Multi-Dimensional Abilities, Task Content of Occupations and Career Choices: A Dynamic Analysis

Dajun Lin*

January 12, 2019

Abstract

This paper presents a dynamic model of schooling, labor supply and occupation choices with two key features. First, abilities are treated as multi-dimensional latent factors (*cognitive*, *socio-emotional*, *manual* and *routine*), and they dynamically generate later life outcomes, including schooling choices, occupation choices and wages. Second, occupations are modeled as bundles of discrete tasks that correspond to the four dimensions of abilities. This approach explicitly accounts for dynamic occupational sorting based on the match between workers' latent abilities and the task content of occupations. The model is estimated with the Maximum Simulated Likelihood method using data from the 1979 National Longitudinal Survey of Youth. Results confirm multi-dimensional sorting and quantify the effects of latent abilities, schooling and human capital accumulated on the job on occupational task choices and wages. Using the structural estimates, I conduct counterfactual simulations to evaluate various human capital development policies proposed to promote welfare (e.g., universal 4-year college subsidies). Simulation results quantify how much each policy affects workers' educational attainment, task-specific labor supply throughout prime working ages and lifetime income.

*PhD candidate, Department of Economics, University of Virginia. Contact: dl5me@virginia.edu. I thank Steve Stern and Leora Friedberg for their guidance and support. I am also grateful for comments from Selcen Cakir, Karim Chalak, Dina Guo, Bill Johnson, Mike LaForest, Siying Liu, John Pepper, Chris Ruhm, Todd Stinebrickner, Paul Sullivan and Eric Young. Special thanks to Steve McClaskie for his help with nuances of the NLSY79 data, Katherine Holcomb for her assistance with parallel computing, and the Advanced Research Computing Services group at the University of Virginia for providing technical support and allocating an unusually large amount of computational resource to support this project. The Bankard Fund of Political Economy and Dissertation Completion Fellowship are acknowledged for financial support. All errors are mine.

1 Introduction

Much of the literature on labor economics has been devoted to answering one question: why are workers paid differently? Multiple factors over workers' careers contribute to the observed variation in income. Early theory suggests three basic ideas to characterize workers' careers: investment in education (Mincer, 1958; Becker, 1964), occupational sorting (Roy, 1951) and learning by doing (Ben-Porath, 1967; Blinder and Weiss, 1976; Heckman, 1976). Keane and Wolpin (1997) incorporate these ideas into a dynamic model of endogenous human capital accumulation and find that such a theoretically restricted model can fit observed data, including schooling attainment, occupation choices and wages, reasonably well.

Recent literature documents two important empirical facts. First, abilities, measured during childhood and adolescence, are multi-dimensional and have different effects on a wide range of later life outcomes, including but not limited to labor market outcomes.[1] Second, occupations can be viewed as bundles of tasks, and the task content of occupations explains a significant portion of observed wage variation, even conditioning on education levels.[2]

Motivated by these empirical facts, this paper presents a dynamic model of schooling, labor supply and occupational choices with multi-dimensional unobserved heterogeneity. The framework combines features from the dynamic human capital accumulation model (Keane and Wolpin, 1997) and factor models (Carneiro et al., 2003; Heckman et al., 2006; Cunha et al., 2010; Heckman et al., 2016). Specifically, each agent is modeled to have four dimensions of latent abilities, *cognitive*, *socio-emotional*, *manual* and *routine*, and these abilities dynamically generate later life outcomes, including schooling choices, occupation choices and wages. Occupations are modeled as bundles of discrete tasks that correspond to the four dimensions of abilities, and occupation choices are results of sorting based on not only level of education and human capital accumulated on the job, but also how well multi-dimensional abilities match up with the task content of occupations. By recognizing the multi-dimensional nature of abilities and accounting for dynamic occupational sorting based on the task content, this approach enriches our understanding of abilities, task content of occupations, dynamic occupational sorting, the development of human capital and how each of them contributes to lifetime income and welfare.

