, so I can separately identify each mechanism.

Once this issue is solved, other parameters are easily identified. For example, from between-jobs wage growth, the productivity effect of job match quality, γ_ϵ , is identified. In addition to the conditional means in log earnings, conditional variances in log earnings also provide useful information for the variance of wage measurement error (σ_v^2) and the uncertainty in the belief on ability.

4.2 Estimation: Indirect Inference

The estimation method is indirect inference, which is broadly a kind of the Method of Simulated Moments (MSM). Indirect inference is characterized by the use of an auxiliary model. The indirect inference estimator is the minimizer of the (optimally-weighted) distance between auxiliary models estimated from real data and simulated data. The estimator is consistent and as efficient as the Maximum Likelihood (ML) estimator if the auxiliary model contains all information of the true model (Gourieroux et al., 1993).

$$\hat{\theta}_{II} = \underset{\theta}{\operatorname{argmin}} \{[\hat{\beta} - \hat{\beta}^s(\theta)]'W[\hat{\beta} - \hat{\beta}^s(\theta)]\}$$

$$\hat{\beta}^s(\theta) = \frac{1}{H} \sum_{h=1}^H \hat{\beta}_h^s(\theta)$$

where W is the weighting matrix, $\hat{\beta}$ is the estimated auxiliary model parameters from the data, $\hat{\beta}^s(\theta)$ is the estimated auxiliary model parameters from the simulated data given θ .

The optimal distance between the estimated auxiliary models is calculated by using the inverse of the data moment covariance matrix as the weighting matrix (W). The covariance matrix is obtained by a block bootstrapping method ($b = 1,000$).

The auxiliary model in this paper is a system of seven regression equations that describes log earnings and log earnings squared for job-movers and job-stayers, respectively, and job change (E-to-E'), employment-to-employment (E-to-E) and nonemployment-to-employment

(N-to-E) transition probabilities. All equations have the same explanatory variables: a constant, a quartic in potential experience, the AFQT score and its interaction terms with the quartic in potential experience, and lagged log earnings.

All seven equations are about changes, so it is difficult to match the levels. Hence, I use additional moments describing initial earnings and employment: cross-sectional mean and variance of initial earnings and initial employment rate at age 19. The regression coefficients from the seven equations and the additional data moments are in total $80(= 7 \times 11 + 3)$.

The resemblance of the auxiliary models between mine and Altonji et al. (2013)'s is quite striking because my model and identification argument are totally different from Altonji et al. (2013)'s. Following Altonji et al. (2013), I use the same set of control variables for all equations because this is in fact a Seemingly Unrelated Regression (SUR) system. I do not implement the GLS estimation because it is computationally too costly to do in this inner loop; my estimates are consistent but might be less efficient.

The auxiliary model is estimated from both real and simulated data. Since the structural model does not have any yearly or regional change, I use predicted log earnings (obtained by first regressing log earnings on a quartic in regional and yearly fixed effects – the base is an urban area, an SMSA, Northeast region in Census and year 1979 – and eliminating all yearly and regional fixed effects) in the estimation of the auxiliary model from real data. Also, I make simulated data equal to real data in all other respects. Whenever an observation is missing in real data, I consider the corresponding observation in simulated data as missing. In particular, the NLSY79 surveys are collected annually until 1994 and biannually afterwards; the same structure is imposed on the simulated sample.

The auxiliary model includes several discrete variables, which makes auxiliary model parameters (regression coefficients) discontinuous in true model parameters. To use a gradient-based minimization algorithm in the estimation procedure, the discrete variables need to be smoothed. Although the auxiliary model I use in this paper is similar to Altonji et al. (2013)'s, the structural model I use heavily depends on past decisions including employment

status and savings.⁴³ I introduce importance sampling weights to the auxiliary model in this indirect inference procedure, following Sauer and Taber (2013).

A rough description of indirect inference with importance sampling weights (Sauer and Taber, 2013) is as follows:

1. Given initial parameters (θ), simulate a data set (Y).
2. Evaluate the likelihood of the simulated data (Y) under a new set of parameters (θ').
3. Estimate a new auxiliary model using the calculated likelihood ratio as sampling weights.
4. By using a gradient-based minimization technique, find a new set of parameters ($\theta'' \in \{\theta'\}$) that minimizes the optimal distance between two auxiliary models – one from the simulated data (Y) using weights and the other from real data.
5. Update the initial set of parameters (θ) with the new minimizer (θ'') and simulate a new data set (Y'').
6. Repeat 2-5 until the distance function shows a sign of convergence.

Although I need to specify a likelihood function in this case, the function in this setup is much easier to write down because I can directly use unobserved variables as noted by Sauer and Taber (2013) – I do not need the integration over unobserved variable, which is the main difficulty in using the Maximum Likelihood (ML) approach. In the evaluation of the likelihood, I include all state and decision variables – unobserved variables such as true ability, job search and savings decisions⁴⁴ and productivity signals as well as observed variables such as work decisions and labor earnings.

The algorithm used for the minimization is a quasi-Newton's method. The derivative-free simplex method is also used to verify the estimates.

⁴³Keane and Smith (2003)'s smoothing method in this case requires me to go back two periods and describe all conditional probabilities between the two periods, which is an extremely difficult task – especially when there are continuous state variables such as posterior beliefs and savings. I tried this method first, but I could not achieve enough smoothing regardless of the choice of smoothing parameters.

⁴⁴Although savings are observed in the NLSY data, the variable is very noisy and available only for limited time periods.

4.3 Estimation Results

The estimation results are summarized in Table 6 and 7. Figure 1 and 2 show how the model fits the data. The risk-neutral workers case ($\rho = 0$) is not meaningfully different from the risk-averse workers with savings case ($\rho = 1$) in earnings growth and job mobility; however, employment levels are quite different between the two cases.

Table 6: Estimates: Risk-Neutral Case ($\rho = 0$)

Parameter	$\hat{\theta}$	S.E. ($\hat{\theta}$)	
γ_0	6.4552	(0.0231)	Initial Average Productivity
γ_1	0.0528	(0.0041)	General Skill Accumulation (1st order)
γ_2	-0.0017	(0.0001)	General Skill Accumulation (2nd order)
γ_3	0.0116	(0.0068)	Job-Specific Skill Accumulation (1st order)
γ_4	-0.0004	(0.0003)	Job-Specific Skill Accumulation (2nd order)
γ_θ	0.1783	(0.0169)	Productivity Effect of (+1SD) Ability
$\gamma_{g,\theta}$	0.0457	(0.0346)	(Learning) Ability Effect (+1SD) - General Skill Production
$\gamma_{s,\theta}$	0.0962	(0.0813)	(Learning) Ability Effect (+1SD) - Specific Skill Production
γ_ϵ	0.3227	(0.0147)	Productivity Effect of (+1SD) Job Match Quality
σ_ν^2	0.4886	(0.0343)	Variance: Wage Measurement Error
κ	0.4899	(0.0960)	Scale Parameter: Second Stage Preference Shock
b	13,493	(3,269)	Leisure Utility
σ_η^2	0.0721	(0.0056)	Variance: Noise in Productivity Signals
λ	0.8804	(0.0358)	Job Offer Arrival Probability
δ	0.0501	(0.0080)	Job Destruction Probability
c	10,351	(3,312)	Job Search Costs

Asymptotic standard errors are in parentheses.

Figure 1 and 2 compare log real weekly earnings, job change and employment status by observed ability between the real data (left panels) and the simulated data (right panels). These data moments are not directly used in the estimation; the auxiliary model comprises a set of equations describing year-to-year changes. The sample is split into two groups by observed ability – high AFQT (\geq median) and low AFQT ($<$ median) groups. In the graphs from the real data, high AFQT group shows not only higher levels but also faster growth of labor earnings than low AFQT group. Also, the former group moves less and works more than the other group throughout their working lives. It is well matched with the regression results in the previous section.

Table 7: Estimates: Risk-Averse Case ($\rho = 1$)

Parameter	$\hat{\theta}$	S.E. ($\hat{\theta}$)	
γ_0	6.4556	(0.0225)	Initial Average Productivity
γ_1	0.0455	(0.0043)	General Skill Accumulation (1st order)
γ_2	-0.0015	(0.0002)	General Skill Accumulation (2nd order)
γ_3	0.0193	(0.0049)	Job-Specific Skill Accumulation (1st order)
γ_4	-0.0004	(0.0002)	Job-Specific Skill Accumulation (2nd order)
γ_θ	0.1691	(0.0109)	Productivity Effect of (+1SD) Ability
$\gamma_{g,\theta}$	0.0321	(0.0152)	(Learning) Ability Effect (+1SD) - General Skill Production
$\gamma_{s,\theta}$	0.0319	(0.0176)	(Learning) Ability Effect (+1SD) - Specific Skill Production
γ_ϵ	0.3504	(0.0196)	Productivity Effect of (+1SD) Job Match Quality
σ_ν^2	0.4783	(0.0345)	Variance: Wage Measurement Error
κ	0.6393	(0.0680)	Scale Parameter: Second Stage Preference Shock
b	0.4602	(0.0485)	Leisure Utility
σ_η^2	0.1154	(0.0123)	Variance: Noise in Productivity Signals
λ	0.8909	(0.0358)	Job Offer Arrival Probability
δ	0.0240	(0.0053)	Job Destruction Probability
c	0.2817	(0.1440)	Job Search Costs

Asymptotic standard errors are in parentheses.

