

Within-Occupation Schooling Dispersion, Over-education and Mismatch in the Labor Market: Theory and Empirics

JIE MA*

November 15, 2018

Abstract

Concerns persist for years about whether individuals acquire more education than is required for their work, a phenomenon known as 'overeducation'. Ever since Duncan and Hoffman's seminal work (1981), much of the previous literature documents mixed evidence and interprets it as evidence for inefficiency and misallocation. To reconcile the contrasting facts, this paper first builds a vertical schooling and occupation sorting model based on a single dimensional human capital index, where education substitutes for ability. Both education and occupation choices are efficient in the theoretical model. I then use simulated data from the calibrated model to show that it reproduces patterns of estimates found in the literature. These estimates are in fact fully consistent with efficient decision making. Finally, I add lifecycle, information frictions and symmetric employer learning to the static model to derive novel and testable implications about the dynamics of education-job match. The paper then turns to the NLSY79 data to demonstrate that empirical evidence in the US from 1982-1994 is consistent with the theoretical model's predictions. Both the theoretical model predictions and the new empirical evidence rationalize the observed overeducation without implications of misallocation.

JEL Codes: I20, I26, J24

Keywords: Human Capital, Employer Learning, Occupational Choice, Mismatch, Overeducation

*Department of Economics, Colgate University, 13 Oak Drive (email: jma@colgate.edu) I am grateful to Fabian Lange, Theodore Papageorgiou, Markus Poschke, for their patience, guidance, encouragement support, and deep generosity of time. Thanks to the conference/seminar participants at EALE, WEAI, CEA, McGill University, Colgate University and CIRANO for their useful comments. All errors are my own.

1 Introduction

Student loan debt in the U.S. has expanded significantly over the past decade or so and stands at historically high levels. As shown in Figure 1, total student loan balances at the end of March 2015 were reported as about 1.2 trillion. The accumulated student debt raises the concern that expensive skills acquired in school may be underutilized in low-paying jobs. For example, a college graduate may work as a barista in Starbucks. According to Figure 2,¹ it is quite common that an individual with a graduate degree works in the same occupation with a high school dropout. Why do workers accept jobs that seemingly do not match their education? How one should interpret this phenomenon relating to education and labor market efficiency? Answers to these questions are of great policy relevance, especially for countries where education is expensive and heavily subsidized. This paper contributes to our understanding of the so-called overeducation phenomenon, or broadly speaking, the within occupation schooling dispersion in three ways. First, I build a educational and occupational sorting model, in which schooling compensates for a worker's lower ability. Education and occupation choices are ex-ante rational and efficient in the model. Then, I calibrate the model and generate simulated data to replicate estimates found in the literature following the seminal work of Duncan and Hoffman (1981). These estimates are often taken as evidence for inefficiency and misallocation. I demonstrate that they are in fact fully consistent with efficient decision making. I provide a theory to explain the mixed empirical facts. Finally, the persistent and transitory nature of education-job match have been widely-documented,² and to my knowledge, are difficult for any previous models to capture. By incorporating information frictions and symmetric employer learning, my model generates dynamic patterns that are consistent with the empirical evidence in the U.S. from 1982-1994.

Ever since Freeman's famous book in 1979, many papers continue documenting the incidence of overeducation across countries and its impact on individual earnings.³ These studies find that returns to an overeducated year are significantly and substantially lower than returns to a required year of education. These findings are taken as evidence that individuals acquire more education than required to perform the tasks on the job. They argue that this represents evidence for mismatch

¹In Figure 2, I group occupations according to the 2-digit OCC1980 code, rank occupations by their mean education, and plot the within-occupation schooling (Q90-Q10 range) using the CPS data.

²Quintini (2011) provides an extensive review of the literature on the persistence of qualification mismatch and suggests that the evidence is mixed: some papers find that overeducation is just a temporary phenomenon that most workers overcome through job mobility, while others find that it is a more stable phenomena, with successful transitions from overeducation to matched job relatively unlikely.

³Leuven et al. (2011) do a meta analysis with 151 existing studies on overeducation.

and inefficient acquisition of education.⁴ Meanwhile, two facts stand out in the empirical work that are not consistent with the misallocation story: (i) unconditionally, there is a positive correlation between ability and education attainment; (ii) conditional on occupation choices, low ability workers tend to have more education than their coworkers.

To reconcile the contrasting evidence, I develop a vertical occupational sorting model. Sorting depends on a single dimensional human capital index, and schooling augments an individual's productivity in a sense that it substitutes for cognitive ability. The occupation-specific wages are output contingent. As in Gibbons and Waldman (1999) and Groes et al. (2014), high human capital workers are assigned to more productive occupations.

The current setup predicts that ability and education correlate negatively conditional on occupation choices. This negative correlation is driven entirely by the vertical sorting mechanism and is the result of efficient decision making. Putting into the overeducation context, this negative correlation implies that overeducation (undereducation) is more common among low (high) ability workers (Chevalier and Lindley, 2009; Green and McIntosh, 2007; Allen and Van der Velden, 2001; Green et al., 1999; Clark et al., 2017).

This model provides an alternative perspective on estimates from the conventional wage regression in the literature. I calibrate the model parameters to match the wage distributions of a series of vertically ranked occupations. With the structural parameters, I then simulate the joint distribution of ability, education attainment, occupational choices and wages. Within my model structure, I find similar patterns of estimates using simulated data as in many other papers using survey data from different countries. My simulation results demonstrate that the lower returns to years of overeducation is mechanically driven by sorting mechanism and by occupation-specific returns to human capital (Lemieux, 2014), which instead are fully consistent with efficiency decision making.

Policy implications of this paper are in sharp contrast to previous studies (Duncan and Hoffman, 1981; Verdugo and Verdugo, 1989; Korpi and Tählin, 2009; Montt, 2017). Had those low ability workers not been able to acquire their current optimal education, they would enter the labor market with lower human capital, consequently sort into low-rank low-pay occupations and be worse off.

This model provides a simple solution to unify the contrasting evidence found by previous studies.

⁴It is conventional in the overeducation literature to follow the footsteps of Duncan and Hoffman (1981)'s specification. The wage specification extends the standard Mincerian equation by dividing individual attained education S_i into three separate components: (i) years of required education S^r , which is occupation-specific; (ii) years of overeducation $S^o = S_i - S^r$ if $S_i > S^r$; and (iii) years of undereducation $S^u = S^r - S_i$ if $S_i \leq S^r$.

But a static human capital theory leaves out the dynamic behavior of education-job match. Recently, the lifecycle dynamics of education-job match has attracted increasing attention (Meroni and Vera-Toscano, 2017; Clark et al., 2017). I thus extend the original model to incorporate information frictions and learning and derive additional implications that can be used to test my model.

Workers are heterogeneous in terms of their cognitive ability. With information frictions, neither the market nor the workers perfectly observe this underlying ability. When employed, worker's productivity performance is revealed, and based on this performance both the market and workers infer unobserved ability. Uncertainty about ability gets resolved with time as in the employer learning literature (Farber and Gibbons, 1996; Altonji and Pierret, 2001; Lange, 2007). Workers can frictionlessly switch occupations upon the arrival of new information.

The dynamic model produces novel and testable implications. The information structure generates negative duration dependence in the exit rates of overeducation and undereducation. New information becomes less precise compared to the prior as labor market experience increases. As a result, the arrival of new information becomes less likely to shift an individual's posterior. Meanwhile, the model predicts selection on unobserved ability in addition to true duration dependence. The overeducation (undereducation) exhibits higher persistence for the low (high) ability workers. I then test these theoretical model implications by exploring the longitudinal patterns in the NLSY79.

This paper is both theoretical and empirical. Theoretically, this paper adds to a large body of overeducation literature by introducing a model of human capital and information frictions that reconcile the contrasting evidence in the literature and disentangles the impact of labor market frictions from ability heterogeneity.

To my knowledge, this is the first paper that explains both static and dynamic patterns of so-called overeducation through a human capital model with information frictions. In the literature, various labor market theories (Human capital, see Sicherman and Galor, 1990, Sicherman, 1991, Kiker et al., 1997;⁵ Job competition, see Gautier et al., 2002;⁶ Heterogeneous preference, see Gottschalk

⁵Sicherman and Galor (1990) and Sicherman (1991) formalize the idea that a worker may prefer to start in a job below his ability / education if this is compensated by an investment opportunity or a higher probability to be promoted. Sicherman (1991) tests the prediction that overschooled workers are more likely to move to higher level occupations. Kiker et al. (1997) find that overeducated workers have faster earning growth with tenure.

⁶Thurow (1976) proposed the job competition theory that wages are solely determined by requirements of the job. One direct implication of the job competition model is that higher educated workers crowd out of lower educated workers during a recession. Gautier et al. (2002) find no support for this theory.

and Hansen, 2003;⁷ Search and friction, see Dolado et al. (2009).⁸) are provided to interpret this phenomenon, but often provide inconsistent evidence and fail to capture its temporary and persistent nature simultaneously. The mechanism in this model shows that the combination of vertical sorting, learning, and human capital formation is sufficient to capture all important empirical regularities, including the cross-sectional patterns of education-job match (individual determinants of overeducation), the dynamics or the career perspective of overeducated (undereducated) workers, and their implications on wages.

Empirically, this paper re-investigates the workhorse wage specification employed in the literature. Many papers have produced estimates using similar specifications (Verdugo and Verdugo, 1989; Dolton and Silles, 2008; Tsai, 2010). The return to years of overeducation is substantially smaller than the return to required years of education.⁹ I demonstrate that these heterogeneous returns may entirely be driven by vertical occupational sorting and occupation-specific returns to education, which does not imply efficiency losses. This is consistent with the idea of Lemieux (2014) that education helps workers get assigned to higher-paying occupations where output is more sensitive to skill and returns to school are higher for higher-paying occupations.

This paper also contributes to the literature on the social value of education and human capital. Testing human capital against job market signaling hypothesis is a difficult endeavor, a fact acknowledged by Lang and Kropp (1986) and Lange and Topel (2006). The conditional negative correlation between ability and overeducation documented in this paper contrasts with the prediction of the pure signaling hypothesis.

The remainder of the paper is organized as follows. Section 2 describes the sample used, and presents contrasting empirical evidence found using the NLSY79 sample. Section 3 provides a solution to the previous inconsistent evidence by introducing a vertical occupation sorting model. Section 4 extends the model to a dynamic setting with information frictions and derives testable implications. Section 5 statistically tests these new model implications. Section 6 briefly discusses other related labor market theories, and Section 7 concludes.

⁷Gottschalk and Hansen (2003) develop a simple model with two sectors (the college and the noncollege sector) and two types of workers (college and noncollege graduates). Workers have heterogeneous preferences with regard to being employed in the two sectors. The heterogeneous preferences lead to the equilibrium in which some college workers voluntarily choose to work in the noncollege sector (being overeducated). The within-sector schooling dispersion in this model does not signal a misallocation of resources or an involuntary assignment.

⁸Dolado et al. (2009) analyze a model with on the job search. In this model, overeducation is a consequence of search friction. Highly-educated workers may end up in unskilled jobs for which they are overqualified but are allowed to engage in on the job search on pursuit of a better job. Skill mismatch has in this model a transitory nature.

⁹It has been noted in the literature that these specifications may suffer from serious measurement error and omitted variable biases, thus can not be interpreted as causal (Leuven et al., 2011).

2 Data, Measurement and Contrasting Existing Evidence

2.1 Data

The main data source in this paper is the NLSY79, a nationally representative sample of 12,686 young men and women who were 14-22 years old when they were first surveyed in 1979. I use the 6111 individuals that comprise the core civilian cross section of the NLSY79, from the 1982 to the 1994 rounds. This results in 79,443 person year observations. I follow the data clean process proposed by Clark et al. (2017). The data is restricted to workers who have permanently entered the labor market.¹⁰ I define permanent entry into the labor market as the first survey year when the individual (1) is in the civilian labor force; (2) works more than 26 weeks out of the year; (3) has reached her highest level of education over the sample period 1982 - 1994; and (4) is not enrolled in school as of May 1st of the survey year.

The main variables of interest are the highest level of completed education, the occupation (measured using the 1980 3-digit Census code) and the hourly wage rate at the time of each interview. Besides these, the variables used in this paper also include age, minority status, gender and place of birth, cognitive and non-cognitive skill measures, geographical location and the corresponding local unemployment rate, family characteristics, a measure of hazards associated with the current occupation and employment history.¹¹

To measure required schooling, I follow Verdugo and Verdugo (1989), Kiker et al. (1997) and Clark et al. (2017). For each given occupation (1980 3-digit Census occupation classification), I compute the average level of education from the pooled monthly samples of the 1989-1991 waves of the Current Population Survey (CPS). Those years were chosen based on the following two considerations. First, they match the waves of the NLSY79 (1982-1994). The match along the time dimension minimizes the extent to which technological change might have altered the education attainment and occupation requirements. Second, pooling three-year data (with an average unemployment rate equals 5.9% between 1989 and 1991) can reduce the impact of business cycles. Meanwhile, in order to obtain required schooling levels that are pertinent to the NLSY79 sample, I restrict the age range

¹⁰In the vertical occupational sorting model, agents enter the labor market are not allowed to go back to school. The permanent entrance restriction is consistent with the model assumption.

¹¹For detailed information on data cleaning and the construction of variables, see Appendix B.

within each year of the CPS to that of the NLSY79 cohorts at that time. Required schooling in a given occupation code is then defined as the average education in that occupation.

To define the over- / under- education, I choose to use a more conservative definition. Specifically, workers are defined to overeducated or undereducated if their education deviates at least 10% from the required education. Not only that the results presented in this paper are robust to the choice of other cutoffs, in the next section, I will show that, despite many critics of the realized match measure I choose in this paper, the realized match measure is consistent with the equilibrium model.

2.2 Conceptual Measurement Issues

The definition of over- / under-education, since it is first introduced has been subject to debate and challenge.¹² Conventionally there have been three ways to measure overeducation, each with its own drawbacks: (i) the objective / job analysis measure;¹³ (ii) the subjective / worker self-assessment measure;¹⁴ (iii) the realized match measure.

The realized match measure uses information from realized worker-occupation matches. In this method, the required amount of schooling for a worker is inferred from the mean / mode of completed schooling of all workers holding the same occupation. A person is over-educated if he/ she has a level of education that is above some arbitrary threshold (related to the required education). Though the use of realized matches is often regarded as inferior because it is the result of demand and supply forces, I choose the realized match measure in this paper. In my model ‘overeducation’ or within-occupation schooling dispersion emerges an equilibrium result of the joint forces of demand and supply. My model and the realized match measure are conceptually consistent. All the model implications hold even without introducing the definition of overeducation. In essence, what the model generates is the within-occupation schooling dispersion as an equilibrium result, wage dynamics, and occupational mobility. Once we formally introduce the overeducation definition to the model setting, this simple model is capable of capturing both the cross-sectional and longitu-

¹²It is explicitly pointed out by Leuven et al. (2011): ‘the conceptual measurement of overeducation has not been resolved’.

¹³In the U.S., Dictionary of Occupation Titles (DOT) and later the Occupational Information Network (O*NET) provide a direct measure of the required level of formal education in an occupation in the form of the General Educational Development (GED) scale by occupational specialists. Updates of this measure are infrequent and not accurate because of the high cost of obtaining such measure.