The model is estimated with 35 years of schooling and employment data on a sample of white men from the 1979 National Longitudinal Survey of Youths (NLSY79). Aside from

[1] See, among others, Murnane et al. (2001); Heckman et al. (2006); Cunha et al. (2010); Lindqvist and Vestman (2011); Nybom (2017); Prada and Urzúa (2017); Speer (2017); Heckman et al. (2018); Guvenen et al. (2018).

[2] See, among others, Autor et al. (2003, 2006); Poletaev and Robinson (2008); Acemoglu and Autor (2011); Yamaguchi (2012); Autor and Handel (2013); Robinson (2018); Yamaguchi (2018).

career-related information, the data also contain respondents' scores from ten batteries of aptitude tests and multiple surveys on socio-emotional abilities, around the time of their labor market entry. Following Heckman et al. (2006) and Heckman et al. (2016), these scores are assumed to be (a) functions of each respondent's schooling and family background at the time of tests to account for reverse causality, and (b) fallible to account for potential measurement errors. The sufficient conditions for identification of such dynamic factor models are established in Heckman et al. (2016) and Freyberger (2018). It draws on the latent factor structure and a set of conditional independence assumptions. The model is estimated using Maximum Simulated Likelihood (MSL), and the joint likelihood of observing the test scores and labor market outcomes is maximized. Parallel computing is employed to overcome the associated computational burden.

The model generates four novel results. First, the model delivers the most thorough estimates of the returns to education by occupations, characterized as task bundles. College education yields a large wage premium for cognitive tasks and a small premium for interpersonal tasks but no premium for manual or routine tasks. Returns to education vary significantly across tasks and hence occupations. Second, the average return to education, calculated by applying the density of observed occupation choices as weights, increases as workers age. In particular, the average return to an additional year of college grows steadily from 2.80% at age 25 to 5.08% at age 40, and stabilizes at around 4.80% afterwards. The change in return to college over time reflects the fact that, as workers with at least some college education age, they gradually move to cognitive- and interpersonal-task-intensive occupations. Third, task content and pairwise task complementarity are important in explaining wage variation. In particular, the cognitive-interpersonal tasks pair and the manual-routine tasks pair are complementary, and performing either pair in the same occupation brings a large additional wage premium. Fourth, while higher abilities to perform cognitive and routine tasks increase the utility from schooling, socio-emotional ability has no direct effect on schooling utility, and higher manual ability decreases schooling utility.

These findings have two implications. First, failure to account for occupational sorting based on multi-dimensional abilities, especially the ability to perform routine tasks, results in omitted variable bias and an overestimation of return to education. On one hand, higher ability to perform routine tasks results in higher levels of education because it directly increases the utility flow of attending school. On the other hand, workers with higher ability to perform routine tasks tend to sort into routine-task-intensive occupations after graduation, and they are paid more as a result of higher ability rather than increased levels of education. Second, in an environment with multi-dimensional sorting, different policies may incentivize different groups of students to gain additional education, and as a result, generate vastly different marginal returns to education. This idea relates to earlier work by Willis and Rosen (1979) and Björklund and Moffitt (1987), and is further developed in recent work by Carneiro et al. (2011), who advocate for quantifying the "marginal policy relevant treatment effect" in order to address specific policy questions.

Using the structural estimates, I conduct counterfactual simulations to evaluate various human capital development policies proposed to promote welfare (e.g., universal 4-year college subsidies). Simulation results quantify the income gain throughout workers' careers, how much of the income gain is due to workers with increased levels of education sorting into occupations with better-paying tasks (*cognitive* and *interpersonal*) and the net impact on welfare.

The paper proceeds as follows. Section 2 describes the data and presents data patterns that motivate a dynamic human capital development model with multi-dimensional latent abilities, which is developed in section 3. Section 4 presents the estimation method. Section 5 presents and discusses the structural estimates. This paper is still a work in progress, and section 6 lays out future extensions. Throughout the rest of the paper, unless specified otherwise, bold letters denote vectors and scripted letters denote sets.