The estimated model (in both cases) fits the most important qualitative features with a relatively small number of parameters (16). First, average labor earnings increase over time at a decreasing rate, and the growth rate is significantly different across ability groups, up to almost .2 log points. Second, job mobility decreases over time at a decreasing rate, and less able workers move relatively more frequently. Third, employment ratio fluctuates around .9-.95. Employment ratio initially rises after labor market entry and then decreases afterwards – especially for low ability workers in the data.

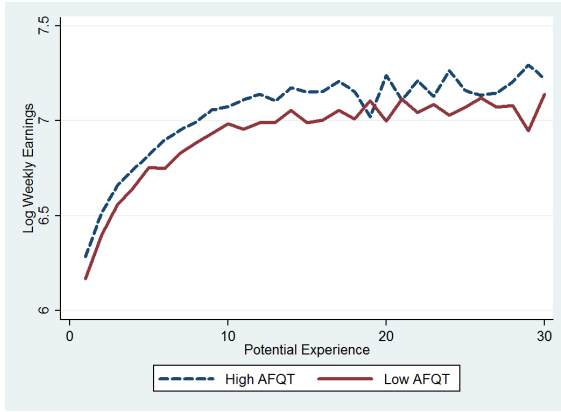
There are several patterns in the data not very well explained by the model, which leave room for improvement. First, the curvature in the simulated age-earnings profile is not flexible enough to capture the rapid earnings growth during early career in the data. This is due to the functional form restriction on skill accumulation (quadratic). Second, the differences in employment ratio across ability groups, especially in late career are not very well matched. One reason is the large variances of employment-related moments in the data; as a result, employment ratio has only very small weights in the estimation procedure.

Another reason is complicated incentives near retirement age. Relatively high employment ratio of senior workers with high ability (or relatively high employment ratio of senior workers with low ability) is likely related to life expectancy, bequest motives and medical expenses (e.g., French and Jones, 2011) or disability insurance, which are all omitted in the current model.

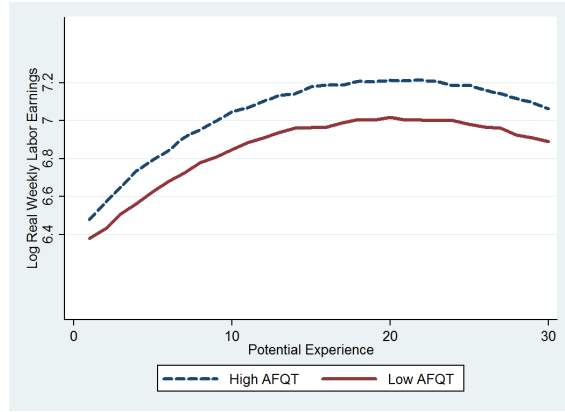
It is noteworthy that both job mobility and employment in the risk-neutral case are affected by *misperceived* productivity as expected. Less able workers move more because of their incorrectly low belief on job match quality. They work less because of their belief on the ability is on average low; however, they work more compared to the full information situation – that is, they are more likely to search and more willing to work because of their incorrectly high belief on ability, according to the model. As time passes on in the labor market, these incentives based on *misperceived* productivity disappear.

Figure 1: Model Fit (Risk-neutral case, $\rho = 0$) (Left: Data, Right: Model)

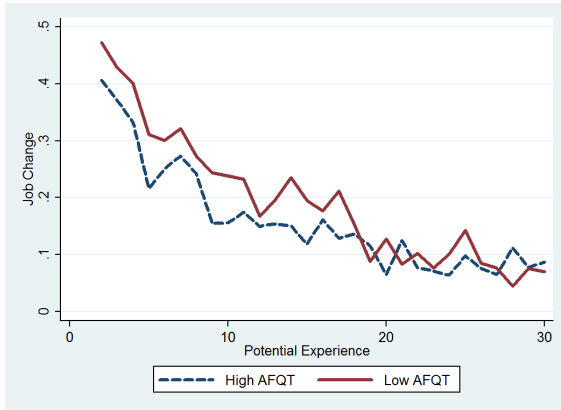
(a) Data: Log Earnings



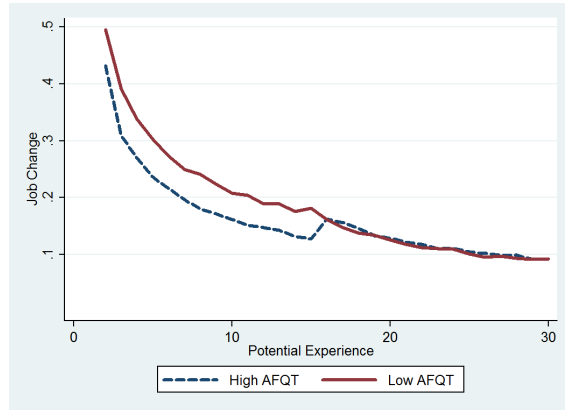
(b) Model: Log Earnings



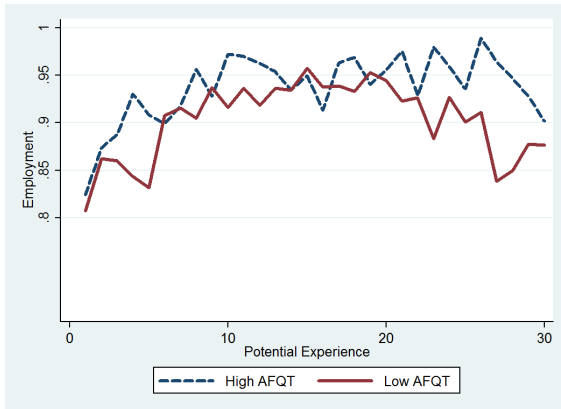
(c) Data: Job Change



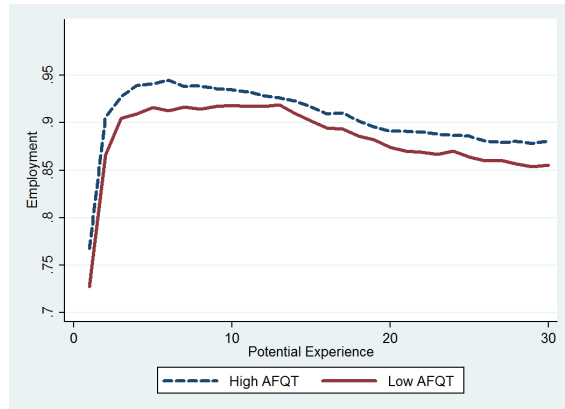
(d) Model: Job Change



(e) Data: Employment



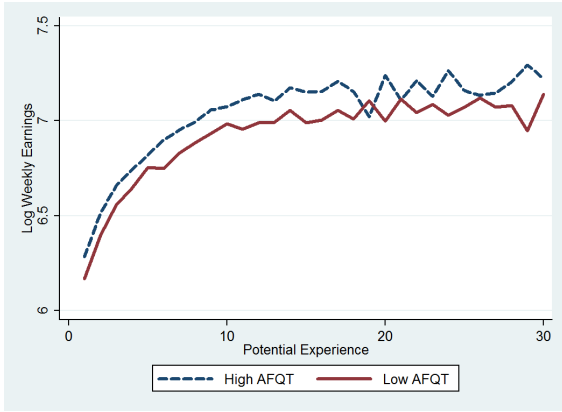
(f) Model: Employment



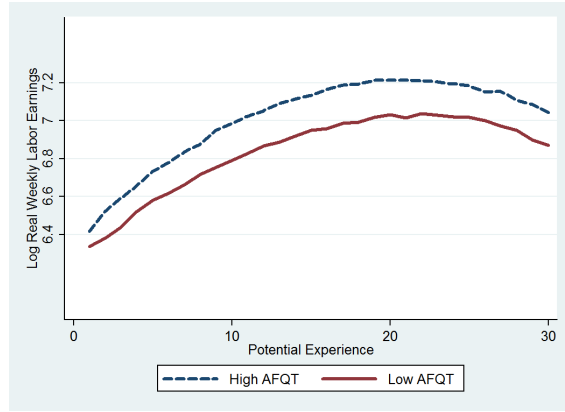
Note: Left panels are from real data. 1. log real weekly earnings is first regressed on a quartic equation in potential experience and its interaction terms with the AFQT score with other controls such as regional dummies (urban, SMSA, and Census regions) and yearly fixed effects. Then the estimated regional and yearly fixed effects are subtracted from the log earnings variable. 2. job change is empirically defined by working on the interview date and having changed to a new employer within 52 weeks. 3. employment variable is an indicator of working on the interview date. Right panels are from simulated data. The graphs show cross-sectional averages over potential experience. High AFQT: $\geq \text{med.}$, Low AFQT: $< \text{med.}$

Figure 2: Model Fit (Risk-averse case with savings, $\rho = 1$) (Left: Data, Right: Model)