¹⁴Individuals are asked directly in surveys to report the required level of education needed in their job. This measure is subject to criticism that there may be misreporting by individuals due to inaccurate and incomplete information.

dinal patterns of education-job match, the transitory and persistent nature found in the previous empirical work.

2.3 Descriptive Statistics

Table 1 presents summary statistics for the NLSY79 variables used in my analysis. Detailed data clean and measure construction are described in the Appendix B. Of the final sample, 92.5% are employed, 48% are female, 64.3% are high school graduates, 13% have 2 years and 18.8% have 4 years of college education.

There are substantial variations of required schooling across 3-digit occupations (Figure 3). The minimum required year of schooling is 9 years while the maximum reaches 17 years with some graduate education.

2.4 Contrasting Evidence

Before moving to theory, I redo the empirical analysis that is standard in the overeducation literature. With a different sample, I find mixed evidence similar to many previous papers.

The impact of overeducation, undereducation, and required education on wages has been extensively evaluated in the literature by estimating the following log-wage equation introduced by Duncan and Hoffman (1981):

$$\log(w_{it}) = \alpha^r S_{it}^r + \alpha^o S_{it}^o + \alpha^u S_{it}^u + X_{it}\beta + \epsilon_{it}$$

where, for any given individual i in year t , S_{it}^r , S_{it}^o and S_{it}^u denote respectively the number of years of required schooling, years of overeducation (years of schooling above the required level), and years of undereducation (years of schooling below the required level), X_{it} a vector of controls (including ability measures, socio-demographic background characteristics, labor market experience, and experience square) and ϵ_{it} an idiosyncratic productivity shock. The estimates of the specification, which nests the standard Mincerian wage regression ($\alpha^r = \alpha^o = -\alpha^u$), allows for different returns to the different components of education.

I estimate the above augmented log-wage regression (Duncan and Hoffman, 1981) using the NLSY79

sample, including a full set of controls (See Table 2).¹⁵ The estimation results are in accordance with the previous studies in the literature. Using the full-time full-year sample, I obtain an estimate of the return for an additional year of education about 4.5%, which is less than half of the returns to required education (11%). And I also find substantial wage penalties for years of undereducation (the return for each additional year of undereducation is -7.9%).

Previous literature interprets the ex-post heterogeneity in returns to education as evidence of skill mismatch and inefficiency. They use these results to reject the human capital model which predicts homogeneous return ($\alpha^r = \alpha^o = -\alpha^u$) and to reject the job competition model that wages depend only on job characteristics ($\alpha^o = \alpha^u = 0$). When I re-estimate the workhorse specification with the NLSY79 data, I find patterns similar as in many previous papers, which seemingly suggests the existence of mismatch and misallocation.

I next estimate a probit model to study the individual determinants of overeducation incidence (see Table 7). Cognitive ability, namely, the AFQT score, correlates negatively with being overeducated. The results are statistically significant and robust after controlling for a wide set of variables and across various sub-samples. Using the NLSY79, I find that overeducation is more common among low ability workers.

Like the previous literature, I find inconsistent evidence with the NLSY79 data. How to reconcile these contrasting findings? In the next section, I develop a human capital theory with vertical occupation sorting that comes to the rescue.

3 Full Information Static Model

To make the main point of this paper, I start our discussion with a simple static full information model. In this full information model, individual ability is public information and fully observed.

¹⁵I present results using three different samples: (1) full sample with valid wage data; (2) select workers whose hourly wage above the federal minimum wage levels and below 200 dollars; (3) on the basis of sample 2, select workers working more than 1500 hours yearly (full time full year workers).

3.1 Model and Theoretical Implication

The Economy

1. The economy is occupied with a finite number of occupations, indexed by $k \in \{0, 1, 2, \dots, K\}$. Each of these discrete occupations has some fixed measure γ_k of available jobs. The fixed measure of labor demand is exogenous. Each unit of the good (or service) produced in occupation k sells in the market at some fixed price P_k . The match revenue of a worker i in occupation k is

$$R_{ik} = P_k H_i,$$

where H_i is the human capital of individual i . Occupations are ranked in the order of increasing output prices such that $P_K > P_{K-1} > \dots > P_1 > P_0$. Therefore, any worker with human capital H_i produces more in a high-rank occupation.

2. The linear wage offer contract in this competitive economy is as follows

$$W_{ik}(H_i) = P_k H_i - \Pi_k.$$

This linear function indicates that wages are output-contingent contracts that specify different wages based on the realized match outputs. (Gibbons and Waldman, 1999; Groes et al., 2014). A firm obtains a profit Π_k from any worker who is willing to take this contract. I assume workers are risk-neutral, and they choose the occupation delivering the highest expected wage.

3. The economy has a unit measure of total workers. The relevant sorting criterion for risk-neutral workers is their expected wage given their human capital. Human capital is a function of ability A_i and schooling S_i .

$$H_i = A_i + \alpha f(S_i)$$

where $f'(S_i) > 0$. In this simple model, ability and schooling enter the human capital production function additively. These two components are substitutes. ¹⁶

¹⁶The perfect substitutability is not the key assumption that drives the theoretical model predictions. As long as both ability and schooling contribute positively to an individual's human capital stock, all model predictions will still hold. One can model these two components as complements, which will not change the qualitative predictions. In mathematical terms, the human capital function can be either supermodular or submodular. In my current setup, I choose an additive human capital function mainly because it makes the Bayesian learning in Section 4 more tractable.

Occupation Sorting

In this model, we maintain the assumption that workers cannot go back to school once entering the labor market. Workers differ by their human capital $H_i = A_i + \alpha f(S_i)$. Workers make occupational choices to maximize their incomes based on their human capital and the linear wage contracts. In particular, a worker with human capital H_i chooses occupation k rather than $k-1$ if $W_{ik}(H_i) > W_{i,k-1}(H_i)$. The linear wage contracts are similar to the ones in Gibbons and Waldman (1999), with increasing productivity ($P_K > P_{K-1} > \dots > P_1 > P_0$) and profits ($\Pi_K > \Pi_{K-1} > \dots > \Pi_1 > \Pi_0$). Workers sort into occupation k if their human capital H_i falls in the interval $[B_k, B_{k+1})$, where $B_k \equiv \frac{\Pi_k - \Pi_{k-1}}{P_k - P_{k-1}}$. I illustrate the vertical sorting mechanism based on expected human capital in Figure 4. In Figure 4, high-rank occupations are more productive and pay higher returns to individual human capital (the slopes are steeper). Meanwhile, these occupations also extract higher profits from the worker and occupation matches (the intercepts are negative and large in absolute value). The equilibrium results indicate that workers with higher human capital sort into higher-ranked occupations.

Lemma. *Within-occupation K , i.e. conditional on individual occupational choice k , workers with more schooling on average are lower ability.*

$$\mathbb{E}(A_i | S_L, k) > \mathbb{E}(A_i | S_H, k) \text{ where } S_L < S_H.$$

Workers sort into occupation k if

$$B_k \leq H_i < B_{k+1}.$$

Human capital is a linear combination of two perfectly substitute components, individual ability A_i and schooling S_i .

$$B_k \leq A_i + \alpha f(S_i) < B_{k+1}$$

Given two distinct educational levels, $S_L < S_H$, this immediately implies that

$$\mathbb{E}[A_i | B_k - \alpha f(S_H)] < A_i < B_{k+1} - \alpha f(S_H) < \mathbb{E}[A_i | B_k - \alpha f(S_L)] < A_i < B_{k+1} - \alpha f(S_L).$$

The lemma above establishes a lower level of ability as an individual determinant of overeducation. This is consistent with the empirical evidence that many others (Chevalier and Lindley, 2009; Green and McIntosh, 2007; Allen and Van der Velden, 2001; Green et al., 1999; Clark et al., 2017) and I

find. In the model, workers are efficiently and optimally assigned to vertically ranked occupations according to their human capital stock, a combination of ability and education. In the theoretical model, the so-call overeducation or undereducation is not associated with mismatch, but is fully efficient in the absence of any frictions.

To understand the other piece of contrasting evidence within my model structure, i.e., the patterns of estimates from the augmented Mincerian equation, I first calibrate the static model.

3.2 Calibration of the Static Model

To identify the parameters of this static model, I assume that data are generated by the stationary competitive equilibrium and match the theoretical moments and estimates from the auxiliary regressions to their empirical counterparts.

The calibration proceeds in two steps. First, the joint distribution of ability and education together with the parameters in human capital function are calibrated independently from the NLSY79 data. Then, I retrieve the remaining parameters specifying the wage contracts by indirect inferences.

In the simple static model, individual education decisions are not explicitly modeled. Instead, I assume ability and education are jointly normally distributed with a positive correlation coefficient. I estimate the mean and variance covariance matrix using the NLSY79 sample. The coefficient of schooling in the human capital function is recovered by using the estimates from the standard wage equation $\alpha = \frac{\beta_{Educ}}{\beta_A}$.¹⁷

The main parameters of interest $\{P_1, P_2, \dots, P_K\}$ is the productivity of vertically ranked occupations. I recover these parameters by matching the simulated moments of conditional wage distribution and the estimates from an auxiliary model(the standard Mincerian equation) to their empirical counterparts. For any given productivity profile $\{P_1, P_2, \dots, P_K\}$, the stationary profit profiles of different occupations are sequentially determined by market clearing conditions for all occupation. The model assumes that the demand for labor in occupation k is fixed with a measure of γ_k . Worker sort into occupation k if their human capital H falls into the interval $(B_k, B_{k+1}]$. The labor market clear condition of occupation k is given by the following equation

$$\gamma_k = F(B_{k+1}) - F(B_k) \text{ for any } k \in \{1, 2, \dots, K\}.$$

¹⁷ $\log(w_i) = P(A_i + \alpha S_i) - \Pi + \epsilon_i = \beta_0 + \beta_A A + \beta_{Educ} S_i + \epsilon_i$ the estimation equation corresponds my theoretical model. In this framework, $\beta_A = P$, $\beta_{Educ} = \alpha P$ thus $\alpha = \frac{\beta_{Educ}}{\beta_A}$

For any known human capital distribution $F(H)$ and exogenous labor demand $\{\gamma_1, \dots, \gamma_K\}$, using the above market clear condition, one can recover the cut-off values (B_k) recursively.¹⁸ Finally, we can derive the profit values using the analytic solution of B_k ,

$$\Pi_k = (P_k - P_{k-1})B_k + \Pi_{k-1}.$$
¹⁹

In the first step, we identify the joint distribution of ability and education and the human capital function. With any wage contract parameters, one can easily map the human capital distribution to a wage distribution. In this step, we identify the wage contract parameter (P_k) by matching the occupation-specific mean wages and the coefficients of the standard Mincerian equation to their empirical counterparts.

To simplify the quantitative analysis, rather than using the detailed 3-digit occupations, I rank occupations according to the mean wages and aggregates the detailed 3-digit occupations into 10 occupation groups.²⁰ With the defined synthetic occupational groups, I subsequently compute the employment share for all occupational groups. The employment shares are then taken as the exogenously determined labor demand for each occupational group.

I use 13 empirical moments (constant, returns to education and ability from the standard Mincerian equation, and 10 occupation mean wages) to identify the 10 occupation-specific productivity parameters P_k . In the process, I impose the restriction as in the theoretical model that productivity is non negative and increases with the occupational rank.

The targeted moments and calibrated results are presented in Table 3 and Table 4 respectively. The calibrated productivity parameters are shown in the last column of Table 4. The productivity parameter increases by about 62% from 0.2263 for the lowest ranked occupation to 0.3663 for the highest ranked occupation. The calibrated P_k s provide some preliminary evidence of heterogeneous returns to education across occupations. However, the calibrated values also suggest alternative cutoffs how we should empirically group occupations.²¹

With the structural parameters, I generate the cross-sectional wage distribution and re-estimate the workhorse wage model in this literature. Table 5 presents the estimated results of the Duncan and

¹⁸We start the solution of this problem from occupation $k=1$. The market clear condition for occupation 1 implies that $F(B_1) = \gamma_1$, then $B_1 = F^{-1}(\gamma_1)$. After finding B_1 , one can solve $B_2 = F^{-1}(F(B_1) + \gamma_2)$.

¹⁹This is done with the normalization that $\Pi_0 = 0$.

²⁰For each 3-digit occupation, I compute the mean wage using only full-time full-year workers in that occupation and rank these occupations by their mean wages. Then the 3-digit occupations are divided into 10 occupational groups by the deciles of the mean wage distribution.

²¹Rather than 10 occupational groups, the calibrated values of P_k suggest a deviation of three major groups: (1) low skilled occupations at the bottom 40% of the skilled distribution; (2) intermediate skilled occupations ranging from 40% to 80%; (3) high skilled occupations corresponding to the top 20%.

Hoffman (1981)'s specification. In the second and third row of Table 5, I report the ranges of these estimates found in previous papers and the Mata analysis (Leuven et al., 2011) results respectively. Using my simulated sample, I find a return of 0.078 to a year of education required for the occupation, a return of 0.048 for a year of surplus education and a negative return of -0.024 for each year of deficit education. The return to an overeducated year is significantly and substantially lower than the return to a required year of education. These estimates are within the ranges of corresponding parameter values found in the literature and are closer to the estimates found using the US and Canada data (See the last row of Table 5).²² All the coefficients are precisely estimated. Similarly as many other papers, my estimates reject the null hypothesis $\alpha^r = \alpha^o = -\alpha^u$ at the 1% significance level.

The test of the joint equality ($\alpha^r = \alpha^o = -\alpha^u$) has been interpreted as evidence against the standard human capital model. The ex-post heterogeneous returns to different education components found in many empirical studies are often taken as evidence for mismatch and inefficiency. In sharp contrast to the conventional conclusions in the overeducation literature, my simulation demonstrates that these empirical patterns are in fact fully consistent with rational choices and efficiency allocation.

3.3 Evidence of Occupation-Specific Returns to Education

Besides the criticism made by Leuven et al. (2011) about omitted variable bias and measurement error, my theoretical model and simulation suggests another potential specification error. All the conclusions made by previous studies are based on one implicit and crucial assumption that all occupations reward individual human capital or education identically. My theoretical model suggests otherwise. Higher ranked occupations pay higher returns to education. In this section, I evaluate these alternative assumptions empirically using both the ACS and the NLSY79 data.

To illustrate the point, I first do the exercise with the ACS data leveraging on its big sample size. I plot the occupation-specific returns to education against various occupation rank measures in Figure 5. Each circle represents an occupation at three-digit level(OCC1990), and the circle size represents the sample size used to obtain these estimates. I use different rankings to robustly show that re-

²²In Leuven et al. (2011) survey paper, they tabulate the findings of previous empirical studies. From all the 151 studies, the estimates of returns to required schooling α^r range from 0.043 to 13.5; the estimates of returns to overschooling α^o are mostly positive but tend to be smaller, ranging from -0.031 to 0.054; the estimates of returns to underschooling α^u , on the other hand, are always positive and fall between -0.056 to -0.025.

turns to education or human capital are positively associated with occupation ranks. In Figure 5, occupations are ranked by their average education, average hourly wage rate, and average annual income respectively. The fitted lines indicate that high-rank occupations reward education more. This point is consistent with the findings of Lemieux (2014).