2 Data

In this section, I describe the data. I start with the primary data source, the NLSY79, which provides detailed schooling and employment information on a sample of individuals for more than 30 years. Crucially, it also contains test scores on the respondents' abilities along multiple dimensions, measured around the time of labor market entry. I use O*NET data to help categorize occupations as bundles of tasks. I then document two empirical data patterns using the NLSY79 data, and the analysis motivates my dynamic human capital accumulation framework.

2.1 NLSY79

The primary data source, the 1979 National Longitudinal Survey of Youth (NLSY79), sampled individuals who were between the ages of 14 and 21 on December 31, 1978. Starting from 1979, 16 interviews were conducted annually through 1994, and 10 interviews biennially since then, with data used here going through 2014. The data provide rich information on education, including enrollment status and educational attainment by year. For those who report to have earned the GED, dates of GED receipt are reported. Employment data are extremely detailed, and two separate pieces are used in this paper. The first piece is a wave-by-wave survey of current employment status, occupation and hourly wage.[3] The second is a retrospective recall of weekly work history, which is used to determine the length of employment spells.

The initial NLSY survey consists of a nationally representative sample (6,111) and supplemental samples of blacks (2,172), Hispanics (1,480), poor whites (1,643) and military respondents (1,280). This study focuses on civilian white men from the representative sample and the supplemental poor whites sample.[4,5] Following Keane and Wolpin (1997), I assume an individual's decision period corresponds to a school year, running from September to August, and decisions are aggregated up to the annual level accordingly. See Appendix C for details on data aggregation and sample selection criteria. The final sample includes 1,379 men for an average of 32.5 years, resulting in 44,759 "person-year" observations.[6]

Table 1 tabulates schooling and labor supply decisions for the final sample. The enrollment rate starts at 88.8% at age 16 and drops to 41.0% at age 18, the typical age for high school graduation. It continues to drop gradually to below 5% at age 25. Graduate education is not uncommon, with 193 (14.0%) of the 1,379 respondents having finished at least one year of graduate school. As an alternative to high school graduation, 115 (8.3%) of the 1,379 respondents report to have ever obtained the GED.

The employment rate climbs from 1.5% at age 16 to 85.1% at age 25. A discrete jump occurs at age 18, due to the large number of students graduating high school and entering

[3] Throughout the paper, wage levels are reported using 2014 dollars, adjusted by the Consumer Price Index published by the Bureau of Labor Statistics. I do not adjust for cost of living at the local labor market level (e.g., MSA level) because location identifiers at finer geographic levels are not publicly available.

[4] The decision to exclude women is because women's careers are heavily affected by fertility and marriage decisions. These features, albeit important, are outside the scope of this paper. See Marianne (2011) for a discussion of these issues.

[5] There are two ways to include minority men in the analysis. The first is "non-structural" with Neal and Johnson (1996) as a prominent example. They include racial dummy variables in the wage equation without interaction terms. This approach assumes that the effects of other control variables (such as AFQT) on wage are the same, which is inconsistent with recent evidence that returns to AFQT are greater for blacks and Hispanics than whites (Lin et al., 2018). The second approach is to recognize that the same characteristics could have different effects for minorities than whites. This is analogous to including a full set of interaction terms between racial dummies and all other explanatory variables included the model. This approach would greatly increase the complexity of the model and computational burden of the estimation procedure. As a result, I follow previous literature (e.g., Willis and Rosen 1979; Heckman and Sedlacek 1985; Keane and Wolpin 1997; Belzil and Hansen 2002; Gould 2002; Sullivan 2010; Prada and Urzúa 2017) and include only whites in the analysis. An additional benefit is that it renders my results directly comparable to previous studies.

[6] In the NLSY79, it is common for respondents to miss certain waves of interview and reappear in later waves. In the final sample, 1,450 (3.2%) of the 44,759 "person-year" observations are missing, creating "holes" in respondents' career history spells. I retain all observations for each respondent, including missing waves, until either the latest survey wave available or the wave the respondent appears in the survey for the last time. This approach preserves the maximum amount of information available from the data. I use a simulation technique introduced later (section 4.5.1) to overcome this missing data problem.

the labor force, and gradual increases ensue. It stabilizes after age 25, hovering at around 90% and peaking at age 37 (92.3%).