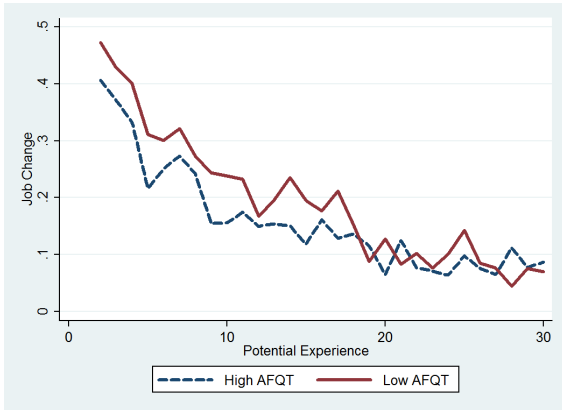
(a) Data: Log Earnings



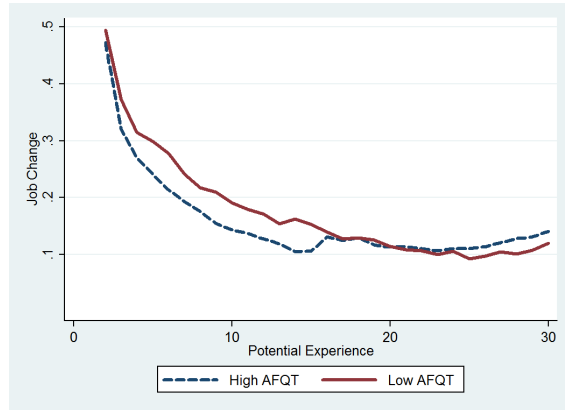
(b) Model: Log Earnings



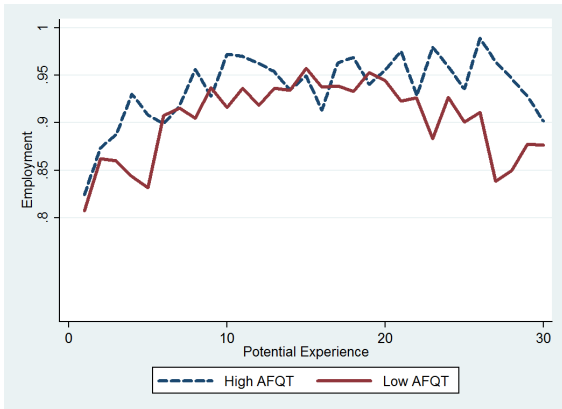
(c) Data: Job Change



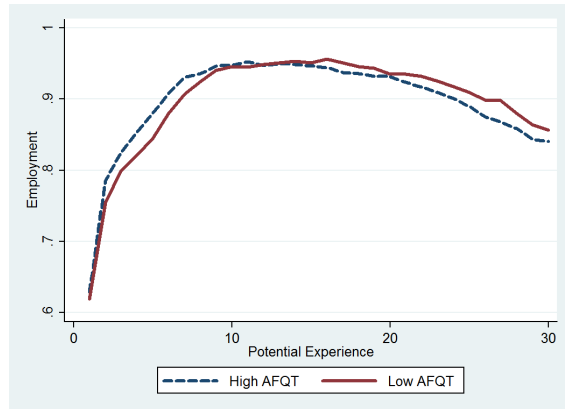
(d) Model: Job Change



(e) Data: Employment



(f) Model: Employment



Note: Left panels are from real data. 1. log real weekly earnings is first regressed on a quartic equation in potential experience and its interaction terms with the AFQT score with other controls such as regional dummies (urban, SMSA, and Census regions) and yearly fixed effects. Then the estimated regional and yearly fixed effects are subtracted from the log earnings variable. 2. job change is empirically defined by working on the interview date and having changed to a new employer within 52 weeks. 3. employment variable is an indicator of working on the interview date. Right panels are from simulated data. The graphs show cross-sectional averages over potential experience. High AFQT: $\geq \text{med.}$, Low AFQT: $< \text{med.}$

5 Counterfactual Analyses: Earnings Dynamics

In this section, I perform several simulations on the sources of average earnings growth and heterogeneity in earnings growth. I only report the simulation results from the risk-averse case because the risk-neutral case is not very different.

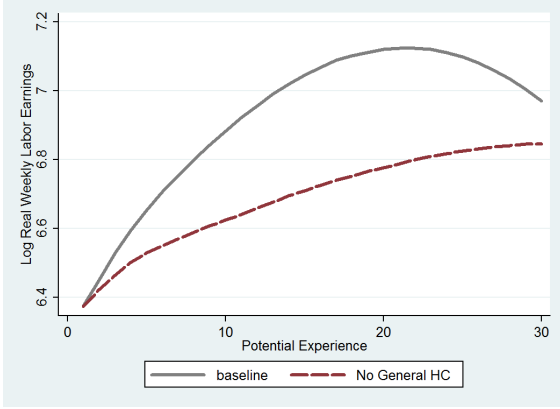
5.1 Average earnings growth

Earnings growth can arise from various channels, according to the model. Earnings may increase because of direct productivity improvement by general and/or job-specific skill accumulation. Earnings can also grow simply due to improved job match quality – on average, job match gets better over the life cycle when workers and employers learn something about their matches and terminate ones perceived as unproductive (job shopping). Individual earnings become more dispersed over time as ability is gradually reassessed after productivity signals in the labor market – individual earnings will diverge to the levels associated with their true ability (learning about ability). Of course, all components are interconnected as well.

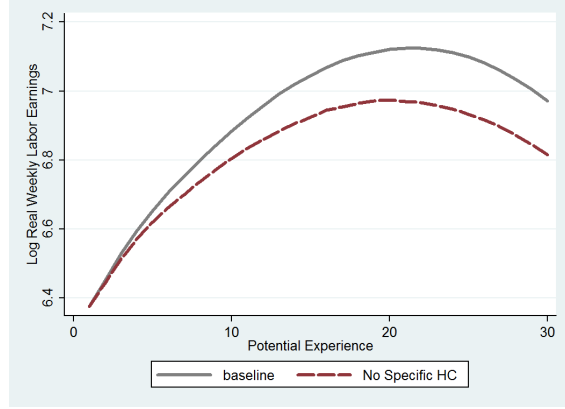
Figure 3 shows how much each channel contributes to average earnings growth, by shutting down each channel one by one. Skill accumulation is the most important contributor to life-cycle earnings growth. General skill accumulation accounts for .12 log points increase in average earnings over the first 5 years and .28 and .34 log point increases over the first 10 and 20 years in the labor market (Figure 3 (a)). Job-specific skill accumulation explains .04, .10 and .14 log point increases over the first 5, 10 and 20 potential experience, respectively (Figure 3 (b)). General skill accumulation is surely more important, but the relative importance of job-specific skill accumulation increases over time. This is partly due to the possibility of recall to a previous employer.

Job shopping also contributes considerably. Average job match quality rises quickly up to .32, .53 and .80 SD during the first 5, 10 and 20 years, respectively, and this is

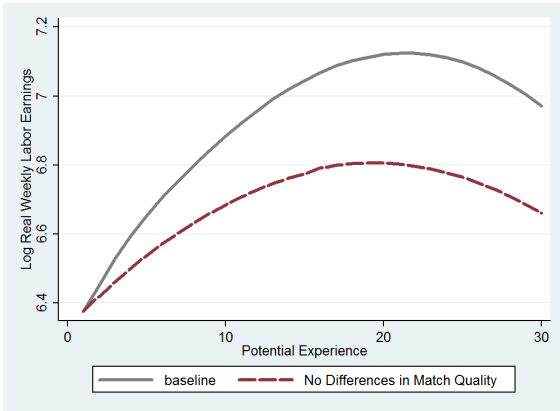
Figure 3: Various Reasons for Earnings Growth: Counterfactuals