In the model, the linear wage contract ($W_{ik}(H_i) = P_k H_i - \Pi_k$) are specified in a way that high-productivity occupations pay higher returns to human capital and meanwhile extract high profits from the match. This implies that in the Mincerian equation returns to education are negatively correlated with constant terms.²³ To show this, I plot the constant terms in the standard Mincerian wage regression against the returns to education by occupations. The results using the ACS and the NLSY79 are shown in Panel (a) and (b) of Figure 6 respectively. In both samples, the occupation-specific returns and the constant terms are negatively related.

4 Dynamic Model with Information Friction

The static model in the previous section reconciles contrasting empirical evidence found in the literature but leaves out interesting lifecycle dynamics of education-job match. I next add information frictions and lifecycle dynamics to the original model.

4.1 Dynamic Model

The dynamic human capital theory with information frictions captures both the transitory and persistent nature of education-job match. In equilibrium, overeducation (undereducation) is not only more common but also more persistent for low (high) ability workers. Meanwhile, this model differentiates selection on unobservables from true duration dependence. The hazard rates out of overeducation (undereducation) decline as labor market experience increases regardless of individual characteristics, indicating true duration dependence. With uncertainty of ability resolved, the learning component incorporates selection on unobservables in addition to true duration dependence. Unlike previous theories, the dynamic model separates the impact of frictions or mismatch from

²³Gibbons and Waldman (1999) make similar assumption that the coefficient of education is negatively associated with the constant term in the wage equation.

unobserved ability.

Information Structure

The information structure of the economy is defined in the following way.

Each worker has an ability level A_i that is drawn at the beginning of his life from some underlying normal distribution. This is a model without human capital accumulation. This ability remains constant throughout a worker's life. Ability A_i is not perfectly observed due to information friction. Instead, outputs of worker and occupation matches are revealed in every period. Output is a noisy measure of human capital,

$$Y_{it} = H_i + \epsilon_{it} = A_i + \alpha f(S_i) + \epsilon_{it}, \text{ where } \epsilon \sim N(0, \sigma_\epsilon^2).$$

Both firms and workers infer the true human capital using the observed output history $\{Y_{it}\}$. With the assumption that both firms and agents are risk neutral, wages are paid based on the expected value of human capital. Let \mathcal{F}_t denote the information sets at the beginning of period t . The information set contains the complete output history $\{Y_{it}\}$.

$$W_{ik,t}(H_i|\mathcal{F}_t) = P_k \mathbb{E}(H_i|\mathcal{F}_t) - \Pi_k$$

Individual Occupational Choices

In this dynamic extension, risk neutral agents work for T periods. Future incomes are discounted using a fixed market interest rate r . We maintain the assumption that workers are not allowed to change their education levels after permanently entering the labor market. Individual's schooling level S_i is thus fixed and fully observable when workers make their occupational choices.

Let t denote an individual's labor market experience. Given the information set \mathcal{F}_t , workers and firms form a prior belief about individual unobserved ability $A_i|F_t$. The prior belief follows a normal distribution $A_i|F_t \sim N(\mu_{it}, \sigma_t^2)$ because of the simple information structure and the normal Bayesian learning. Both workers and the market then form the expectation of individual human capital $\mathbb{E}(H_i|\mathcal{F}_t) = \mu_{it} + \alpha f(S_i)$. Standard results on updating of normal distributions imply that the belief at the beginning of $t + 1$ evolves as the weighted average of the prior mean μ_{it} and the new output observation y_{it+1} :

$$\mu_{it+1} = \frac{\sigma_\epsilon^2}{\sigma_t^2 + \sigma_\epsilon^2} \mu_{it} + \frac{\sigma_t^2}{\sigma_t^2 + \sigma_\epsilon^2} (y_{it+1} - \alpha f(S_i)).$$

Workers sort into different occupations based on their expected human capital. The wage of an individual i with expected human capital $\mathbb{E}(H_i|\mathcal{F}_t)$ working in occupation k equals

$$W_{ik,t}(H_i|\mathcal{F}_t) = P_k \mathbb{E}(H_i|\mathcal{F}_t) - \Pi_k = P_k(\mu_{it} + \alpha f(S_i)) - \Pi_k.$$

From the point of view of the individual, the evolution of the expected human capital is a martingale with decreasing variance: the weight on the prior increases as more observations have already been observed in the past. Correspondingly, the weight on the most recent observation decreases with years in the labor market.²⁴ The direct implication of the martingale process is that human capital is expected to remain unchanged for the remaining $T - t$ periods. Intuitively, agents would keep their prior beliefs unchanged unless they receive new information. This property simplifies our analysis to a large extent. If an occupation is optimal for the current period, it is also expected to be optimal for the remaining periods unless we have more information to shift the beliefs. Thus, the dynamic decision process degenerates to a static one. We obtain a simple sorting rule similar to the one in a static full information model.

The expected lifetime income for an individual with a prior belief $N(\mu_{it}, \sigma_t^2)$ and schooling S_i , who are currently working in occupation k is

$$\mathbb{E}(I_{ikt}|\mathcal{F}_t) = \beta_t [P_k(\mu_{it} + \alpha S_i) - \Pi_k]$$

where β_t is the multiplier.

Given the profits $\Pi = (\Pi_0, \Pi_1, \dots, \Pi_K)$, workers choose the occupation that maximizes their lifetime expected income. In particular, a worker chooses occupation k rather than $k - 1$ if $\mathbb{E}(I_{ikt}|\mathcal{F}_t) > \mathbb{E}(I_{ik-1t}|\mathcal{F}_t)$. The multiplier drops out since it appears on both sides of the inequality. We obtain a simple sorting rule which is independent of individual labor market experience. Define $B_k \equiv \frac{\Pi_k - \Pi_{k-1}}{P_k - P_{k-1}}$ similarly as in the static model. Workers sort into occupation k if their expected human $\mathbb{E}(H_i|\mathcal{F}_t)$ falls in the interval $[B_k, B_{k+1})$.

²⁴Define a new random variable $m_{it} = \mathbb{E}(H_i|\mathcal{F}_t)$. Then $\mathbb{E}(m_{i(t+1)}|\mathcal{F}_t) = m_{it}$. The property is easy to establish.

$$\begin{aligned} \mathbb{E}(m_{it+1}|\mathcal{F}_t) &= \mathbb{E}[\mathbb{E}(H_i|\mathcal{F}_{t+1})|\mathcal{F}_t] \\ &= \mathbb{E}\left[\frac{\sigma_\epsilon^2}{\sigma_t^2 + \sigma_\epsilon^2} (A_i + \alpha f(S_i)) + \frac{\sigma_t^2}{\sigma_t^2 + \sigma_\epsilon^2} y_{it+1} \middle| \mathcal{F}_t\right] \\ &= \frac{\sigma_\epsilon^2}{\sigma_t^2 + \sigma_\epsilon^2} (\mu_{it} + \alpha f(S_i)) + \frac{\sigma_t^2}{\sigma_t^2 + \sigma_\epsilon^2} \mathbb{E}[y_{it+1}|\mathcal{F}_t] \\ &= \frac{\sigma_\epsilon^2}{\sigma_t^2 + \sigma_\epsilon^2} (\mu_{it} + \alpha f(S_i)) + \frac{\sigma_t^2}{\sigma_t^2 + \sigma_\epsilon^2} \mathbb{E}[A_i + \alpha f(S_i) + \epsilon_{it+1}|\mathcal{F}_t] \\ &= \mu_{it} + \alpha f(S_i) \\ &= m_{it} \end{aligned}$$

Individual Educational Choices

Before entering the labor market, agents choose their schooling level optimally. Initially, every worker only knows that his ability is distributed with mean $\mathbb{E}(A_i|\mathcal{F}_0) = \mu_{i0}$ and variance σ_0^2 . The initial prior belief (the mean) is individual-specific. Without loss of generality, I assume that the mean of individual initial prior is a noisy measure of the true ability $\mu_{i0} = A_i + \epsilon_{i0}$.

I adopt an educational cost function that is common in the literature. Educational cost is an increasing and convex function of education (S), and is higher for low ability workers (μ_{i0}). The cost takes the following functional form,

$$C(S_i, \mu_{i0}) = \frac{cS_i^2}{2\mu_{i0}}.$$

Given the individual-specific prior belief (μ_{i0}), workers choose optimal schooling to maximize their lifetime incomes

$$\max_{S_i} [P_k(S_i)(\mu_{i0} + \alpha f(S_i)) - \Pi_k(S_i)]\beta_T - C(S_i, \mu_{i0})$$

$P_k(S_i)$ and $\Pi_k(S_i)$ are written as function of education because occupation sorting depends on individual education level.

Equilibrium

We consider a standard stationary competitive equilibrium like the one proposed by Groes et al. (2014). As market prices, one can use either profits or wages, as one determines the other. Stationary means that the profit schedule $\Pi = (\Pi_0, \Pi_1, \dots, \Pi_K)$ and the associated wage offers are constant over time.

Let $F(H)$ denote the cross-sectional distribution of beliefs among workers about their human capital level. $F(H)$ is the measure of workers with belief below H across all cohorts at any point of time. $F(H)$ can be computed prior to any analysis of occupational choice. This simplifies our specification of an equilibrium.

Definition *Given the productivity schedule $P = (P_0, \dots, P_K)$ and the exogenous labor demand $\gamma = (\gamma_0, \dots, \gamma_K)$, an equilibrium is a vector of profits $\Pi = (\Pi_0, \dots, \Pi_K)$ and a vector of optimal cutoff (B_1, B_2, \dots, B_K) such that the following equations hold*

$$B_k \equiv \frac{\Pi_k - \Pi_{k-1}}{P_k - P_{k-1}} \text{ for } k \in \{1, 2, \dots, K\}$$

$$\gamma_k = F(B_{k+1}) - F(B_k) \text{ for } k \in \{1, 2, \dots, K\}.$$

4.2 Model Implications

In my setup, within-occupational schooling dispersion appears an equilibrium result and is consistent with ex-ante efficient decision making. Obtaining education is more costly for workers with a lower level of ability, the an-ante optimal education decision predicts a positive correlation between ability and educational attainment. Sorting into vertically ranked occupations depends instead on human capital, a combination of ability and education. The vertical sorting mechanism generates a conditional negative correlation between ability and overeducation, which is confirmed by previous studies.

This model also produces some novel implications regarding the dynamics of education-job matches. With Bayesian updating, switching probability declines since new information gradually becomes less precise. The declining switching probability holds regardless of individual ability, indicating a true duration dependence. However, as true ability revealed, the overeducation (undereducation) becomes increasingly persistent for low (high) ability workers. Learning in the current setup leads to selection on unobservable ability.

Proposition 1. *More able workers sort into high-pay high-rank occupations.*

Proof. Cost of obtaining an additional unit of education is lower for higher ability workers. High ability workers obtain more schooling. Schooling and ability enter additively into human capital function. Expected human capital is an increasing function of ability. The vertical occupation sorting rules make sure that more able individuals have more human capital and sort into higher-rank occupations. See detailed proof in Appendix C. \square

Proposition 2. *Conditional on occupational choice k , and labor market experience t , overeducated workers on average are lower ability. For two distinct educational levels $S_L \leq S_H$, this means*

$$\mathbb{E}(A_i | S_L, t, k) > \mathbb{E}(A_i | S_H, t, k) \quad \forall t$$

Proof. For workers with t years of experience, they sort into occupation k if $B_k \leq \mathbb{E}(H_i | \mathcal{F}_t) < B_{k+1}$. The expected human capital depends on the true signal and a noise term e_t

$$\mathbb{E}(H_i|\mathcal{F}_t) = A_i + \alpha f(S_i) + e_t,$$

where e_t is a weighted average of all past normal errors $e_t = \frac{\sigma_\epsilon^2}{t\sigma_0^2 + \sigma_\epsilon^2} \epsilon_0 + \frac{\sigma_0^2}{t\sigma_0^2 + \sigma_\epsilon^2} (\sum_{s=1}^t \epsilon_s)$.²⁵ By the law of iterated expectation,

$$\begin{aligned} \mathbb{E}(A_i|S_i, t, k) &= \mathbb{E}_{e_t}[\mathbb{E}(A_i|S_i, t, k, e_t)] \\ &= \mathbb{E}_{e_t}[\mathbb{E}[A_i|B_k - (e_t + \alpha f(S_i)) < A_i < B_{k+1} - (e_t + \alpha f(S_i))]] \end{aligned}$$

For any given e_t ,

$$\begin{aligned} \mathbb{E}[A_i|B_k - (e_t + \alpha f(S_H)) < A_i < B_{k+1} - (e_t + \alpha f(S_H))] \\ < \mathbb{E}[A_i|B_k - (e_t + \alpha f(S_L)) < A_i < B_{k+1} - (e_t + \alpha f(S_L))]. \end{aligned}$$

The above inequality holds because the integration domain for workers with more education is below that of undereducated worker. By the law of iterated expectation, the inequality maintains after integrating out e_t . □

Proposition 3. *Overeducation (undereducation) is more persistent for low (high) ability workers.*

In other words, the hazard rates out of overeducation and undereducation depend on an individual's unobserved ability. Selection depends on unobservables.

Let P_{kt}^+ denote the probability that a worker switches to a higher occupation and out of overeducation. $P_{kt}^+ = \text{Prob}(\mathbb{E}(H_{it+1}) > B_{k+1}|\mathcal{F}_t)$. Conditional on being overeducated, the model predicts that $\frac{\partial P_{kt}^+}{\partial A_i} > 0$. High-ability workers have a higher exiting probability out of overeducation.

Similarly, I can define $P_{kt}^- = \text{Prob}(\mathbb{E}(H_{it+1}) \leq B_k|\mathcal{F}_t)$ as the switch-down probability that results in out of undereducation. Undereducation exhibits higher persistence for high ability workers that $\frac{\partial P_{kt}^-}{\partial A_i} < 0$.

Proof. See Appendix C. □

Proposition 4. *The hazard rate out of overeducation (undereducation) is decreasing in labor market experience t .*

²⁵ e_t is normally distributed $e_t \sim N(0, \frac{(t+1)\sigma_\epsilon^2\sigma_0^2}{t\sigma_0^2 + \sigma_\epsilon^2})$ and is orthogonal to the signal A_i .

Proof. As labor market experience t increases, the new information becomes less precise compared to the prior. It becomes increasingly difficult to shift the posterior belief and trigger changes in over-/under- education status. This proposition holds regardless of individual’s ability level, which is consistent with a true duration dependence. See Appendix C for detailed proof.

□

5 Test Model Implications: Empirics

In this section, I test the above model propositions using the NLSY79 data.

Test Proposition 1

To robustly show that high-ability workers sort into high-rank occupations, I use multiple rank measures.

In the model, occupations are ranked by their productivity, which unfortunately is not observed in the data. Instead, I empirically rank occupations by their average incomes. The linear wage contract establishes a rank-preserving mapping from productivity to average income at the occupational level. In Figure 7 and Figure 8, I plot the average cognitive ability against average education and income scores. Each dot represents an occupation at the three-digit level (OCC1980).