2.2 Test Scores

Ten Armed Services Vocation Aptitude Battery (ASVAB) tests were administered to NLSY79 respondents in 1980, and approximately 94% of all respondents completed the tests. The ten tests are arithmetic reasoning, word knowledge, paragraph comprehension, mathematics knowledge, numerical operations, coding speed, general science, auto and shop information, electronics information and mechanical comprehension.

The word knowledge, paragraph comprehension, mathematics knowledge and arithmetic reasoning tests are used to construct the Armed Forces Qualification Test (AFQT) score, which has been widely used in the literature as a measure for ability. Earlier literature, however, does not recognize the multi-dimensional nature of abilities, hence resulting in ambiguous interpretations of what AFQT measures, such as "skill" (Neal and Johnson, 1996) and "ability" (Altonji and Pierret, 2001). Later research gradually recognizes the multi-dimensional nature of abilities (e.g., Murnane et al. 2001; Heckman et al. 2006), with more recent literature converging on using AFQT as a measure for cognitive ability (Carneiro and Lee, 2009; Prada and Urzúa, 2017; Speer, 2017; Heckman et al., 2016, 2018; Guvenen et al., 2018). Following recent literature, I interpret the four tests that constitute AFQT as measures for cognitive ability.

I follow Prada and Urzúa (2017) and interpret test scores from auto and shop information, electronics information and mechanical comprehension as measuring one's ability to perform manual tasks.

I interpret test scores from the numerical operations and coding speed batteries as measuring one's ability to quickly complete small routine tasks. These two tests differ significantly from the rest. Whereas the rest of ASVAB tests are "power tests," for which the scores are based only on one's ability and knowledge, and test takers are offered ample time to answer each question, numerical operations and coding speed tests are "speed tests," and the scores are based not only on the accuracy of the answers, but also on the time it takes for one to complete the test.[7] See figure A1 for sample questions from numerical operation and coding speed tests. The ability that these two tests measure tends to be overlooked in the existing literature. The exception is Murnane et al. (2001), who interpret it as the ability to complete "routine mental tasks quickly and accurately." They find that this ability

[7]For a more detailed description, see NLSY79 Codebook Supplement Attachment 106: Profiles of American Youth (https://www.nlsinfo.org/content/cohorts/nlsy79/other-documentation/codebook-supplement/nlsy79-attachment-106-profiles).

is rewarded in the labor market, net of cognitive and self-esteem scores.

Following Heckman et al. (2006), I interpret scores from the Rotter Locus of Control test and the Rosenberg Self-Esteem test as measuring one's socio-emotional ability. I add a third component, which directly asks about each respondent's sociability in 1985 (four-point scale answer: 1. extremely shy; 2. shy; 3. outgoing; 4. extremely outgoing). This sociability measure has been increasingly used in the literature as a proxy for interpersonal skills (Deming, 2017; Speer, 2017).

For each of the four abilities introduced above (cognitive, socio-emotional, manual, routine), I use all relevant raw test scores to create one composite score.[8] Table 2 presents the pairwise correlation coefficients of the composite scores among the four ability dimensions. The cognitive score is moderately correlated with the socio-emontional score and highly correlated with manual and routine scores.[9] This empirical fact motivates the structural model presented later to explicitly account for correlations between test scores.

2.3 Task Content of Occupations

I use O*NET job descriptors to define occupations as task bundles. O*NET (Occupational Information Network), published and periodically updated by the U.S. Department of Labor (DOL), is the nation's primary source on occupational information.[10] For each occupation, O*NET provides a list of descriptors with ratings on importance, level, relevance or extent. The ratings are aggregate measures of two surveys: a survey of workers who rate descriptors of their own occupations and a survey of occupation analysts who rate all descriptors of all occupations.

For each occupation, I use O*NET data to rate the intensity of four tasks: cognitive, interpersonal, manual and routine.[11] The descriptors I choose for each task follow Acemoglu and Autor (2011) and Deming (2017). Appendix B lists descriptors by task. I follow Acemoglu and Autor (2011) in taking the simple mean of the descriptors within each task category and then standardizing the sum to create a composite measure for each task. Table

[8] The procedure is as follows: 1. standardize all relevant raw scores (such that each score has zero mean with a standard deviation of one); 2. sum up the standardized scores; 3. standardize the sum.