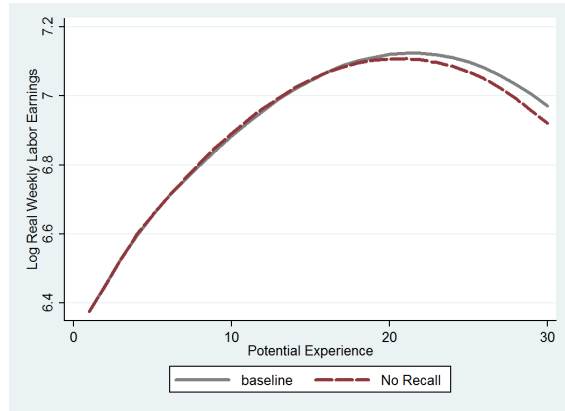
(a) No General Human Capital



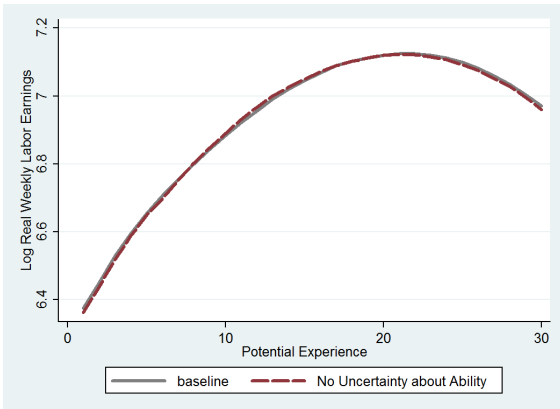
(b) No Job-Specific Human Capital



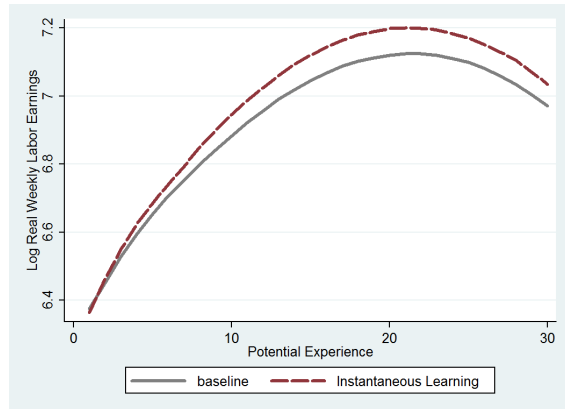
(c) No Job Shopping



(d) No Recall



(e) No Uncertainty about Ability

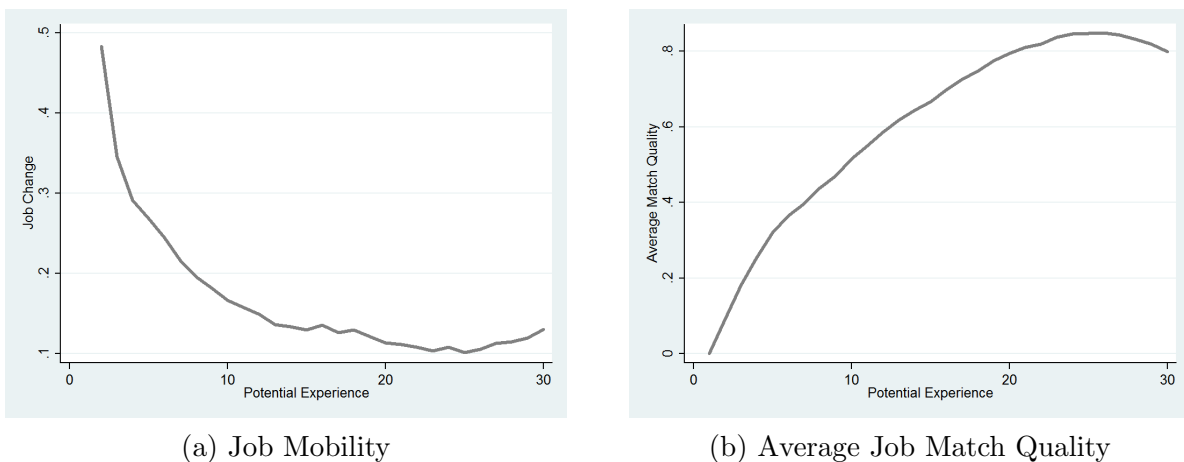


(f) No Info. Friction (Full Info.)

clearly associated with job mobility (Figure 4). That is, young workers select into stable employment relations through job changes as described in Topel and Ward (1992). This average match enhancement explains .10, .22 and .32 log point increases in earnings for 5, 10 and 20 years of potential experience (Figure 3 (c)). The contribution of job shopping to earnings growth is almost two times higher than other recent estimates (e.g., Altonji et al., 2013). The matching gains are partly associated with the possibility of recall, especially after 10 years of labor market experience (Figure 3 (d)). That is, good matches are saved thanks to recalls. The skill- and match- preserving effect of recall is very small during the first 10 years, but increasingly important, about .01 and .05 after 20 and 30 years in the labor market, respectively.

Information frictions significantly restrict average earnings growth. If there were no information friction, earnings would be up to .07 log points higher according to panel (f) in Figure 3. On the contrary, the uncertainty in ability has almost no effect on average earnings growth (Figure 3 (e)), but it clearly affects the distribution of earnings growth – the next subsection explains how it affects individual heterogeneity in earnings growth.

Figure 4: Job Mobility and Average Job Match Quality over the Life Cycle

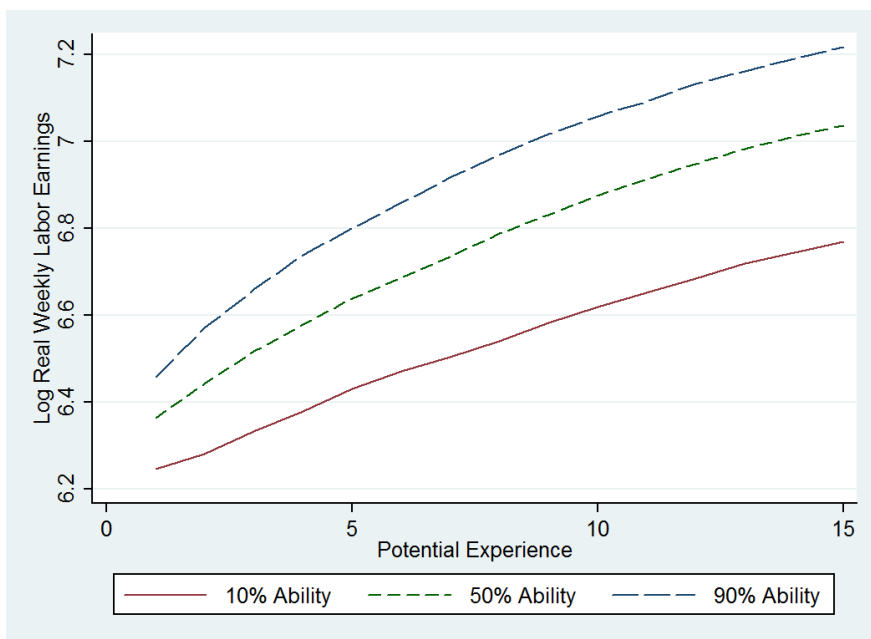


5.2 Individual heterogeneity in earnings growth

Decomposing average earnings growth into various sources provides useful information, but it can be even more interesting for policymakers to understand the reasons for heterogeneous earnings growth. If skill accumulation and information-updating processes are heterogeneous across individuals, the implications on labor market policies and income transfer programs can be surprisingly different from the policy implications of the previous analysis on average earnings growth. For example, job mobility contributes much to young workers' average earnings growth, but the effect can also be widely different across individuals. Labor market flexibility might have much lower or even negative impact on the earnings growth of less able workers.

Figure 5 shows how earnings growth is different by ability. The differences in ability are associated with not only entry-level earnings but also earnings growth over the life cycle.

Figure 5: Heterogeneous Earnings Growth by Worker Ability



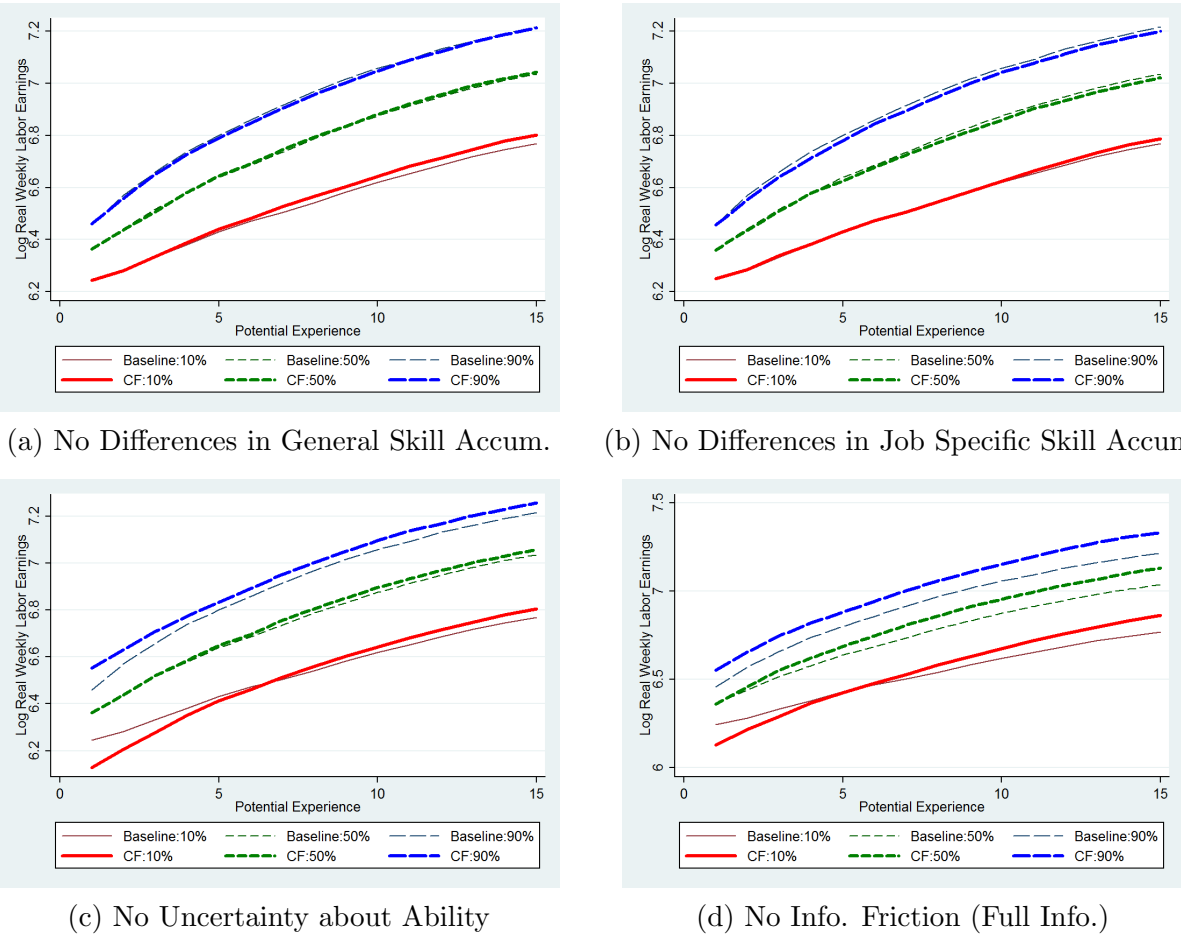
The model decomposes the distribution of earnings growth into various channels (Figure 6). Ability affects earnings growth mainly through information-updating channels.