Figure 7 shows a strong positive relationship between average ability and average education across occupations, which is consistent with the model prediction that high-ability workers obtain more education. These workers then sort into higher ranked occupations and on average earn more as shown in Figure 8.²⁶ The positive relationship found in Figure 8 are robust to the chosen rank measures. The CPS provides multiple other measures of occupational ranking, e.g. Duncan Socioeconomic Index (SEI), Siegel score, Nakato and Treas prestige score, and Nam-Powers-Boyd (NPB) Status score. Detailed information about these occupation ranking measures can be found in Table 6. Table 6 presents the correlation coefficients between average AFQT and different occupation rankings. For all rank measures used, we find strong evidence of positive sorting depending on cognitive ability.

As a falsification test, in Appendix D, I plot the rank measure against other non-cognitive measures.

²⁶In Figure 8, I use income score as the continuous measure of occupational ranking. Income Score is a constructed variable that assigns continuous scores to each occupation. Income Score assigns each occupation a value representing the median total income (in hundreds of 1950 dollars) of all persons with that particular occupation in 1950. Income Score thus provides a continuous measure of occupations, according to the economic rewards.

No strong relationship is detected, which justifies my model choice that the occupational sorting relies mainly on individual cognitive ability.

Test Proposition 2

To test proposition 2, I report in Table 7 the estimation results from a Probit model, which intend to identify the individual determinants of the overeducation incidence. The probit model allows the overeducation status to depend on a set of socio-demographic characteristics, family characteristics, employment history, occupation characteristics, and most important cognitive and non-cognitive ability measures. I present the results from the pooled cross-sectional and also from stratified regression to show the heterogeneous effect across the skill distribution.

AFQT scores exhibit a negative and significant relationship with overeducation across all samples. This negative relationship between negative ability and the likelihood of overeducation, which is in line with prior findings in the literature (see, e.g. Green and Zhu 2010 and Chevalier and Lindley, 2009), is consistent with ability and schooling attainment being substitutes. Interestingly, the impact of AFQT increases as we move to the right of skill distribution, with the effect being largest for the most skilled (-1.15) in the last 2 columns of Table 7. The AFQT effect for the most skilled is four times bigger than what we find using the full sample. The substitutability between education and ability is strongest among the most competent workers, which is consistent with the superstar dropout story. Non-cognitive ability measured by the Rotter and Sociability score, on the other hand, does not significantly predict overeducation. None of these coefficients is statistically significant. Work Experience, except for the most skilled workers, negatively predicts overeducation. The direction of this effect is in accordance with the theory that overeducation is partially frictional and transitory.

As mentioned in the introduction section, the model, in essence, generates within-occupation dispersion without relying on any particular definition of over- / under- education. In other words, the model implication is robust to arbitrary thresholds chosen. As a robustness check, instead of indicators for overeducation, I analyze the determinants of the within-occupation schooling ranking. The within-occupation ranking measure is generated as a real number between 0 and 1, equal to the percentage of schooling in its frequency distribution that are equal to or lower than it.

To analyze the determinants of the within-occupation schooling ranking, I estimate the following

linear model that includes individual, occupational, and economic characteristics, X , as controls:

$$y = \beta_0 + \beta_1 AFQT + X\Gamma + \epsilon,$$

where y is the within-occupation percentile rank of schooling. X includes a set of socio-demographic characteristics, family characteristics, employment history, occupational characteristics, and macroeconomic indicators. My theoretical model predicts that β_1 is negative. As predicted by the model, the effects of AFQT presented in Table 8 are negative across all samples with the largest effect found for the most skilled workers. These results add to the previous probit estimation that the previous result is not driven by my definition of overeducation.

Test Proposition 3

In the following section, I will explore the longitudinal dimension of the NLSY79 in detail to show that low ability workers stay overeducated longer.

I begin by showing some descriptive evidence first. I report the hazard rates out of overeducated employment as a function of the duration of overeducation and ability (see Figure 9).²⁷ The hazard rate out of overeducation drops by more than 50% during the first 5 years spent overeducated. Further, the hazard rate out of overeducation is higher for workers with AFQT score above the mean. To illustrate this idea more formally, I estimate the effect of cognitive ability on the spell length of overeducation,

$$y_i = \beta_0 + \beta_1 AFQT_i + X_i\Gamma + \epsilon_i$$

where y in this specification is the spell length of overeducation, and X_i only includes time invariant covariants, including a set of socio-demographic characteristics, non-cognitive ability measures and other family characteristics.²⁸

The estimation results are presented in Table 9. For overeducation spell, I include results using the full sample with all complete spells and the restricted sample with only the first complete spell for

²⁷The hazard rates are computed as the number of individuals leaving overeducated employment at t , divided by the number of individuals who are still overeducated at the beginning of t

²⁸The unit of measure in the above specification is a spell. An individual can potentially have multiple overeducation spells.

each individual. Across all specification and all samples, AFQT scores exhibit a negative and significant relationship with the length of overeducation spell. The results are robust after controlling for demographic and regional variables. In our preferred specification with the most extensive controls, we find the effect of AFQT on overeducation spell is stronger for the restricted sample. Overeducation spell is 6 months shorter if a worker’s AFQT score increases by one standard deviation.²⁹ This effect is larger (about 8.5 months shorter) if we focus only on the first overeducation spell.³⁰

I repeat the above exercise using the undereducation spells in the sample. My theoretical model predicts longer undereducation spells for high ability workers ($\frac{\partial \text{Undereducation Spell}}{\partial \text{AFQT}} > 0$). The effects of AFQT in Table 10 have the predicted positive sign, but are not always statistically significant.

Rotter score as a measure of non-cognitive ability is not a determinant of either over- or under-education spell. However, an individual’s high sociability predicts shorter overeducation spells, providing suggestive evidence that selection out of overeducation also depends on an individual’s communicative skills. Lastly, all my specification in Table 9 and Table 10 provide strong supporting evidence that there is scarring effects of entering the labor market during a recession as the high unemployment unambiguously leads to a longer mismatch period (Kahn, 2010; Oreopoulos et al., 2012; Altonji et al., 2016; Liu et al., 2016).³¹ The estimates indicate that high unemployment rates when entering the mismatch occupations predicts longer mismatch spells. Overeducation is likely to be more persistent when workers entering the mismatch during recessions.

Test Proposition 4

The results discussed above provide some suggestive evidence of duration dependence, with a strongly decreasing hazard rate out of overeducation. To study this duration dependence formally, in addition to the previous graphic illustration, I assume that the duration of the over- and under- education is determined by a proportional hazard model, a Cox proportional hazard model, where the baseline duration is estimated non-parametrically. I estimate the following model:

$$\lambda(t|X_i) = \lambda_0(t)exp(X_i\beta),$$

²⁹The standard deviation of AFQT is 0.27 (See Table 1). The last two columns of Table 9 shows that the coefficient of AFQT equals -1.921. Hence, if someone increases his AFQT by one standard deviation, this reduces his overeducated spell by about 6 months ($\approx 0.27 * (-1.921) * 12$).

³⁰In our sample, the average of individual first overeducation spell equals 4.25 years or equivalently 51 months.

³¹The unemployment rates entering these regression are unemployment rates at the beginning of each over-/under-education spells.

where $\lambda(t|X_i)$ denotes the baseline hazard rate, and X_i includes only time invariant covariants (socio-demographic characteristics, cognitive and non-cognitive ability measure, and family characteristics).

The non-parametric estimates of the hazards out of over- / under- education are shown in Figure 10 and Figure 11. These figures indicate that the hazard rate out of overeducation (undereducation) is strongly decreasing in the duration even after controlling for observed heterogeneity. In other words, while part of this negative duration dependence is attributable to selection on observables, the exit rate is strongly decreasing with the duration of mismatch spell (true duration dependence) after controlling for an extensive set of observed characteristics. The model predicts that hazard rate out of overeducation (undereducation) is decreasing as the mismatch spell increases since new information becomes less precise compared to the prior. The true duration dependence predicted by the model is consistent with the estimates of the Cox model.

In Table 11, I present the estimated results of two Cox models for both over and under education. The coefficients on observable characteristics provide additional evidence to support Proposition 3. The positive effect of AFQT scores on overeducation exit carries over from the descriptive analysis into the proportional hazard model after controlling for a full set of observables. The AFQT coefficient for undereducation is negative when the full set of observable characteristics is included. This is consistent with the idea that undereducation is more persistent for high ability workers.

6 Other Related Labor Market Theories

6.1 Testing Against Pure Signaling Hypothesis

The empirical part of this paper also contributes to the ongoing debate between Becker's theory of human capital and Spence's signaling approach towards education. The fundamental difficulty in distinguishing between signaling and human capital theory is that both models imply an unconditional positive correlation between years of education and labor market earnings.

The paper finds that a lower level of ability is a statistically significant determinant of overeducation (see Table 7). This result provides additional evidence against the pure signaling hypothesis in Spence's original paper. The signaling story assumes that education only serves as a signal for otherwise unobservable abilities, importantly without increasing a worker's productivity. According

to the theory of signaling, the only function of education is to allocate more productive workers to more productive jobs. Then, the signaling theory would predict that a higher level of cognitive ability (measured by the AFQT score) is a determinant of overeducation. The empirical results of this paper reject the pure signaling hypothesis and are consistent with a human capital model where education substitutes for cognitive ability.

To push this direction further, I exploit the dynamics of ability as a determinant of the overeducation incidence. The current human capital model with information friction implies that a lower level of ability becomes a stronger predictor of overeducation as the true ability revealed with labor market experience. To formally show this, I estimate a similar probit as in Table 7 but add an additional explanatory variable, the interaction term of ability and labor market experience.

$$Prob(y = Overeducation) = \beta_0 + \beta_1 AFQT + \gamma(AFQT * Exp) + X\Gamma + \epsilon$$

The information friction and human capital theory predict that γ is negative. As the true ability revealed, ability becomes a more negative predictor of overeducation. The estimation results of the above equation are presented in Table 12. As predicted by the theory of employer learning, the point estimator of γ is always negative, even though not precisely estimated.

This would not be true under the pure signaling assumption. Education acting as a signal of unobserved ability allocates high ability workers to higher ranked occupations. In a simple setting, this implies at least a positive (separating equilibrium) or zero (pooling equilibrium) within-occupation correlation between education and ability. The negative coefficients of the AFQT score are inconsistent with the prediction of a pure signaling model.

6.2 Brief Discussion of Other Related Labor Market Theories

In the literature, various labor market theories (Human capital, see Sicherman and Galor, 1990, Sicherman, 1991, Kiker et al., 1997; Job competition, see Gautier et al., 2002; Heterogeneous preference, see Gottschalk and Hansen, 2003; Search and friction, see Dolado et al., 2009.) are provided to interpret the overeducation phenomenon, but often with inconsistent conclusions and evidence.

The career mobility theory (stepping stone) are consistent with the empirical fact that overeducated workers are more likely to move to a higher level-occupation than workers with the same level of attained schooling who are not over or undereducated (Sicherman and Galor, 1990). However, it

fails to explain why undereducated workers are even more likely to be promoted (with the effect being even bigger than that of the overeducation) which is also documented in Sicherman and Galor (1990). This paper reconciles these puzzling facts by modeling ability as a determinant of over-/under- education. Search frictions fail to capture the persistence in overeducation. Individual persistence and duration dependence of overeducated employment are well documented in the literature (Clark et al., 2017; Meroni and Vera-Toscano, 2017; Mavromaras and McGuinness, 2012). However, if overeducation is due only to search frictions, one would expect this type of mismatch to be transitory and concentrated early in one’s career. It can not explain why overeducation is persistent for experienced and low ability individuals. Preference heterogeneity, on the other hand, generates persistence in overeducation patterns but leaves out the empirical regularity that a fraction of educational mismatch is temporary.

Unlike the previous theories, I use information frictions to generate transitory overeducation or undereducation. I follow the recent literature that has argued that occupational mobility and switch is largely due to information frictions (e.g. Pastorino, 2013; Papageorgiou, 2014; Groes et al., 2014). To my knowledge, this is the only paper that uses information frictions and learning to explain the dynamics of education-job match over lifecycle. With vertical occupational sorting based on individual human capital stock, this paper is also the first one that attempts to separate the impact of frictions from unobserved ability. The current model setup and predictions are also broadly consistent with the assimilation trajectory of immigrants after arrival.³²

7 Conclusion

Economists and policy-makers have long been concerned about the determinants and wage impacts of so-called overeducation. However, previous studies find mixed empirical evidence, and there lacks consensus in the literature regarding how to interpret the inconsistent evidence, the underlying mechanism, and the related implication on efficiency losses.

In this paper, I first propose a human capital theory with vertical occupational sorting to explain

³²Overeducation and underemployment is a common phenomenon among immigrants. Ability heterogeneity in literacy accounts for a portion of the observed overeducation or underemployment (Ferrer et al., 2006). As their local labor market experience increases in the destination nations, a fraction of immigrants workers gradually find jobs that match their education. One can explain the assimilation trajectory either through human capital accumulation, learning the true underlying ability of immigrants, or the combination of both mechanisms. My current setup can be extended to incorporate human capital accumulation through learning-by-doing.

the inconsistency, in which education substitutes for an individual's cognitive ability. To further test my model, I then add information frictions and symmetric employer learning with experience to the baseline model. The model generates novel and testable implications regarding the lifecycle dynamics of education-job match. Finally, I turn to the NLSY79 to show that empirical evidence in the US is consistent with the theoretical model predictions.

In the previous studies, the patterns obtained by estimating the augmented Mincerian equation (Duncan and Hoffman, 1981) are commonly taken as evidence for misallocation and mismatch. However, previous papers also find individuals with a lower level of ability are more likely to be overeducated. In my setup, workers sort into vertically ranked occupations based on their human capital stock, in which ability and education are perfect substitutes. The model predicts that overeducation is more common for low ability workers. I then calibrate the model and show by simulation that the patterns of estimates found in previous papers are in fact fully consistent with efficient decision making.

I then extend the model to a dynamic one with information frictions. Uncertainty about ability is resolved when both workers and the market learn their true ability. This dynamic model generates some interesting predictions about the lifecycle dynamics of education-job match. The model predicts true duration dependence that hazard rates out of overeducation (undereducation) decline regardless of individual characteristics. Meanwhile, the learning about true ability incorporates selection on unobservables in addition to true duration dependence that workers with a lower level of ability tend to stay overeducated longer. I test these new predictions and validity of my model using the longitudinal patterns in the NLSY79.

Cross-sectional, longitudinal patterns of education-job match, and patterns of wage estimates are in fact fully consistent with ax-ante efficient decision making. Policy implications of this paper are in sharp contrast to previous studies and instead alleviate the public concerns regarding so-called overeducation. Had those low ability workers obtained less education, they would end up in lower ranked occupation and be worse off.

The observed within-occupation schooling dispersion, namely, the so-called overeducation and undereducation, is a complex phenomenon. This paper contributes to our understanding of this by introducing a human capital model with information frictions that explains the inconsistent empirical evidence in a uniform framework. To my knowledge, this is the first paper that uses information frictions and learning to disentangle impact of market imperfectness from ability heterogeneity in

the overeducation context.

There might be other labor market theories that can potentially explain this phenomenon. In particular, a possible competing theory can be a multidimensional mismatch model with search friction (Postel-Vinay et al., 2015). In the following work, first, I would like to explore the data more to see if I can statistically test against an alternative model with search friction. Second, from a policy standpoint, information friction in this paper contributes to the frictional overeducation. It would be interesting in the next step to quantify its impacts prior to making further policy suggestions.