The first two panels, (a) and (b), in Figure 6 show two counterfactual experiments on skill

accumulation. In the first counter-factual situation that everyone accumulates general skill at the same median speed (50 percentile in ability), we see only slight differences between the baseline and the counterfactual age-earnings profiles (Figure 6, (a)). In a similar experiment regarding job-specific skill accumulation (Figure 6, (b)), we also observe almost no difference between the baseline and the counterfactual profiles.

The third panel (Figure 6 (c)) shows another counterfactual experiment on uncertainty about ability. If ability were certain from the beginning of labor market experience, labor earnings would grow at almost the same speed for all workers. It is noteworthy that this would be achieved by increased earnings inequality among young workers.

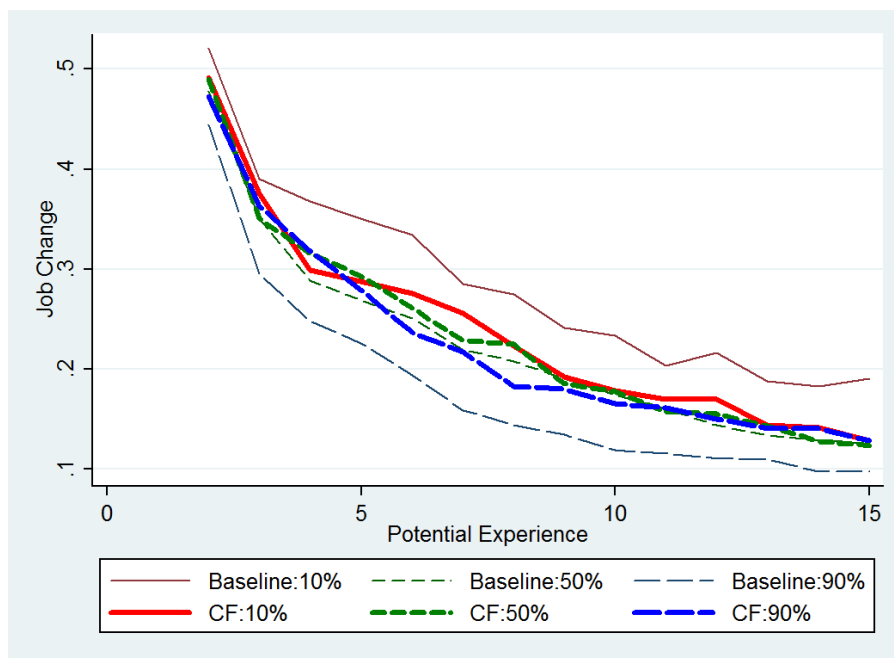
Figure 6: Various Reasons for Differential Earnings Growth by Ability: Counterfactuals



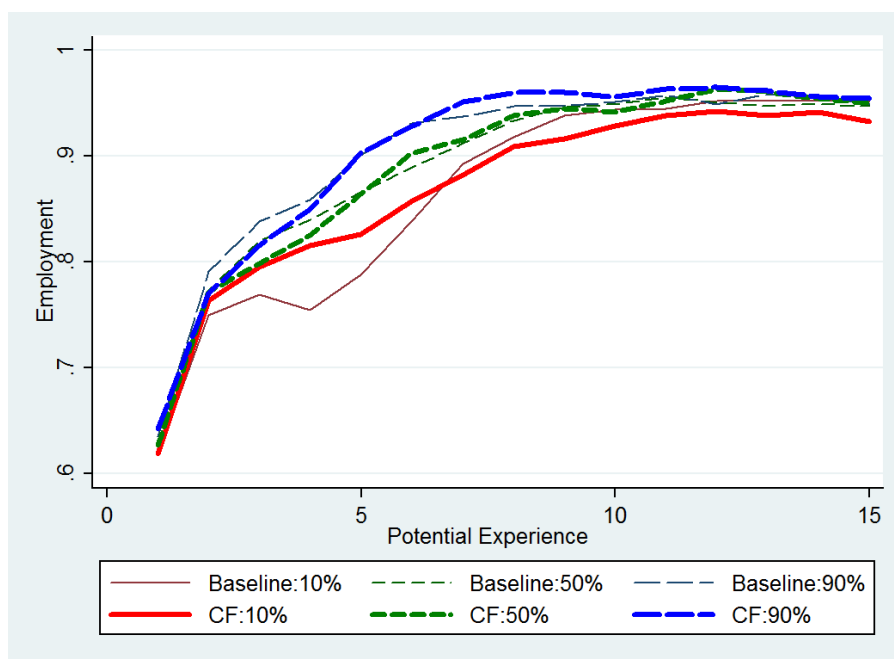
Two different effects contribute to the increased earnings inequality among young work-

Figure 7: Job Mobility and Employment by Ability (CF: No Uncertainty about Ability)

(a) Job Mobility



(b) Employment



Note: Job change indicates working at a new job conditional on working.

ers with full information on ability. The first effect is a direct reassessment effect. With only partial information on ability, individual earnings level gradually converges to the level associated with true ability, which is similar to the story in employer learning (e.g., Altonji and Pierret, 2001). With full information on ability, earnings level would jump to the true level from the beginning.

The second effect is an indirect matching efficiency effect. When workers (and employers) cannot directly distinguish between ability and job match effects, negative selection into job mobility can arise as a result. For example, high ability workers misinterpret their good productivity signals in favor of their job match because they perceive themselves as mediocre workers (conditional on certain characteristics such as a test score). As a result, they tend to stay at a job even when it is desirable to move to another job. On the contrary, low ability workers tend to move to another job even when it is actually better for them to stay there. They misinterpret their bad productivity signals as signs of a wrong match because of their (partially) incorrect prior belief. This matching inefficiency gradually disappears over time as workers (and employers) learn more about their ability with more signals.

Figure 7 shows that the differences in job mobility across ability groups would disappear if initial uncertainty about ability were not there. More information would increase the overall efficiency in terms of utility for all workers, but the realized job match quality improvement could be highly heterogeneous across ability groups. If there were no uncertainty about ability, high ability workers would improve job match quality by more aggressive job shopping; low ability workers would either save moving costs or improve job match quality by reducing unnecessary job moves.⁴⁵ The median ability group would not show any differences in job match quality.

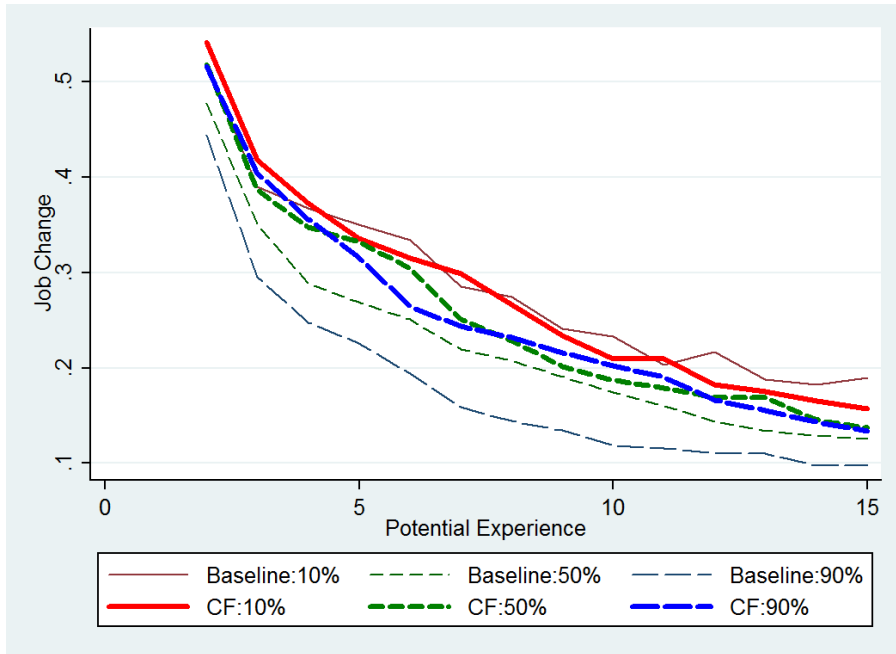
The panel (d) in Figure 6 suggests that full information would initially increase earnings inequality among young workers but improved job matching efficiency would eventually raise

⁴⁵The risk-neutral case is meaningfully different from the risk-averse case only in this part. Less able workers in risk-neutral case do not gain much from more information on their ability while the same workers in risk-averse case do benefit from better information.

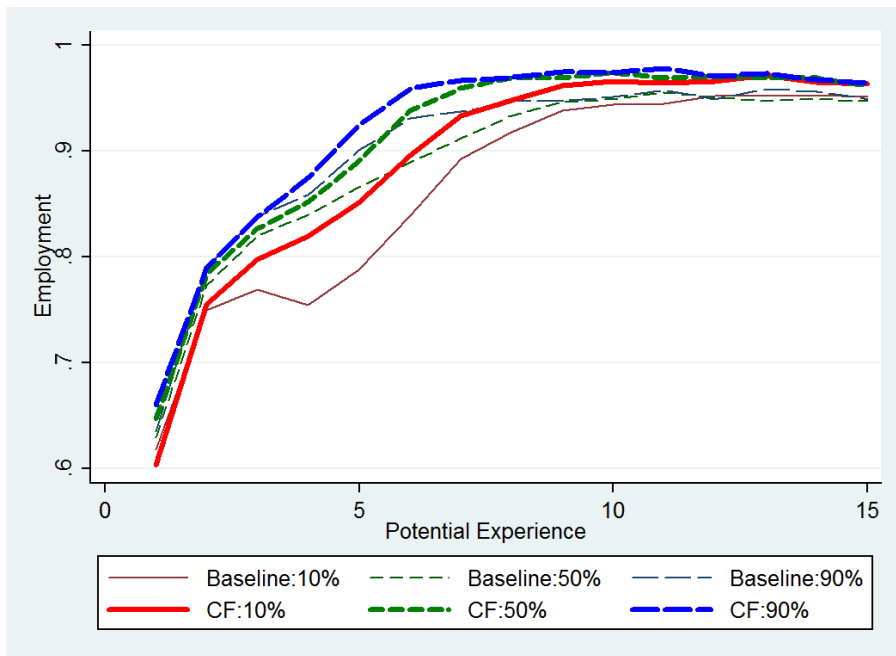
labor earnings of all workers above the baseline level. As predicted, job mobility would not be different across ability groups, but the average job mobility would be initially higher than in the previous case, by almost 10 percent point (Figure 8 (a)). This is related to increased amount of information in the labor market, and it leads to increases in average job match quality.

Figure 8: Job Mobility and Employment by Ability (CF: Full Information)

(a) Job Mobility



(b) Employment



Note: Job change indicates working at a new job conditional on working.

6 Conclusion

This paper develops a model of life-cycle career decisions under various types of uncertainty, focusing on the roles of skill accumulation and information-updating in earnings dynamics. Workers accumulate general and job-specific skills over the life cycle, possibly at different speeds according to their ability; workers learn about both work ability and job match quality by trial and error. In this setup, skill accumulation and information updating processes can be interconnected; more importantly, various uncertainties can jointly affect work decisions, producing distinctive predictions such as a negative but disappearing correlation between ability and job change.

With an additional identifying assumption that the AFQT score carries over some information on unknown ability, I estimated the model from a sample of white male high school graduates who did not attain any post-secondary education nor have any military experience throughout the NLSY79 surveys.

The estimated model shows that average life-cycle earnings growth develops from various sources: general skill accumulation accounts for approximately 33 percent points of earnings growth over the first 10 years; job-specific skill accumulation, 10 percent points; and job shopping, 24 percent points during the same period. The contribution of job shopping to earnings growth is almost two times higher than previous estimates, which suggests the importance of individual heterogeneity in understanding job mobility and life-cycle earnings growth. Recall to a previous employer plays a role during the later career. Information friction restricts average earnings growth considerably, about 7 percent points.

The model also provides an opportunity to look at individual heterogeneity in earnings growth. Information and uncertainty explain most of the difference in earnings growth across individuals: first, individual earnings converge to the level associated with true ability; second, job mobility is heterogeneous, and the improvement in job match quality through job shopping is different across ability groups. Skill accumulation is fairly homogeneous across individuals and explains little about individual heterogeneity in earnings growth.

References

- Altonji, Joseph G and Charles R Pierret (2001), “Employer learning and statistical discrimination.” *The Quarterly Journal of Economics*, 116, 313–350.
- Altonji, Joseph G and Robert A Shakotko (1987), “Do wages rise with job seniority?” *The Review of Economic Studies*, 54, 437–459.
- Altonji, Joseph G, Anthony A Smith, and Ivan Vidangos (2013), “Modeling earnings dynamics.” *Econometrica*, 81, 1395–1454.
- Antonovics, Kate and Limor Golan (2012), “Experimentation and job choice.” *Journal of Labor Economics*, 30, 333–366.
- Arcidiacono, Peter (2004), “Ability sorting and the returns to college major.” *Journal of Econometrics*, 121, 343–375.
- Arcidiacono, Peter, Patrick Bayer, and Aurel Hizmo (2010), “Beyond signaling and human capital: Education and the revelation of ability.” *American Economic Journal: Applied Economics*, 2, 76–104.
- Becker, Gary S (1964), “Human capital theory.” *Columbia, New York*.
- Farber, Henry S (1999), “Mobility and stability: The dynamics of job change in labor markets.” *Handbook of labor economics*, 3, 2439–2483.
- Farber, HS and R Gibbons (1996), “Learning and wage dynamics.” *Quarterly Journal of Economics*, 111, 1007–1047.
- French, Eric and John Bailey Jones (2011), “The effects of health insurance and self-insurance on retirement behavior.” *Econometrica*, 693–732.
- Fujita, Shigeru and Giuseppe Moscarini (2013), “Recall and unemployment.” Technical report, National Bureau of Economic Research.

- Gathmann, Christina and Uta Schönberg (2010), “How general is human capital? a task-based approach.” *Journal of Labor Economics*, 28, 1–49.
- Gladden, Tricia and Christopher Taber (2009), “Turnover and wage growth for low skilled young men.” *Unpublished manuscript*.
- Gourieroux, Christian, Alain Monfort, and Eric Renault (1993), “Indirect inference.” *Journal of applied econometrics*, 8, S85–S85.
- Greenwald, Bruce C (1986), “Adverse selection in the labour market.” *The Review of Economic Studies*, 325–347.
- Heckman, James J, Jora Stixrud, and Sergio Urzua (2006), “The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior.” *Journal of Labor Economics*, 24, 411–482.
- James, Jonathan (2011), “Ability matching and occupational choice.” *FRB of Cleveland Working Paper*.
- Jovanovic, Boyan (1979), “Job matching and the theory of turnover.” *The Journal of Political Economy*, 87, 972–990.
- Jovanovic, Boyan (1984), “Matching, turnover, and unemployment.” *The Journal of Political Economy*, 108–122.
- Keane, Michael and Anthony A Smith (2003), “Generalized indirect inference for discrete choice models.”
- Kennan, John (2006), “A note on discrete approximations of continuous distributions.” *University of Wisconsin Madison*, 211–251.
- Kennan, John (2015), “Spatial variation in higher education financing and the supply of college graduates.” Technical report, National Bureau of Economic Research.

- Lange, Fabian (2007), “The speed of employer learning.” *Journal of Labor Economics*, 25, 1–35.
- Miller, Robert A (1984), “Job matching and occupational choice.” *The Journal of Political Economy*, 1086–1120.
- Moscarini, Giuseppe (2005), “Job matching and the wage distribution.” *Econometrica*, 73, 481–516.
- Murphy, Kevin M and Finis Welch (1990), “Empirical age-earnings profiles.” *Journal of Labor economics*, 202–229.
- Nagypál, Éva (2007), “Learning by doing vs. learning about match quality: Can we tell them apart?” *The Review of Economic Studies*, 74, 537–566.
- Neal, Derek (1999), “The complexity of job mobility among young men.” *Journal of Labor Economics*, 17, pp. 237–261.
- Papageorgiou, Theodore (2014), “Learning your comparative advantages.” *The Review of Economic Studies*, 81, 1263–1295.
- Pavan, Ronni (2011), “Career choice and wage growth.” *Journal of Labor Economics*, 29, 549–587.
- Pinkston, Joshua C (2009), “A model of asymmetric employer learning with testable implications.” *The Review of Economic Studies*, 76, 367–394.
- Sanders, Carl (2014), “Skill uncertainty, skill accumulation, and occupational choice.” *Washington University, St. Louis*.
- Sanders, Carl and Christopher Taber (2012), “Life-cycle wage growth and heterogeneous human capital.” *Annual Review of Economics*, 4, 399–425.

Sauer, Robert M. and Christopher Taber (2013), “Importance sampling in indirect inference.”

Schönberg, Uta (2007), “Testing for asymmetric employer learning.” *Journal of Labor Economics*, 25, 651–691.

Topel, Robert H (1991), “Specific capital, mobility, and wages: Wages rise with job seniority.” *Journal of Political Economy*, 99, 145–76.

Topel, Robert H and Michael P Ward (1992), “Job mobility and the careers of young men.” *The Quarterly Journal of Economics*, 107, 439–479.