References

- Jim Allen and Rolf Van der Velden. Educational mismatches versus skill mismatches: effects on wages, job satisfaction, and on-the-job search. *Oxford Economic Papers*, 53(3):434–452, 2001.
- Joseph G Altonji and Charles R Pierret. Employer learning and statistical discrimination. *The Quarterly Journal of Economics*, 116(1):313–350, 2001.
- Joseph G Altonji, Lisa B Kahn, and Jamin D Speer. Cashier or consultant? entry labor market conditions, field of study, and career success. *Journal of Labor Economics*, 34(S1):S361–S401, 2016.
- Harminder Battu, Clive R Belfield, and Peter J Sloane. Overeducation among graduates: a cohort view. *Education Economics*, 7(1):21–38, 1999.
- Felix Büchel and Matthias Pollmann-Schult. Overeducation and skill endowments: the role of school achievement and vocational training quality. 2001.
- Arnaud Chevalier and Joanne Lindley. Overeducation and the skills of uk graduates. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(2):307–337, 2009.
- Brian Clark, Clément Joubert, and Arnaud Maurel. The career prospects of overeducated americans. *IZA Journal of Labor Economics*, 6(1):3, 2017.
- Juan J Dolado, Marcel Jansen, and Juan F Jimeno. On-the-job search in a matching model with heterogeneous jobs and workers. *The Economic Journal*, 119(534):200–228, 2009.
- Peter Dolton and Anna Vignoles. The incidence and effects of overeducation in the UK graduate labour market. *Economics of Education Review*, 19(2):179–198, 2000.
- Peter J Dolton and Mary A Silles. The effects of over-education on earnings in the graduate labour market. *Economics of Education Review*, 27(2):125–139, 2008.
- Greg J Duncan and Saul D Hoffman. The incidence and wage effects of overeducation. *Economics of Education Review*, 1(1):75–86, 1981.
- Henry S Farber and Robert Gibbons. Learning and wage dynamics. *The Quarterly Journal of Economics*, 111(4):1007–1047, 1996.

- Ana Ferrer, David A Green, and W Craig Riddell. The effect of literacy on immigrant earnings. *Journal of Human Resources*, 41(2):380–410, 2006.
- Pieter A Gautier, Gerard J Van den Berg, Jan C Van Ours, and Geert Ridder. Worker turnover at the firm level and crowding out of lower educated workers. *European Economic Review*, 46(3): 523–538, 2002.
- Robert Gibbons and Michael Waldman. A theory of wage and promotion dynamics inside firms. *The Quarterly Journal of Economics*, 114(4):1321–1358, 1999.
- Peter Gottschalk and Michael Hansen. Is the proportion of college workers in noncollege jobs increasing? *Journal of Labor Economics*, 21(2):449–471, 2003.
- Francis Green and Steven McIntosh. Is there a genuine under-utilization of skills amongst the over-qualified? *Applied Economics*, 39(4):427–439, 2007.
- Francis Green and Yu Zhu. Overqualification, job dissatisfaction, and increasing dispersion in the returns to graduate education. *Oxford Economic Papers*, 62(4):740–763, 2010.
- Francis Green, Steven McIntosh, Anna Vignoles, et al. Overeducation and skills-clarifying the concepts. Technical report, Centre for Economic Performance, LSE, 1999.
- Fane Groes, Philipp Kircher, and Iouri Manovskii. The u-shapes of occupational mobility. *The Review of Economic Studies*, 82(2):659–692, 2014.
- Lisa B Kahn. The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics*, 17(2):303–316, 2010.
- Billy Frazier Kiker, Maria C Santos, and M Mendes De Oliveira. Overeducation and undereducation: evidence for portugal. *Economics of Education Review*, 16(2):111–125, 1997.
- Tomas Korpi and Michael Tåhlin. Educational mismatch, wages, and wage growth: Overeducation in sweden, 1974–2000. *Labour Economics*, 16(2):183–193, 2009.
- Kevin Lang and David Kropp. Human capital versus sorting: the effects of compulsory attendance laws. *The Quarterly Journal of Economics*, 101(3):609–624, 1986.
- Fabian Lange. The speed of employer learning. *Journal of Labor Economics*, 25(1):1–35, 2007.

- Fabian Lange and Robert Topel. The social value of education and human capital. *Handbook of the Economics of Education*, 1:459–509, 2006.
- Thomas Lemieux. Occupations, fields of study and returns to education. *Canadian Journal of Economics/Revue Canadienne d'Économique*, 47(4):1047–1077, 2014.
- Edwin Leuven, Hessel Oosterbeek, et al. Overeducation and mismatch in the labor market. *Handbook of the Economics of Education*, 4:283–326, 2011.
- Kai Liu, Kjell G Salvanes, and Erik Ø Sørensen. Good skills in bad times: Cyclical skill mismatch and the long-term effects of graduating in a recession. *European Economic Review*, 84:3–17, 2016.
- Kostas Mavromaras and Seamus McGuinness. Overskilling dynamics and education pathways. *Economics of Education Review*, 31(5):619–628, 2012.
- Elena Claudia Meroni and Esperanza Vera-Toscano. The persistence of overeducation among recent graduates. *Labour Economics*, 48:120–143, 2017.
- Guillermo Montt. Field-of-study mismatch and overqualification: labour market correlates and their wage penalty. *IZA Journal of Labor Economics*, 6(1):2, 2017.
- Philip Oreopoulos, Till Von Wachter, and Andrew Heisz. The short-and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics*, 4(1):1–29, 2012.
- Theodore Papageorgiou. Learning your comparative advantages. *Review of Economic Studies*, 81(3):1263–1295, 2014.
- Elena Pastorino. Careers in firms: estimating a model of job assignment, learning, and human capital acquisition. Technical report, 2013.
- Fabien Postel-Vinay, Jeremy Lise, et al. Multidimensional skills, sorting, and human capital accumulation. In *2015 Meeting Papers*, number 386. Society for Economic Dynamics, 2015.
- G Quintini. Over-qualified or under-skilled. a review of existing literature. OECD social, employment and migration, 2011.
- Nachum Sicherman. "overeducation" in the labor market. *Journal of Labor Economics*, 9(2):101–122, 1991.

Nachum Sicherman and Oded Galor. A theory of career mobility. *Journal of Political Economy*, 98 (1):169–192, 1990.

Lester C Thurow. *Generating inequality*. Macmillan London, 1976.

Yuping Tsai. Returns to overeducation: A longitudinal analysis of the us labor market. *Economics of Education Review*, 29(4):606–617, 2010.

Richard R Verdugo and Naomi Turner Verdugo. The impact of surplus schooling on earnings: Some additional findings. *Journal of Human Resources*, pages 629–643, 1989.

Appendix A

Figure 1: Student Loan

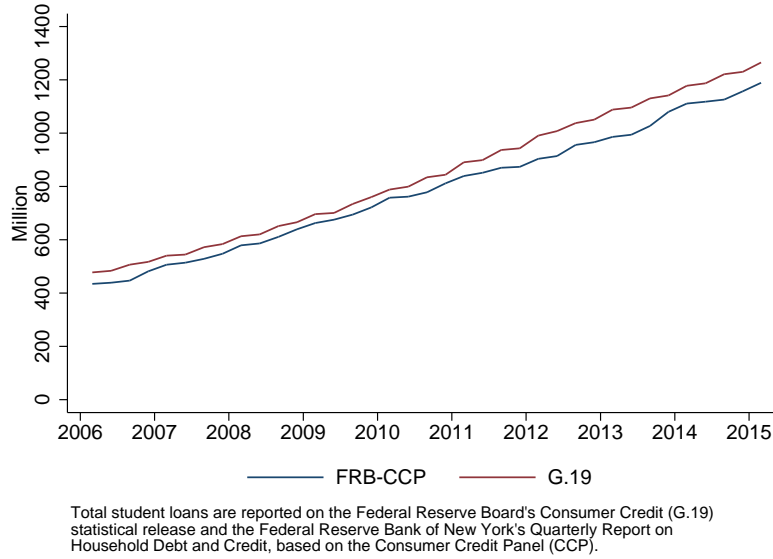
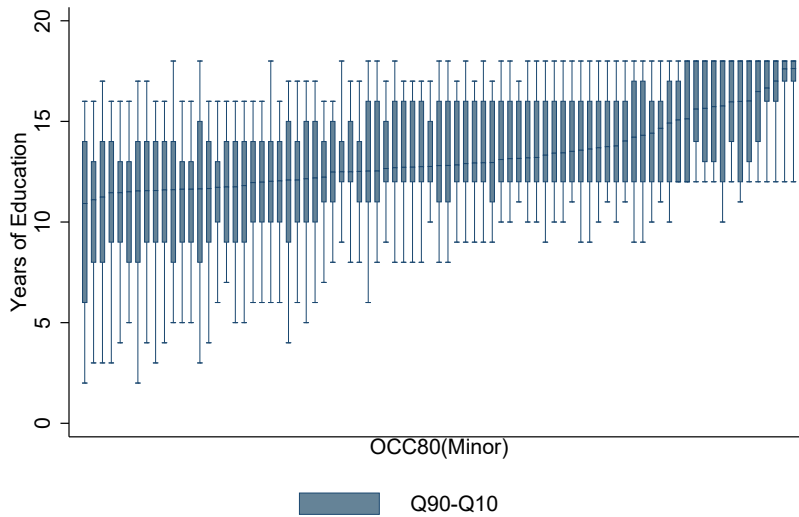
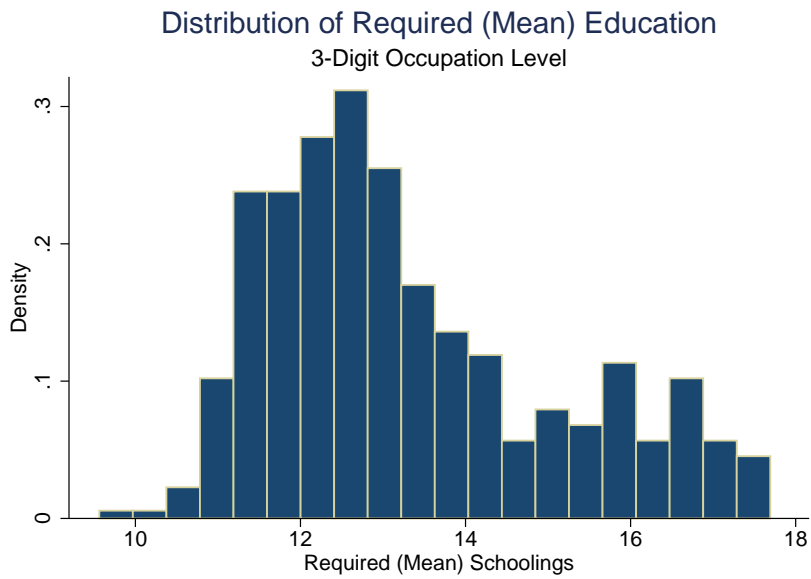


Figure 2: Within-Occupation Schooling Dispersion



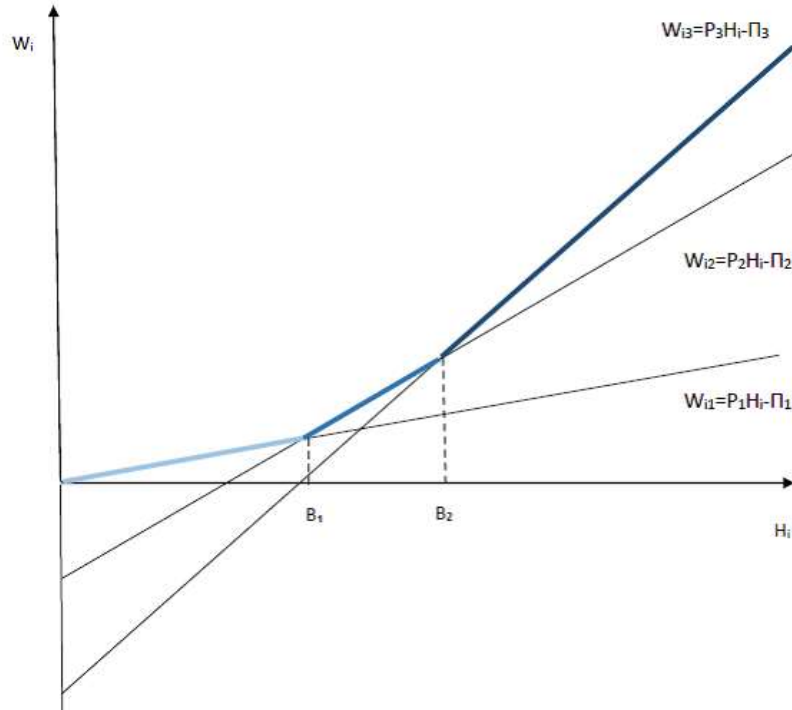
I group occupations according to the 2-digit OCC1980 code, rank occupations by their mean education, and plot the within-occupation schooling using the CPS data. The blue bar indicates the Q90-Q10 range for each occupational group. Outliers are excluded from this box plot.

Figure 3: Distribution of Required Education (CPS)



For each given occupation (1980 3-digit Census occupation classification), I compute the required level of education from the pooled monthly samples of the 1989-1991 waves of the Current Population Survey (CPS). Those years are chosen to match my main data, the NLSY79 (1982-1994), and to reduce the impact of business cycles by pooling data from different years. In order to obtain required schooling levels that are pertinent to the NLSY79 sample, I further restrict the age range within each year of the CPS to that of the NLSY79 cohorts at that time to minimize the possible cohort effect. Required schooling in a given occupation code is then defined as the mean of the distribution of the levels of education among the individuals working in that occupation.

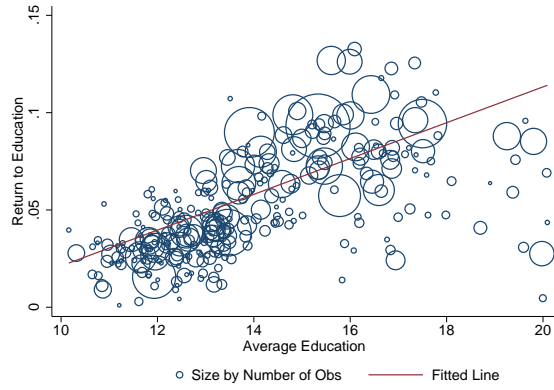
Figure 4: Occupational Sorting



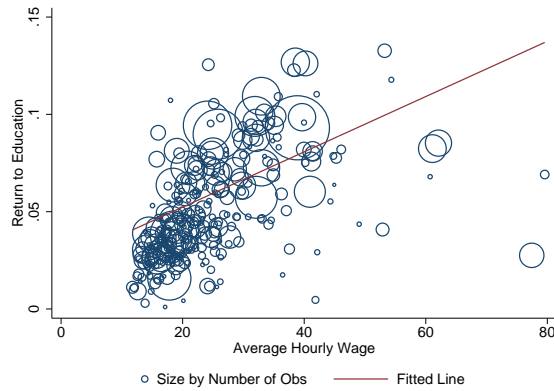
Occupations are ranked according to their productivity, with occupation 3 being the most productive occupation. Occupation 3 pays the highest return to human capital with the steepest slope and retains the largest profit (the lowest intercept) from the work-occupation match. In equilibrium, workers with higher human capital sort into higher-ranked occupations.