Appendix

Appendix 1: NLSY79

For cognitive ability measures, I use age-adjusted Armed Forces Qualification Test (AFQT) and Armed Services Vocational Aptitude Battery (ASVAB) scores.⁴⁶ The AFQT score is a weighted sum of four section scores (1981), which are Arithmetic Reasoning, Word Knowledge, Paragraph Comprehension and Numerical Operations with a half weight on the last one. I follow Lange (2007)'s construction of age-adjusted AFQT score. The ASVAB Math and Verbal scores are constructed from the original answer sheets by the NLS in 2010. These scores are maximum likelihood estimates of true abilities by the Item-Response Theory (IRT). The ASVAB scores are standardized within 4-month age intervals by the NLS, and I re-standardize each score to have mean 0 and s.d. 1 in each birth year group.

For non-cognitive ability measures, I use the Rotter's Locus of Control Scale and the Rosenberg Self-Esteem Scale in 1980. The Rotter's Locus of Control scale measures the extent that individuals attributes events in their life to uncontrollable factors. A low score (a negative z-score) indicates more internal, and a high score (a positive z-score) signifies more external locus of control. The Rosenberg Self-Esteem scale is a measure of self-worth based on a series of questions on how one feels about the self. Each item has four alternatives – from strongly agree to strongly disagree. I re-standardize each score to have mean 0 and s.d. 1 in each birth year group.

For education and work histories, I mostly follow the conventions developed in the previous literature.(e.g., Altonji and Pierret, 2001; Lange, 2007; Arcidiacono et al., 2010). Some measures (education groups, employment, job change) are redefined for the purpose of this paper.

⁴⁶The ASVAB was administered to 11,914 NLS respondents (94 percent) during July through October of 1980 for the purpose of establishing a new national norm of the test (NLSY Attachment 106). ASVAB score are used to determine eligibility and assignment qualifications for specific military jobs for new enlistees, and the AFQT score, the sum of four subsection scores (word knowledge, paragraph comprehension, arithmetic reasoning and numeric operations), is a general measure of trainability and a primary criteria of enlisted eligibility for the Armed Forces (NLSY Attachment 106).

First, I use three schooling measures: $hgc0$, hgc and $hgcf$. hgc is the highest grade completed on the interview date, and $hgc0$ is the highest grade completed on May 1 in the first graduation year, and $hgcf$ is the highest grade completed ever throughout the surveys. There are very few cases that breaks the relationship $hgc0 \leq hgc \leq hgcf$, and $hgcf$ is on average 0.65 year higher than $hgc0$ – almost 30 percent in the sample enrolled in college again after they had once stopped enrollment because of completion or graduation. Since extra college education choice can bring biases into estimation results, I use only a sample of high school graduates ($hgc0 = hgcf$) who are without a G.E.D. or any post-secondary education in the estimation of the model.

Second, the first entry into the labor market is defined by their first graduation year in the surveys, which is the earliest year that the respondent answers as a graduation year. The first graduation year is equal to the year of high school graduation for high school graduates ($hgc0 = hgcf$). The first graduation year is almost always the same with the last year before enrollment status (May 1) changes to completion/graduation for the first time in the surveys – the latter definition is not applicable to some respondents who had already graduated by their first interview in 1979.

Third, potential experience is simply the number of years after the first entry into the labor market. This is roughly equal to age minus 18 for high school graduates ($hgc0 = hgcf$).

Fourth, I focus on the main job which is the current or most recent job (job 1/CPS job) to determine one's employment status. A person is employed if and only if he answers as working at his main job on the interview date. Since multiple jobs are not uncommon, nonemployment by this definition may not correspond to actual unemployment or non-participation.

Fifth, a job change is also defined by the main job. A job is equivalent to an employer in the NLSY data, and I empirically define a job change (=employer change) as working on the interview date for a different employer from the employer before 52 weeks.

Sixth, I use weekly earnings from the main job as the measure of labor earnings. It is

constructed from the interview responses about the rate of pay and the time unit. Then the number is converted to the equivalent value in 2010 dollars using the average CPI index. Wages and earnings are used inter-changeably throughout this paper, and they always mean weekly earnings from the main job.⁴⁷

Appendix 2: Bayesian Updating with Two (or More) Unknown Objects – When One of the Objects is Renewable.

When a signal y_τ is a weighted sum of two unknown objects, Θ , and a pure noise term (η), the posterior beliefs on the unknown variables after t signals (no changes in the objects) computed as follows:

$$\begin{aligned}
 y_\tau &= \gamma'_\tau \Theta + \eta_\tau, \tau = 1, \dots, t \\
 y &= \Gamma \Theta + \eta && \text{(Matrix notation)} \\
 f(\Theta|y) &\propto \int f(y|\Theta)f(\Theta) && \text{(Bayes' rule)}
 \end{aligned}$$

where γ'_τ are the deterministic weights on the unknown objects at period τ .

When all distributions are normally distributed, the beliefs after observing t signals are given by:

prior beliefs: $\Theta \sim N(\Theta_0, \Sigma_0)$, signal: $\eta_t \sim N(0, \sigma_\eta^2)$.

$$\begin{aligned}
 \Sigma_t &= (\Sigma_0^{-1} + (\Gamma'\Gamma)/\sigma_\eta^2)^{-1} \\
 \Theta_t &= \Sigma_t (\Sigma_0^{-1}\Theta_0 + (\Gamma'y)/\sigma_\eta^2)
 \end{aligned} \tag{2}$$

This procedure is equivalent to Bayesian linear regression – intuitively, unknown objects (Θ) are “parameters,” and observed signals (y_t) and time-varying weights (γ_t) are “data” from the perspective of a Bayesian-rational agent. This expression holds when there is no

⁴⁷This paper do not separately model working hour decision – hours worked is very hard to match in a behavioral model. Instead of separately modeling hourly wage determination and hours decision, I focus on weekly earnings.

change in the unknown objects.

If one of the unknown objects – for example, job match quality – can be renewed, we need a modified expression. It may be possible to write a fully general expression including job changes, but it can also be concisely expressed using a concept of updating “priors” at job changes. Whenever a worker moves to another job, the same expression applies but with a new set of “prior” (or initial) beliefs. For example, at the first job change, the current belief on ability at time $t = T_{1t}$ becomes a new prior belief on ability at time t – the belief on job match quality is simply reset, including the covariance. At the second job change, the current belief on ability at time $t = T_1 + T_{2t}$ becomes a new prior belief on ability at time t – again, the belief on job match quality and the covariance are simply reset. This two-way updating process can go on and on.

Hence, it is easy to see that this learning process, as a whole, is a Markov process. In calculating the beliefs in the next period $(\Theta_{j,t+1}, \Sigma_{j,t+1})$, the past signals up to period t (y_1, \dots, y_t) and the history of job changes (d_1, \dots, d_t) do not add any more information conditional on current beliefs $(\Theta_{jt}, \Sigma_{jt})$, current job tenure (T_{jt}) , and the “prior beliefs” at current job entry $(\Theta_{j0}, \Sigma_{j0})$. In fact, the sufficient statistics for past history can be further reduced: the mean of current beliefs (Θ_{jt}) , current job tenure (T_{jt}) , and the uncertainty of the belief on unchanging object at current job entry $(\sigma_{\theta,j0}^2)$ because Σ_{jt} is easily backed out from $\sigma_{\theta,j0}^2$ and T_{jt} .

In practice, it is very hard to accurately calculate the uncertainty of current beliefs (Σ_{jt}) – small approximation errors in the initial uncertainty on ability $(\sigma_{\theta,j0}^2)$ can become non-negligible approximation errors if we back out Σ_{jt} from approximated $\sigma_{\theta,j0}^2$ and T_{jt} . In this version, I choose to include full employment history (tenure spells at each work experience) in state variables in order to calculate exact posterior variances – at the cost of exponentially growing state space in time in the labor market.

Proof of Proposition 1: $y_{t+1} | \Theta_t, \Sigma_t \sim N(\gamma'_{t+1} \Theta_t, \gamma'_{t+1} \Sigma_t \gamma_{t+1} + \sigma_\eta^2)$

$y_{t+1}|\Theta_t, \Sigma_t$ is normally distributed because the prior beliefs on Θ follows a normal distribution, $N(\Theta_t, \Sigma_t)$, and the signal (y_{t+1}) is also normally distributed around Θ (or because the prior beliefs follows a normal distribution, $N(\Theta_0, \Sigma_0)$, and all signals (y_1, \dots, y_{t+1}) are jointly normally distributed around Θ). Also,

$$\begin{aligned} E[y_{t+1}|\Theta_t, \Sigma_t] &= E[\gamma'_{t+1}\Theta + \eta_{t+1}|\Theta_t, \Sigma_t] \\ &= \gamma'_{t+1}\Theta_t \\ \text{Var}[y_{t+1}|\Theta_t, \Sigma_t] &= \text{Var}[\gamma'_{t+1}\Theta + \eta_{t+1}|\Theta_t, \Sigma_t] \\ &= \gamma'_{t+1}\Sigma_t\gamma_{t+1} + \sigma_\eta^2 \quad \text{Q.E.D.} \end{aligned}$$

Proof of Proposition 2: $\Theta_{t+1}|\Theta_t, \Sigma_t \sim N(\Theta_t, (\Sigma_{t+1}\gamma_{t+1}/\sigma_\eta^2)(\gamma'_{t+1}\Sigma_t\gamma_{t+1} + \sigma_\eta^2)(\gamma'_{t+1}\Sigma_{t+1}/\sigma_\eta^2))$

$\Theta_{t+1}|\Theta_t, \Sigma_t$ is normally distributed because Θ_{t+1} is a linear sum of Θ_t and y_{t+1} , and y_{t+1} is normally distributed – uncertainty about future signal is the only source of variation in posterior beliefs. Also,

$$\begin{aligned} E[\Theta_{t+1}|\Theta_t, \Sigma_t] &= E[\Sigma_{t+1}(\Sigma_t^{-1}\Theta_t + \gamma_{t+1}y_{t+1}/\sigma_\eta^2)|\Theta_t, \Sigma_t] \\ &= \Sigma_{t+1}(\Sigma_t^{-1}\Theta_t + \gamma_{t+1}\gamma'_{t+1}\Theta_t/\sigma_\eta^2) \\ &= \Sigma_{t+1}(\Sigma_t^{-1} + \gamma_{t+1}\gamma'_{t+1}/\sigma_\eta^2)\Theta_t \\ &= \Theta_t \end{aligned}$$

$$\begin{aligned} \text{Var}[\Theta_{t+1}|\Theta_t, \Sigma_t] &= \text{Var}[\Sigma_{t+1}(\Sigma_t^{-1}\Theta_t + \gamma_{t+1}y_{t+1}/\sigma_\eta^2)|\Theta_t, \Sigma_t] \\ &= (\Sigma_{t+1}\gamma_{t+1}/\sigma_\eta^2)(\gamma'_{t+1}\Sigma_t\gamma_{t+1} + \sigma_\eta^2)(\Sigma_{t+1}\gamma_{t+1}/\sigma_\eta^2)' \\ &= (\Sigma_{t+1}(\gamma_{t+1}\gamma'_{t+1}/\sigma_\eta^2)\Sigma_t(\gamma_{t+1}\gamma'_{t+1}/\sigma_\eta^2)\Sigma_{t+1} + \Sigma_{t+1}(\gamma_{t+1}\gamma'_{t+1}/\sigma_\eta^2)\Sigma_{t+1}) \quad \text{Q.E.D.} \end{aligned}$$

As a special case, when there are two hidden objects and one of them has time-varying weights in signal, the posteriors for each unknown object can be computed as follows: (γ_t

are deterministically known weights)

$$y_t = \gamma_t \theta + \epsilon + \eta_t$$

$$f(\theta|y_1, \dots, y_t) \propto \int f(y_1, \dots, y_t|\theta) f(\theta) \quad (\text{Bayes' rule})$$

$$\propto \int f(y_1, \dots, y_t|\theta, \epsilon) f(\epsilon|\theta) d\epsilon f(\theta) \quad (\text{Integrate out})$$

$$f(\epsilon|y_1, \dots, y_t) \propto \int f(y_1, \dots, y_t|\epsilon) f(\epsilon) \quad (\text{Bayes' rule})$$

$$\propto \int f(y_1, \dots, y_t|\theta, \epsilon) f(\theta|\epsilon) d\theta f(\epsilon) \quad (\text{Integrate out})$$

When all distributions are normally distributed, i.e., $\theta \sim N(\bar{\theta}, \sigma_\theta^2)$, $\epsilon \sim N(\bar{\epsilon}, \sigma_\epsilon^2)$, $\eta \sim N(0, \sigma_\eta^2)$, it is straightforward to show that the posterior means (θ_t, ϵ_t) and variances $(\sigma_{\theta,t}^2, \sigma_{\epsilon,t}^2)$ after t signals are given by the following formulas.

$$\begin{aligned} \theta_t &= \sigma_{\theta,t}^2 \left(\frac{\bar{\theta}}{\sigma_\theta^2} + \frac{\sum \gamma_\tau y_\tau}{\sigma_\eta^2} - \frac{\sum \gamma_\tau \frac{\sum y_\tau}{\sigma_\eta^2} + \frac{\bar{\epsilon}}{\sigma_\epsilon^2}}{\frac{t}{\sigma_\eta^2} + \frac{1}{\sigma_\epsilon^2}} \right) \\ &= \sigma_{\theta,t}^2 \left(\frac{\bar{\theta}}{\sigma_\theta^2} + \frac{1}{\sigma_\eta^2} \frac{\sigma_\eta^2 (\sum \gamma_\tau (y_\tau - \bar{\epsilon})) + t \sigma_\epsilon^2 \sum \gamma_\tau (y_\tau - \bar{y})}{t \sigma_\epsilon^2 + \sigma_\eta^2} \right) \end{aligned}$$

$$\begin{aligned} \sigma_{\theta,t}^2 &= \left(\frac{1}{\sigma_\theta^2} + \frac{\sum \gamma_\tau^2}{\sigma_\eta^2} - \frac{\sum \gamma_\tau \frac{\sum \gamma_\tau}{\sigma_\eta^2}}{\frac{t}{\sigma_\eta^2} + \frac{1}{\sigma_\epsilon^2}} \right)^{-1} \\ &= \left(\frac{1}{\sigma_\theta^2} + \frac{1}{\sigma_\eta^2} \frac{\sigma_\eta^2 (\sum \gamma_\tau^2) + t \sigma_\epsilon^2 \{ \sum \gamma_\tau^2 - (\sum \gamma_\tau)^2 / t \}}{t \sigma_\epsilon^2 + \sigma_\eta^2} \right)^{-1} \end{aligned}$$

$$\begin{aligned}
\epsilon_t &= \sigma_{\epsilon,t}^2 \left(\frac{\bar{\epsilon}}{\sigma_\epsilon^2} + \frac{\Sigma y_\tau}{\sigma_\eta^2} - \frac{\Sigma \gamma_\tau \frac{\Sigma \gamma_\tau y_\tau}{\sigma_\eta^2} + \frac{\bar{\theta}}{\sigma_\theta^2}}{\sigma_\eta^2 \frac{\Sigma \gamma_\tau^2}{\sigma_\eta^2} + \frac{1}{\sigma_\theta^2}} \right) \\
&= \sigma_{\epsilon,t}^2 \left(\frac{\bar{\epsilon}}{\sigma_\epsilon^2} + \frac{1}{\sigma_\eta^2} \frac{\sigma_\eta^2 (\Sigma y_\tau - \Sigma \gamma_\tau \bar{\theta}) + \Sigma \gamma_\tau^2 \sigma_\theta^2 (\Sigma y_t - \Sigma \gamma_\tau \frac{\Sigma \gamma_\tau y_\tau}{\Sigma \gamma_\tau^2})}{\Sigma \gamma_\tau^2 \sigma_\theta^2 + \sigma_\eta^2} \right) \\
\sigma_{\epsilon,t}^2 &= \left(\frac{1}{\sigma_\epsilon^2} + \frac{t}{\sigma_\eta^2} - \frac{\Sigma \gamma_\tau \frac{\Sigma \gamma_\tau}{\sigma_\eta^2}}{\sigma_\eta^2 \frac{\Sigma \gamma_\tau^2}{\sigma_\eta^2} + \frac{1}{\sigma_\theta^2}} \right)^{-1} \\
&= \left(\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_\eta^2} \frac{\sigma_\eta^2 t + \Sigma \gamma_\tau^2 \sigma_\theta^2 (t - \frac{(\Sigma \gamma_\tau)^2}{\Sigma \gamma_\tau^2})}{\Sigma \gamma_\tau^2 \sigma_\theta^2 + \sigma_\eta^2} \right)^{-1}
\end{aligned}$$

This can be understood as two-stage updating process. The first stage is to update beliefs on the hidden component given signals. The second stage is to update beliefs on the variable of interest given first-stage beliefs of another variable, the hidden component. The resulted posteriors are equivalent to a weighted average of two different sets of posteriors which are separately calculated based on two different assumptions – only priors vs. only data.

When $\gamma_t = 1$ (or a constant), $\forall t$, the above expressions become much more simplified and look very similar to the expressions in the previous learning literature:

$$\begin{aligned}
\theta_t &= \sigma_{\theta,t}^2 \left(\frac{\bar{\theta}}{\sigma_\theta^2} + \frac{\bar{y}_t - \bar{\epsilon}}{\sigma_\epsilon^2 + \sigma_\eta^2/t} \right) \\
\sigma_{\theta,t}^2 &= \left(\frac{1}{\sigma_\theta^2} + \frac{1}{\sigma_\epsilon^2 + \sigma_\eta^2/t} \right)^{-1} \\
\epsilon_t &= \sigma_{\epsilon,t}^2 \left(\frac{\bar{\epsilon}}{\sigma_\epsilon^2} + \frac{\bar{y}_t - \bar{\theta}}{\sigma_\theta^2 + \sigma_\eta^2/t} \right) \\
\sigma_{\epsilon,t}^2 &= \left(\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_\theta^2 + \sigma_\eta^2/t} \right)^{-1}
\end{aligned}$$

The posterior mean on the sum of the two objects ($x = \theta + \epsilon$) must be equal to $\theta_t + \epsilon_t$.