Figure 5: ACS Evidence of Heterogeneous Returns to Education

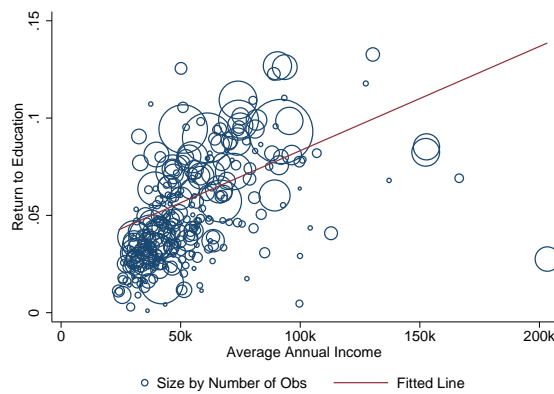
(a) Occupations Ranked by Average Education



(b) Occupations Ranked by Average Wage



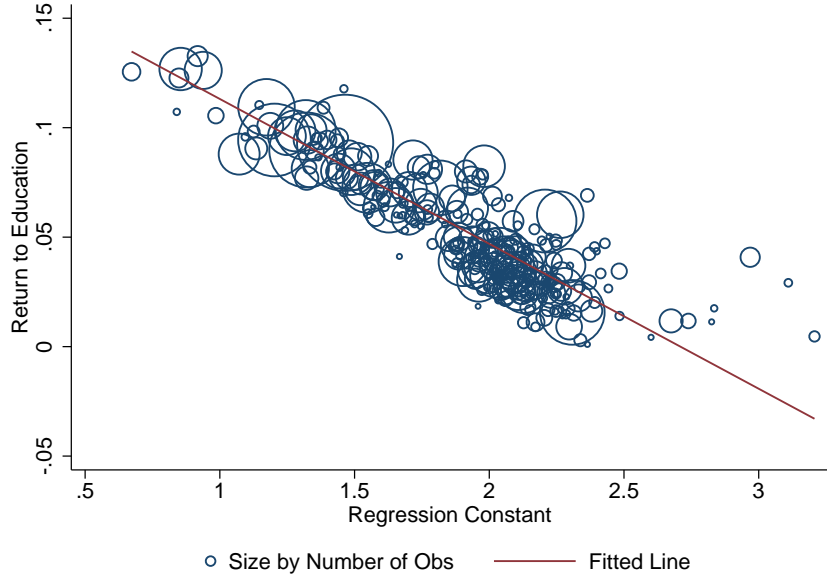
(c) Occupations Ranked by Average Income



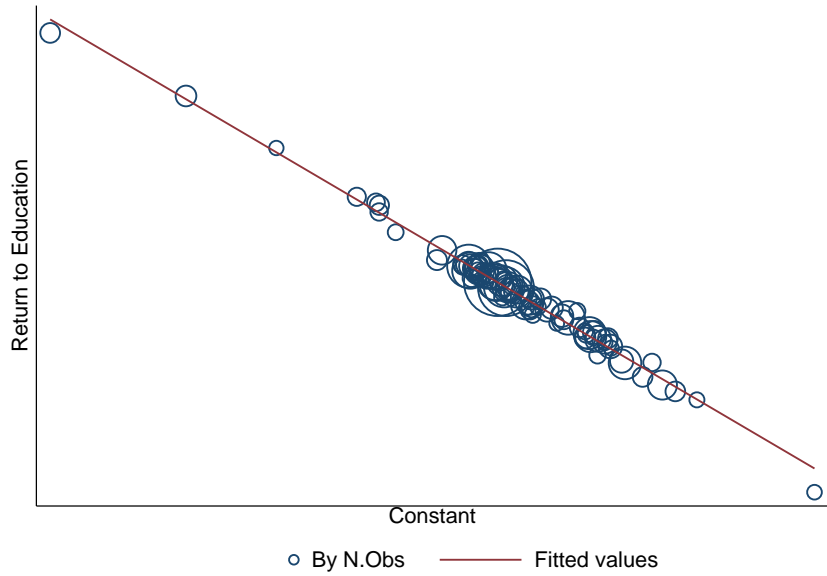
The standard Mincerian equations are estimated separately for each occupation, allowing returns to education to differ across occupations. The occupation-specific returns to education are plotted against various occupation rank measures. Each circle represents an occupation at three-digit level (OCC1990), and the circle size represents the sample size used to obtain these estimates. Occupations are ranked by the average education, by average wages and by average incomes of employed workers in panel (a), (b) and (c) respectively.

Figure 6: Evidence of Linear Wage Contract (ACS & NLSY79)

(a) Constant vs Return to Education (ACS)

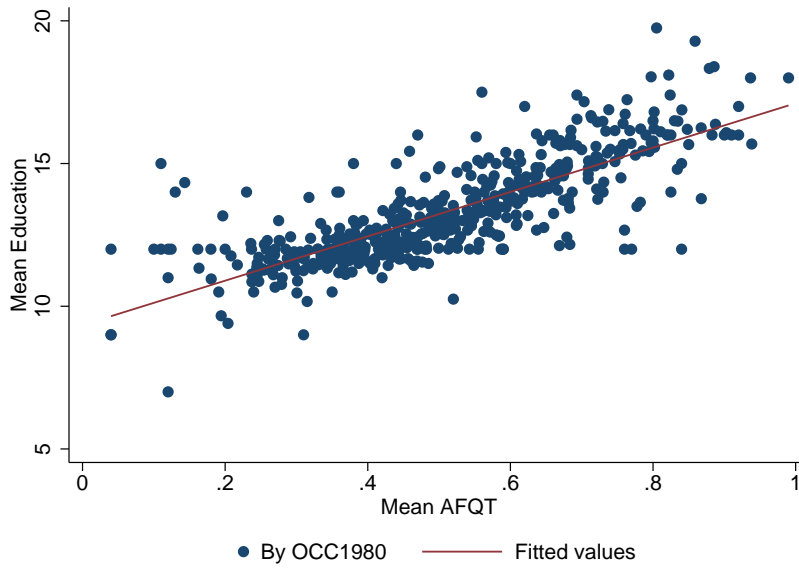


(b) Constant vs Return to Education (NLSY79)



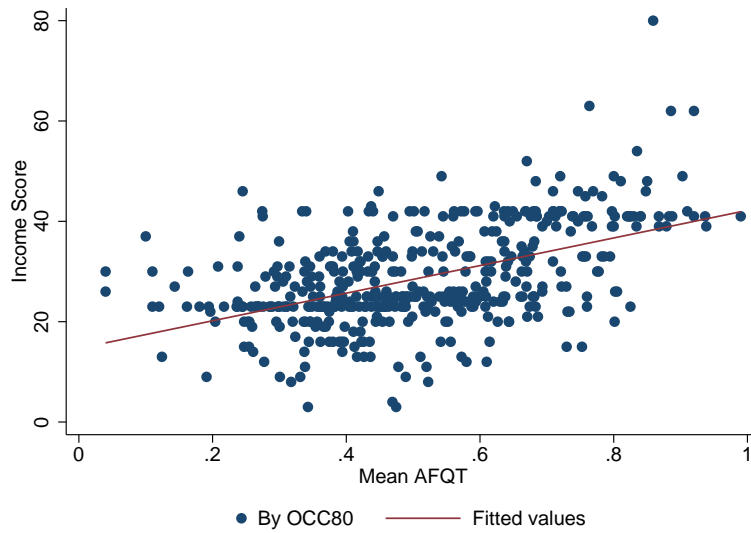
The standard Mincerian equations are estimated separately for each occupation, allowing returns to education and constant to differ across occupations. The occupation-specific returns to education are plotted against constant terms in the Mincerian wage regression. Each circle represents an occupation at three-digit level (OCC1990), and the circle size represents the sample size used to obtain these estimates. The results using the ACS and the NLSY79 are presented in panel (a) and (b) respectively.

Figure 7: Mean Education v.s. Mean Ability By Occupation



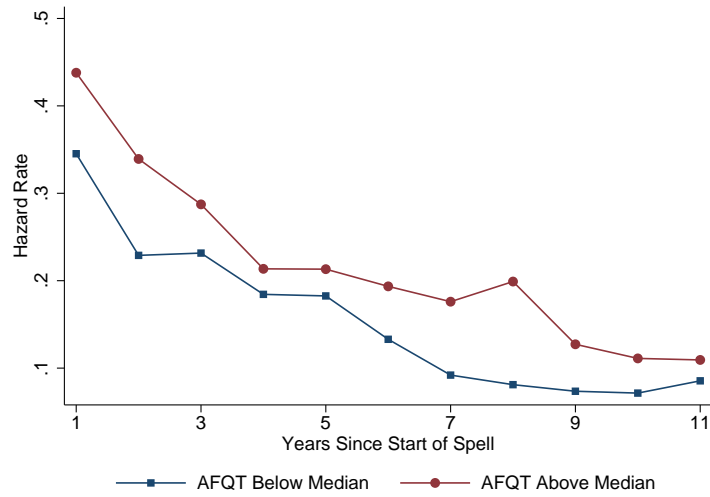
The average cognitive ability is plotted against average education at occupation level using the NLSY79 data. Each dot represents an occupation at the three-digit level (OCC1980).

Figure 8: Mean Education v.s. Income Score By Occupation



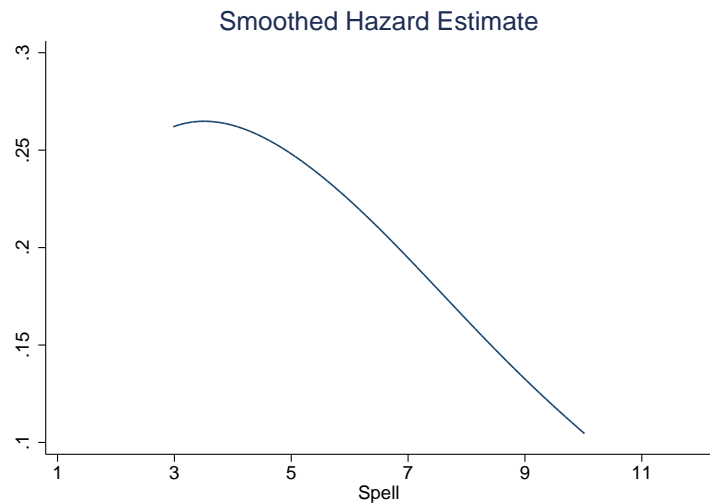
The average cognitive ability is plotted against income score at occupation level using the NLSY79 data. Income Score is a constructed variable that assigns continuous scores to each occupation. Income Score assigns each occupation a value representing the median total income (in hundreds of 1950 dollars) of all persons with that particular occupation in 1950. Income Score provides a continuous measure of occupations, according to the economic rewards. Each dot represents an occupation at the three-digit level (OCC1980).

Figure 9: Hazard Out of Overeducation By AFQT



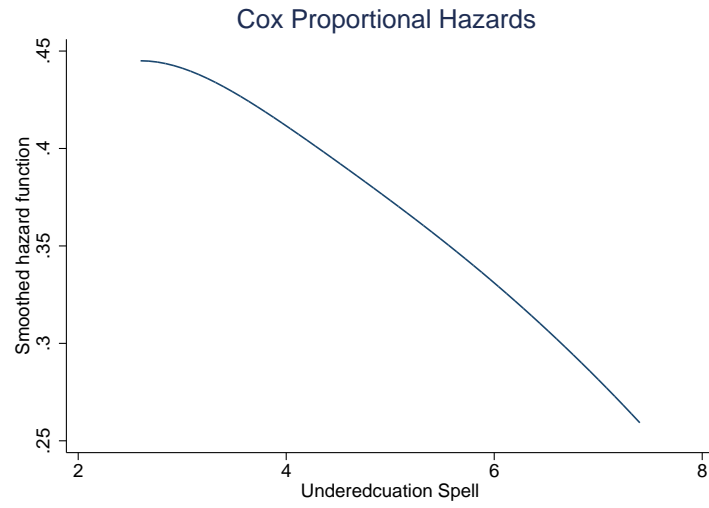
The hazard rates are computed as the number of individuals leaving overeducated employment at t , divided by the number of individuals who are still overeducated at the beginning of t . The computed hazard rate is plot against the length of overeducation measured by the number of years since the start of the current overeducation spell. The exercises are done separately for high (above median AFQT score) and low (below median AFQT score) ability workers.

Figure 10: Cox Proportional Hazard - Base Hazard (Overeducation)



The duration of overeducation is assumed to be determined by a proportional hazard model. The baseline hazard rate is estimated non-parametrically. The smoothed hazard rate out of overeducation is plotted against time.

Figure 11: Cox Proportional Hazard - Base Hazard (Undereducation)



The duration of undereducation is assumed to be determined by a proportional hazard model. The baseline hazard rate is estimated non-parametrically. The smoothed hazard rate out of undereducation is plotted against time.

Table 1: Summary Statistics: Pooled Cross Section 1982-1994

Variable	Mean	(Std. Dev.)	N
Required Education (Year)	13.072	(1.590)	37444
Overeducation (Year)	0.172	(1.616)	37444
Age	27.768	(4.112)	37600
Black	0.103	(0.304)	37600
Hispanic	0.056	(0.230)	37600
Female	0.480	(0.500)	37600
Born in the United States	0.963	(0.188)	37600
AFQT	0.517	(0.270)	36159
Rotter Scale	8.4	(2.357)	37357
Sociability (Adult)	2.902	(0.664)	36757
Highest Grade Attained	13.421	(1.933)	37600
12 Years of Education	0.643	(0.479)	37600
14 Years of Education	0.130	(0.336)	37600
16 Years of Education	0.188	(0.391)	37600
18 Years of Education	0.038	(0.192)	37600
Northeast	0.197	(0.398)	37426
South	0.330	(0.470)	37426
West	0.167	(0.373)	37426
Urban	0.786	(0.410)	36697
HH in SMSA	0.772	(0.420)	35726
Mother's Edu (Yrs.1979)	11.768	(2.517)	36127
Father's Edu (Yrs.1979)	12.037	(3.344)	36127
Employed	0.925	(0.263)	37600
Out of Labor Force	0.038	(0.191)	37600
Unemployed	0.037	(0.189)	37600
Tenure (1k Wks)	0.185	(0.177)	37153
Number of Jobs	1.504	(0.809)	37600
Work Experience (1k Hrs)	17.179	(9.080)	37600
Hourly Wage	10.397	(18.202)	37600
Weeks Unemployed	1.278	(3.906)	37038
Occupational Hazards	2.105	(0.688)	36888

Table 2: Augmented Log-wage Regression

	All Sample	Restricted Smaple Without Wage Outliers	Restricted Smaple Only Full Time Workers
Required Education (Yrs)	0.121*** (0.0020)	0.117*** (0.0015)	0.113*** (0.0017)
Year Above Required Education	0.040*** (0.0021)	0.039*** (0.0016)	0.045*** (0.0019)
Year Below Required Education	-0.086*** (0.0033)	-0.084*** (0.0025)	-0.079*** (0.0028)
Experience (1000h)	0.062*** (0.0010)	0.0525*** (0.0008)	0.051*** (0.0010)
Experience ²	-0.0009*** (0.00003)	-0.0007*** (0.00002)	-0.007*** (0.00002)
Adjust R ²	0.3873	0.4802	0.4398
F	1453.60	2018.14	1202.85
N.Obs	45969	43674	30665
Demographic Controls	X	X	X
Ability Controls	X	X	X
Regional Controls	X	X	X
Occupation Characteristics	X	X	X

The demographic variables control for race, gender and a dummy for born in US; the ability measures include AFQT scores, rotter score, and sociability score; regional controls are a set of dummies for living in an urban area at the time of interview, a set of dummies for census regions. I present results using three different samples: (1) the full sample with valid wage data; (2) the subsample with selected workers whose hourly wage above the federal minimum wage levels and below 200 dollars; (3) on the basis of sample 2, further restrict workers to those who work more than 1500 hours yearly (full-time full-year workers).

* p<0.10. ** p<0.05. *** p<0.01

Table 3: Calibration of the Model to the NLSY 79 Data

Parameter	Value	Target Moment	
μ_A	=0.5474	average AFQT scores (NLSY 79 Sample)	0.5474
σ_A	=0.2682	Std. of AFQT scores (NLSY 79 Sample)	0.2682
μ_S	=13.01	average Education (NLSY 79 Sample)	13.01
σ_S	=2.31	Std. of Education scores (NLSY 79 Sample)	2.31
ρ_{AS}	=0.5813	Correlation Coefficient of Education and Ability (NLSY 79 Sample)	0.5813
α	=0.1955	ratio of return to education to return to cognitive ability (Standard Mincerian Equation)	0.1955
γ_k		employment share of occupational groups (Exogenous Labor Demand Demand)	
P_k		conditional wage distribution & OLS estimates of the Standard Mincerian Equation	

Table 4: Employment Share, Mean Log Wage,& Productivity

Occupational Group	Employment Share	Mean Log Wage	Calibrated Productivity
k	γ_k		P_k
1	10.1%	0.8947	0.2263
2	10.8%	1.3992	0.2263
3	9.2%	1.6025	0.2263
4	10.0%	1.7684	0.2263
5	11.0%	1.9371	0.3548
6	9.0%	2.0934	0.3548
7	10.0%	2.2463	0.3548
8	10.0%	2.4183	0.3548
9	10.0%	2.6340	0.3663
10	9.9%	3.1093	0.3663

Table 5: Estimates of Duncan and Hoffman's Augmented Mincerian Specification (Simulated Sample)

	α^r		α^o		α^u	
Simulated Sample	0.0781***	(0.001)	0.0479***	(0.002)	-0.0244***	(0.002)
Range in Literature	[0.043, 0.135]		[-0.031, 0.054]		[-0.056, -0.025]	
Mata Analysis (151 Studies)	0.089	(0.003)	0.043	(0.002)	-0.036	(0.002)
Mata Analysis (US/Canada)	0.083	(0.006)	0.046	(0.004)	-0.027	(0.004)

The second to the forth rows are taken from a meta-regression by Leuven et al. (2011) with 151 studies. Leuven et al. (2011) break the sample of results down by continent, decade, method to measure required schooling, estimation method and gender. See Leuven et al. (2011) for detail. Std. Err presented in the parentheses. * $p < 0.10$. ** $p < 0.05$. *** $p < 0.01$

Table 6: Correlation Coefficient between Average Ability and Occupation Scores

	Mean Education	Income Score	Socioeconomic Score(SEI)	Prestige Score (Siegel)	Prestige Score (Nakao-Treas)	Nam-Powers-Boyd Status
AFQT	0.8024	0.5085	0.7414	0.6749	0.6723	0.7324

Income Score is a constructed variable that assigns occupational income scores to each occupation. Income Score assigns each occupation a value representing the median total income (in hundreds of 1950 dollars) of all persons with that particular occupation in 1950. Income Score thus provides a continuous measure of occupations, according to the economic rewards enjoyed by people working at them in 1950. SEI is a constructed measure that assigns a Duncan Socioeconomic Index (SEI) score to each occupation using the 1950 occupational classification scheme. The SEI is a measure of occupational status based upon the income level and educational attainment associated with each occupation in 1950.

The Siegel score variable is based on the series of surveys conducted at National Opinion Research Center during the 1960s. In all surveys, respondents were asked to evaluate either "general standing" or "social standing" of occupations. Siegel transformed occupational prestige rating data into a common metric.

The prestige score assigned by Nakato and Treas, using data from the 1989 General Social Survey.

The NPB Status score is a measure of occupational status based upon the median earnings and median educational attainment associated with each category in the occupational. Occupational status score gives equal weights to education and earnings, and can be interpreted as the percentage of persons in the civilian labor force who are in occupations having combined levels of education and earnings below that occupation.

Table 7: Probit Model of Overeducation

	All Sample	>=12 Years	>=16 Years	>=18 Years
AFQT	-0.2627*** (0.0537)	-0.3491*** (0.0543)	-0.4652*** (0.0929)	-1.1504*** (0.3657)
Rotter Score	-0.0034 (0.0047)	-0.0038 (0.0048)	0.0044 (0.0075)	-0.0044 (0.0246)
Sociability	-0.0050 (0.014)	-0.0132 (0.0167)	0.0202 (0.0262)	0.0047 (0.0781)
Black	0.1399*** (0.0407)	0.1241*** (0.0414)	0.1825** (0.0780)	0.0007 (0.2721)
Hispanic	-0.0566 (0.0528)	-0.0414 (0.0541)	-0.2564** (0.1010)	-0.3140 (0.3026)
Work Experience (1000h)	-0.0217*** (0.0051)	-0.0133 (0.0095)	-0.028*** (0.0101)	0.0183 (0.0317)
Work Experience ²	0.0004*** (0.0001)	0.0004*** (0.0001)	0.0004* (0.0002)	-0.0003 (0.0008)
Occupational Hazards	0.5026*** (0.1940)	0.5157*** (0.0195)	0.8391*** (0.0423)	2.538*** (0.2239)
Unemployment Rate	1.6690* (0.8898)	1.6710* (0.8971)	2.6237* (1.4683)	-1.3.994*** (5.2625)
χ^2	11137.46	10218.94	772.28	141.28
LL	-8973.09	-8871.84	-3940.49	-436.58
N.Obs	29011	26625	6426	994

Model also includes controls for mother and father's education level in 1979, a dummy for born in US, a dummy for living in an urban area at the time of interview, a set of dummy variables for census regions, number of jobs had, weeks unemployed last year, gender. Like Battu et al. (1999), Dolton and Vignoles (2000), Green and Zhu (2010), and Büchel and Pollmann-Schult (2001), I also control for the degree effect. * $p < 0.10$. ** $p < 0.05$. *** $p < 0.01$

Table 8: Determinants of Within-Occupation Schooling Ranking

	All Sample	>=12 Years	>=16 Years	>=18 Years
AFQT	-0.0779*** (0.0043)	-0.0840*** (0.0045)	-0.0806*** (0.0119)	-0.1626*** (0.0343)
Rotter Score	0.0010** (0.0004)	0.0011*** (0.0004)	0.0017* (0.0010)	-0.0022 (0.0023)
Sociability	-0.0010 (0.0013)	-0.0021 (0.0014)	-0.0025 (0.0038)	0.0062 (0.0076)
Black	0.0250*** (0.0033)	0.0215*** (0.0036)	0.0099 (0.0097)	0.0141 (0.0249)
Hispanic	-0.0126*** (0.0040)	-0.0130*** (0.0046)	-0.0254** (0.0130)	-0.0398 (0.0307)
Work Experience (1000h)	-0.0028*** (0.0004)	-0.0029*** (0.0005)	-0.0018 (0.0012)	0.2230 (0.0030)
Work Experience ²	0.00002** (0.00001)	0.00002*** (0.00001)	0.00004 (0.00003)	-0.0000 (0.0000)
Occupational Hazards	0.0920*** (0.0015)	0.0979*** (0.0016)	0.0882*** (0.0047)	0.1209*** (0.0113)
Unemployment Rate	0.2751*** (0.0703)	0.3195*** (0.0753)	0.0879 (0.1876)	-0.1890** (0.4901)
R ²	0.6673	0.5768	0.1223	0.2080
N.Obs	29011	26625	6426	994

Model also includes controls for mother and father's education level in 1979, a dummy for born in US, a dummy for living in an urban area at the time of interview, a set of dummy variables for census regions, number of jobs had, weeks unemployed last year, gender. Like Battu et al. (1999), Dolton and Vignoles (2000), Green and Zhu (2010), and Büchel and Pollmann-Schult (2001), I also control for the degree effect. * p<0.10. ** p<0.05. *** p<0.01

Table 9: Overeducation Spell Length v.s. Ability Measure

Overeducation		All Spell	
AFQT	-1.887*** (0.2228)	-1.845*** (0.2268)	-1.793*** (0.2253)
Unemployment Rate	84.532*** (3.1829)	83.753*** (3.1717)	71.337*** (3.9321)
Wks Uemployment	0.018** (0.0098)	0.017** (0.0097)	0.008 (0.0098)
Rotter	-0.007	-0.012 (0.0255)	-0.008 (0.0252)
Sociability	-0.250***	-0.255*** (0.0474)	-0.249*** (0.0469)
Work Experience (1000h)		-0.054*	-0.025 (0.0277)
Work Experience ²		-0.0002	-0.0008 (0.0009)
Demographic Controls			ⓧ
Regional Controls			ⓧ
Overeducation		Only First Spell	
AFQT	-2.525*** (0.3203)	-2.443*** (0.3258)	-2.466*** (0.3256)
Unemployment Rate	90.299*** (4.9323)	89.671*** (4.9016)	85.334*** (5.5558)
Wks Uemployment	0.020 (0.0130)	0.018 (0.0129)	0.017 (0.0131)
Rotter	0.007	0.007 (0.0359)	0.003 (0.0359)
Sociability	-0.310***	-0.318*** (0.0620)	-0.319*** (0.0620)
Work Experience (1000h)		0.037	0.058 (0.0472)
Work Experience ²		-0.003	-0.0029 (0.0019)
Demographic Controls			ⓧ
Regional Controls			ⓧ

The demographic controls include race, gender, born in US; the regional variables control for regional dummy, dummy for urban and smsa area. * p<0.10. ** p<0.05. *** p<0.01

Table 10: Undereducation Spell Length v.s. Ability Measure

Undereducation		All Spell		Only First Spell	
AFQT	-0.154 (0.1712)	-0.096 (0.1798)	0.392** (0.1699)	0.274* (0.1498)	
Unemployment Rate	88.241*** (2.5229)	90.949*** (3.2.5547)	56.616*** (2.7339)	60.348*** (2.8279)	
Wks Uemployment	0.058*** (0.0096)	0.060*** (0.0097)	0.011 (0.0093)	0.002 (0.0092)	
Rotter		0.003 (0.0202)	-0.016 (0.0190)	-0.008 (0.0187)	
Sociability		-0.163** (0.0696)	-0.061 (0.0654)	-0.048 (0.0644)	
Work Experience (1000h)			-0.193*** (0.0185)	-0.183*** (0.0183)	
Work Experience ²			0.0007 (0.0005)	0.0003 (0.0005)	
Demographic Controls					⊗
Regional Controls					⊗
AFQT	0.199 (0.3258)	0.184 (0.3383)	0.435 (0.3178)	0.262 (0.3360)	
Unemployment Rate	51.784*** (5.4245)	54.925*** (5.5991)	29.453*** (5.8327)	30.293*** (5.9987)	
Wks Uemployment	0.001 (0.0228)	0.025 (0.0234)	0.0007 (0.0222)	-0.006 (0.0222)	
Rotter		-0.044 (0.0356)	-0.060 (0.0334)	-0.056* (0.0333)	
Sociability		-0.191 (0.1225)	-0.138 (0.1150)	-0.130 (0.1146)	
Work Experience (1000h)			-0.070* (0.0373)	-0.071* (0.0375)	
Work Experience ²			-0.0017* (0.0010)	-0.0020* (0.0010)	
Demographic Controls					⊗
Regional Controls					⊗

The demographic controls include race, gender, born in US; the regional variables control for regional dummy, dummy for urban and smsa area.
p<0.10. ** p<0.05. *** p<0.01

Table 11: Cox Proportional Hazard Model

	Overeducation			Undereducation		
	Coef.	Std. Err	Haz. Ratio	Coef.	Std. Err	Haz. Ratio
AFQT	0.204*	(0.1205)	1.226	-0.158*	(0.1026)	0.854
Rotter Score	0.012	(0.0116)	1.012	0.004	(0.0101)	1.004
Sociability	0.043*	(0.0241)	1.043	0.007	(0.0186)	1.007
Black	-0.116	(0.1057)	0.890	0.045	(0.0908)	1.046
Hispanic	0.090	(0.1308)	1.094	0.044	(0.1099)	1.045
Born in the US	-0.123	(0.164)	0.884	0.026	(0.1247)	1.026
Work Experience (1000h)	0.043***	(0.0131)	1.043	0.008	(0.0094)	1.008
Work Experience ²	-0.0008**	(0.0004)	0.999	-0.00003	(0.0002)	0.999
Unemployment Rate	-15.74***	(2.5194)	0.00	-1.139	(2.093)	0.320
χ^2	113.74			41.55		
LL	-9425.26			-11910.98		
N.Obs	4698			2824		

Model also includes controls for mother and father's education level in 1979, a dummy for living in an urban area at the time of interview, a set of dummy for census regions, degree type, gender. * p<0.10. ** p<0.05. *** p<0.01

Table 12: Dynamic Impact of AFQT on Overeducation

		All Samples	>= 12 Years	>=16 Years	>=18 Years				
		Incidence of Overeducation							
Probit Model	AFQT	-0.2045**	(0.09310)	-0.3034***	(0.0960)	-0.3120	(0.2059)	-1.059	(0.8595)
	Exp	-0.0198***	(0.0056)	-0.0212***	(0.0058)	-0.0060	(0.0129)	0.0231	(0.0518)
	AFQT*Exp	-0.0036	(0.0047)	-0.0028	(0.0049)	-0.0098	(0.0118)	-0.0056	(0.0478)

Model also includes controls for mother and father's education level in 1979, a dummy for born in US, a dummy for living in an urban area at the time of interview, a set of dummy variables for census regions, number of jobs had, weeks unemployed last year, gender. Like Battu et al. (1999), Dolton and Vignoles (2000), Green and Zhu (2010), and Büchel and Pollmann-Schult (2001), I also control for the degree effect. * p<0.10. ** p<0.05. *** p<0.01

Appendix B

CPS data

The 1989-1991 monthly CPS survey has a sample target of 50,000 households split into eight representative subsamples, each of which is interviewed for the first and last four months of a 16-month period. In any given month, a new sample of 6250 households is surveyed for the first time. As a result, the pooled monthly CPS data from January of 1989 through December of 1991 contain 453866 unique households.

From these pooled cross-sections, I keep only observations in the age range spanned by the NLSY79 cohort, which leaves 210567 unique individuals. Then I drop observations where an individual is unemployed, does not report a Census occupation code, has a missing level of education, did not report the level of education, or is enrolled in college.

After making these cuts, I have 615,070 observations. From this restricted sample, the required level of education for each occupation identified by its 3-digit Census occupation code is computed as the sample mode of the distribution of education levels among workers in that occupation. Then, each individual observation in NLSY79 is matched to the required level of education using the 3-digit occupation code (OCC1980). For those occupations that are observed less than 100 times in the CPS pooled sample, I collapse these low-frequency occupations using 2-digit codes rather than 3-digit codes before applying the match procedure. The 1980 Census Occupational Code structure provides a detailed mapping between the 2-digit and the 3-digit code.

Regional Unemployment Rates

The BLS Local Area Unemployment Statistics data provide monthly estimates of total employment and unemployment for each Census region (Northeast, South, West, Midwest). The regional unemployment rate is computed by aggregating monthly data to compute the annual unemployment rate.

Measure Occupational Hazards

I derive the occupational hazards measure from the Occupational Information Network (O*NET), developed by the US Department of Labor. O*NET provides detailed descriptive information for each of over 900 occupations in the US Standard Occupational Classification (SOC) system. Despite several methodological problems in earlier studies, O*NET is a promising source of job characteristics that impact workers' health. O*NET data describe occupation characteristics through rating obtained from randomly selected current job holders and occupational analysts. They assessed each job using a standardized rating system, which consisted of 277 items describing various aspects of the occupation.

To construct the occupational hazards measure, I calculate the mean of 7 items that address common physical hazards traditionally studied in occupational safety and health. These 7 items asked the frequency of exposure to the following conditions: sounds and noise levels that are distracting and uncomfortable, very hot (above 90F) or very cold (under 32F) temperatures, extremely bright or inadequate lighting conditions, high places (e.g., working on poles, scaffolding, catwalks, or ladders), an environment that is not controlled (i.e., without air conditioning), outdoors under cover, and outdoors exposed to all weather conditions.

After obtaining the occupational hazards measure, I use the crosswalk provided by the BLS to compute measures for OCC1980 codes.

Appendix C

Proof of Proposition 1

Educational Choices

Workers choose optimal education to maximize lifetime income.

$$\max_{S_i} [P_k(S_i)(\mu_{i0} + \alpha f(S_i)) - \Pi_k(S_i)]\beta_T - C(S_i, \mu_{i0})$$

This problem involves both continuous and discrete optimization and can be solved by the following two-step procedure.

Step 1: To keep the functional form simple and the solution tractable, I impose additional functional

form on the human capital function with $f(S) = S$. For any given intended occupation k identified by P_k , we first find the corresponding optimal schooling S_k^* for this potential occupation.

$$S_k^* = \frac{\alpha\beta_T P_k \mu_{i0}}{c}$$

Step 2: After computing the optimal schooling level S_k^* for every occupation k , individuals subsequently choose the best intended occupation based on expected lifetime income $\mathbb{E}(I_{ik0}|\mathcal{F}_0)$.

$$\begin{aligned}\mathbb{E}(I_{ik0}|\mathcal{F}_0) &= [P_k(\mu_{i0} + \alpha S_k^*) - \Pi_k]\beta_T - C(S_k^*, \mu_{i0}) \\ &= (P_k \mu_{i0} - \Pi_k)\beta_T + \frac{\alpha^2 \mu_{i0} P_k^2}{2c} \beta_T^2 \\ &= [(P_k + \frac{\alpha^2 \beta_T}{2c} P_k^2) \mu_{i0} - \Pi_k] \beta_T\end{aligned}$$

Occupation k is optimal rather than $k-1$ or $k+1$ if the expected lifetime income is higher:

$$\begin{aligned}\mathbb{E}(I_{ik0}|\mathcal{F}_0) &\geq \mathbb{E}(I_{ik-10}|\mathcal{F}_0) \\ \Rightarrow \tilde{P}_k \mu_{i0} - \Pi_k &\geq \tilde{P}_{k-1} \mu_{i0} - \Pi_{k-1} \\ \mathbb{E}(I_{ik0}|\mathcal{F}_0) &\geq \mathbb{E}(I_{ik+10}|\mathcal{F}_0) \\ \Rightarrow \tilde{P}_k \mu_{i0} - \Pi_k &\geq \tilde{P}_{k+1} \mu_{i0} - \Pi_{k+1}\end{aligned}$$

where $\tilde{P}_k = P_k + \frac{\alpha^2 \beta_T}{2c} P_k^2$. \tilde{P}_k maintain the ranking of P_k because $\frac{\partial \tilde{P}_k}{\partial P_k} \geq 0$. We would obtain a vector of optimal cutoff $(\tilde{B}_1, \tilde{B}_2, \dots, \tilde{B}_K)$ where \tilde{B}_k is defined as

$$\beta_k \equiv \frac{\Pi_k - \Pi_{k-1}}{\tilde{P}_k - \tilde{P}_{k-1}} \text{ for } k \in \{1, \dots, K\}$$

The sorting in Step 2 implies that $\frac{\partial P_k^*}{\partial \mu_{i0}} \geq 0$ and $\frac{\partial S_k^*}{\partial \mu_{i0}} \geq 0$. Not surprisingly, these two inequality immediately implies that high ability workers obtain more education and sort into high-paying high-rank occupations.

Proof of Proposition 3

Consider an overeducated worker in occupation k with t years of experience, this worker exists overeducation when he moves up the occupation ladder. Proposition 1 illustrates that high ability

workers sort into higher occupations which on average more educated.

$$\begin{aligned}
P_{kt}^+ &= Prob(\mathbb{E}(H_{it+1}) > B_{k+1} | \mathcal{F}_t) \\
&= Prob\left(\frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_t^2} \mu_{it} + \frac{\sigma_t^2}{\sigma_\epsilon^2 + \sigma_t^2} (A_i + \epsilon_{it+1}) + \alpha S_i > B_{k+1} | \mathcal{F}_t\right) \\
&= Prob\left(\epsilon_{t+1} > \left(1 + \frac{\sigma_\epsilon^2}{\sigma_t^2}\right) (B_{k+1} - \alpha S_i) - \frac{\sigma_\epsilon^2}{\sigma_t^2} \mu_{it} - A_i\right) \\
&= 1 - \Phi\left(\frac{1}{\sigma_\epsilon} \left[\left(1 + \frac{\sigma_\epsilon^2}{\sigma_t^2}\right) (B_{k+1} - \alpha S_i) - \frac{\sigma_\epsilon^2}{\sigma_t^2} \mu_{it} - A_i\right]\right)
\end{aligned}$$

Therefore, the switch-up probability P_{kt}^+ is an increasing function of ability A_i .

In addition, one can define the switch-down probability $P_{kt}^- = Prob(\mathbb{E}(H_{it+1}) < B_{k+1} | \mathcal{F}_t)$ in a similar way. We can show that the switch-down probability is a decreasing function of ability $\frac{\partial P_{kt}^-}{\partial A_i} < 0$.

$$\begin{aligned}
P_{kt}^- &= Prob\left(\frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_t^2} \mu_{it} + \frac{\sigma_t^2}{\sigma_\epsilon^2 + \sigma_t^2} (A_i + \epsilon_{it+1}) + \alpha S_i \leq B_k\right) \\
&= \Phi\left(\frac{1}{\sigma_\epsilon} \left[\left(1 + \frac{\sigma_\epsilon^2}{\sigma_t^2}\right) (B_k - \alpha S_i) - \frac{\sigma_\epsilon^2}{\sigma_t^2} \mu_{it} - A_i\right]\right)
\end{aligned}$$

As ability revealed low-ability workers are more likely to move down the occupational ladder to where they are persistently considered as overeducated. Both $\frac{\partial P_{kt}^+}{\partial A_i} > 0$ and $\frac{\partial P_{kt}^-}{\partial A_i} < 0$ imply that overeducation (undereducation) is more persistent for low (high) ability workers.

Proof of Proposition 4

In Proposition 3, we show that the probability of existing overeducation P_{kt}^+ is an increasing function of ability A_i . In this section, we show that P_{kt}^+ is an decreasing function of experience t , $\frac{\partial P_{kt}^+}{\partial t} < 0$.

$$\begin{aligned}
P_{kt}^+ &= Prob(\mathbb{E}(H_{it+1}) > B_{k+1} | \mathcal{F}_t) \\
&= 1 - \Phi\left(\frac{1}{\sigma_\epsilon} \left[\left(1 + \frac{\sigma_\epsilon^2}{\sigma_t^2}\right) (B_{k+1} - \alpha S_i) - \frac{\sigma_\epsilon^2}{\sigma_t^2} \mu_{it} - A_i \right]\right) \\
&= 1 - \Phi\left(\frac{1}{\sigma_\epsilon} \left[(B_{k+1} - \alpha S_i - A_i) + \frac{\sigma_\epsilon^2}{\sigma_t^2} (B_{k+1} - \alpha S_i - \mu_{it}) \right]\right)
\end{aligned}$$

where $\sigma_t^2 = \frac{\sigma_\epsilon^2 \sigma_0^2}{\sigma_\epsilon^2 + t \sigma_0^2}$, is a decreasing function of experience t , $\frac{\partial \sigma_t^2}{\partial t} < 0$.

To find the sign of $\frac{\partial P_{kt}^+}{\partial t} = \frac{\partial P_{kt}^+}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial t}$, we need to determine the sign of $\frac{\partial P_{kt}^+}{\partial \sigma_t^2}$,

$$\frac{\partial P_{kt}^+}{\partial \sigma_t^2} = \phi\left(\frac{1}{\sigma_\epsilon} \left[(B_{k+1} - \alpha S_i - A_i) + \frac{\sigma_\epsilon^2}{\sigma_t^2} (B_{k+1} - \alpha S_i - \mu_{it}) \right]\right) \frac{\sigma_\epsilon^2}{\sigma_t^4} (B_{k+1} - \alpha S_i - \mu_{it}) > 0.$$

The last term in the previous equation $(B_{k+1} - \alpha S_i - \mu_{it})$ is positive because this individual currently works in occupation k with $B_k \leq \alpha S_i + \mu_{it} < B_{k+1}$. This immediately implies that $\frac{\partial P_{kt}^+}{\partial t} = \frac{\partial P_{kt}^+}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial t} < 0$. Similarly, we can show $\frac{\partial P_{kt}^-}{\partial t} = \frac{\partial P_{kt}^-}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial t} < 0$. The existing probability of undereducation declines as experience increases as well.

$$\begin{aligned}
P_{kt}^- &= Prob\left(\frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_t^2} \mu_{it} + \frac{\sigma_t^2}{\sigma_\epsilon^2 + \sigma_t^2} (A_i + \epsilon_{it+1}) + \alpha S_i \leq B_k\right) \\
&= \Phi\left(\frac{1}{\sigma_\epsilon} \left[\left(1 + \frac{\sigma_\epsilon^2}{\sigma_t^2}\right) (B_k - \alpha S_i) - \frac{\sigma_\epsilon^2}{\sigma_t^2} \mu_{it} - A_i \right]\right) \\
&= \Phi\left(\frac{1}{\sigma_\epsilon} \left[(B_k - \alpha S_i - A_i) + \frac{\sigma_\epsilon^2}{\sigma_t^2} (B_k - \alpha S_i - \mu_{it}) \right]\right)
\end{aligned}$$

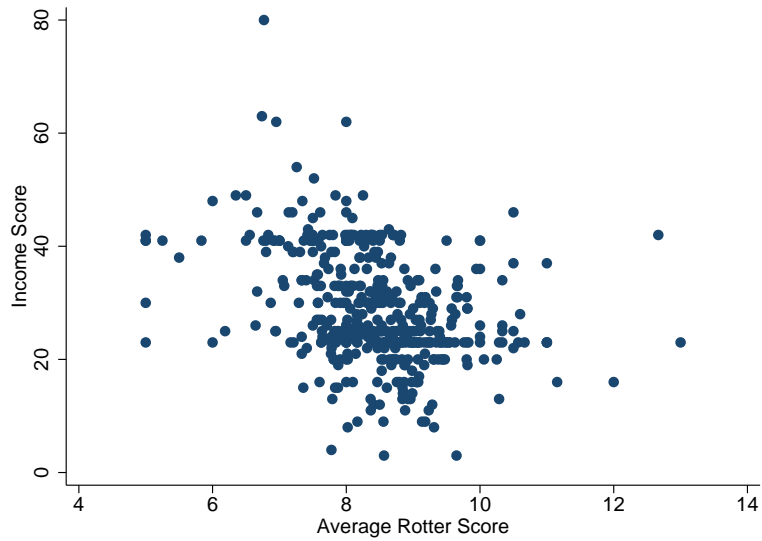
$$\frac{\partial P_{kt}^-}{\partial t} = \frac{\partial P_{kt}^-}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial t} < 0 \text{ because } \frac{\partial P_{kt}^-}{\partial \sigma_t^2} > 0$$

$$\frac{\partial P_{kt}^-}{\partial \sigma_t^2} = -\phi\left((B_k - \alpha S_i - A_i) + \frac{\sigma_\epsilon^2}{\sigma_t^2} (B_k - \alpha S_i - \mu_{it})\right) * \frac{\sigma_\epsilon^2}{\sigma_t^4} (B_k - \alpha S_i - \mu_{it}) > 0$$

The last term $(B_k - \alpha S_i - \mu_{it})$ is negative for the same reason why $(B_{k+1} - \alpha S_i - \mu_{it})$ is positive. The worker's current occupation k indicates $B_k \leq \alpha S_i + \mu_{it} < B_{k+1}$. The hazard rates out of both over- and under- decline with labor market experience t .

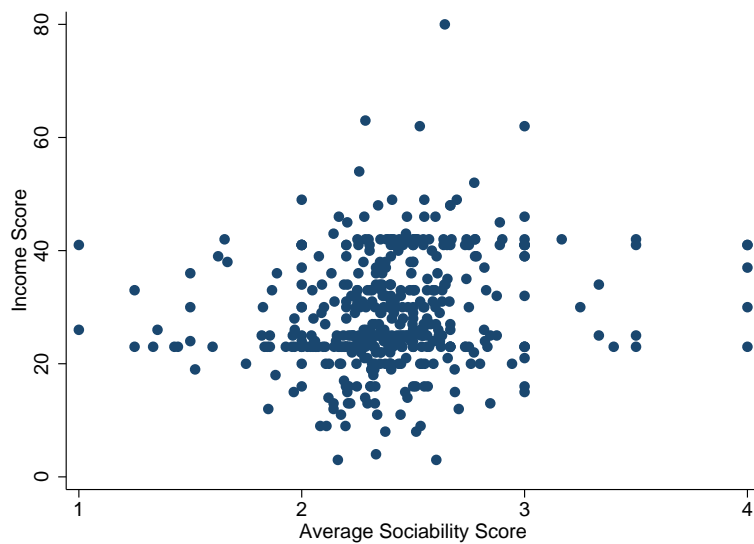
Appendix D

Figure D.1: Rotter Score v.s. Occupational Ranking



The average non-cognitive ability (Rotter Score) is plotted against income score at occupation level using the NLSY79 data. Each dot represents an occupation at the three-digit level (OCC1980).

Figure D.2: Sociability v.s. Occupational Ranking



The average sociability is plotted against income score at occupation level using the NLSY79 data. Each dot represents an occupation at the three-digit level (OCC1980).