[9] Prada and Urzúa (2017) document similar patterns in these test scores.

[10] Its first version was released in 1998. The latest version is O*NET 23, as of October, 2018. In this paper, I use O*NET 4, which was released in April 2002. The reason for using O*NET 4 is twofold. First, this is consistent with the choice of Acemoglu and Autor (2011). Second, task content of occupations may change over time. Since I observe workers from 1979 to 2014, it is helpful to minimize this concern by selecting a year in the middle. Additionally, emails from staff members at DOL confirm that version-to-version task content changes for occupations are minor.

[11] The NLSY79 uses census occupation codes, while the O*NET uses Standard Occupation Classification (SOC) codes. Acemoglu and Autor (2011) map both classifications to 326 occupations, based on the 1990 census code. The crosswalk is available at `https://economics.mit.edu/faculty/dautor/data/acemoglu`.

A3 shows the top- and bottom-ranked occupations by task intensity.

Recognizing occupations as bundles of tasks provides a new way to categorize occupations. Based on whether the standardized composite task intensity measure is greater than zero, I classify each task of each occupation to be of either high or low intensity. Using 1 to denote high task intensity, 0 to denote low task intensity and arranging the tasks in the order of *cognitive*, *interpersonal*, *manual* and *routine*, I collapse all occupations into 16 distinct categories. Table 3 shows the distribution of occupation choices using this categorization from the 2000 census and the NLSY79 sample. The patterns from the NLSY79 sample largely agree with those from the 2000 census.[12] The most striking pattern is that two task-bundle-defined occupations, (1,1,0,0) and (0,0,1,1), account for around 60% of the civilian labor force, suggesting that the cognitive-interpersonal tasks pair and the manual-routine tasks pair may be complementary.[13]

2.4 Data Patterns

In this section, I analyze the NLSY79 data to confirm patterns documented in recent literature. I first analyze the associations between observed wages and measures of cognitive and socio-emotional abilities by age using OLS regressions.[14] Because of high correlations between AFQT and the constructed manual and routine scores, I omit manual and routine scores from this preliminary analysis.[15] Figure 1 plots percentage gains in wages associated with a 0.1 standard deviation increase in AFQT, the Rosenberg Self-Esteem, and the (reverse) Rotter Locus of Control scores, respectively, for men in the representative sample of NLSY79. As a reminder, AFQT measures cognitive ability; a higher Rosenberg score indicates higher self-esteem; and a higher (reverse) Rotter score indicates higher internal control. The wage gain associated with higher cognitive performance increases from close to zero at age 20 to 1.8% at age 36, and fluctuates around 2% afterwards with a peak of 2.2% at age 44. The wage gains associated with socio-emotional measures are positive and consistent

[12] Differences are not surprising since the NLSY79 sample is a panel sample of one cohort spanning 35 years, whereas the 2000 census provides a cross-section snapshot of the entire labor force. Moreover, part of the difference may be due to change in labor demand over time. Recent literature documents a gradual structural shift away from routine-intensive occupations, due to increased supply of college educated workers (Autor et al., 2008), automation (Autor et al., 2003; Autor and Dorn, 2013) and trade shocks (Autor et al., 2013). However, with panel data of one cohort, it is impossible to separate cohort effects from age effects. Hence, consideration of change in labor demand over time is outside the scope of this paper.

[13] (1,1,0,0) largely corresponds to a typical white-collar occupation, such as various types of managers, while (0,0,1,1) largely corresponds to an occupation of a typical factory worker.

[14] See notes to figure 1 for details on other controls included in the model.

[15] Prada and Urzúa (2017) find that higher manual ability reduces the likelihood of 4-year college attendance, contrary to cognitive and socio-emotional abilities. Controlling for AFQT and self-esteem scores, Murnane et al. (2001) find that the ability to perform routine tasks quickly and accurately brings large wage returns at ages 27-28, on the same order as AFQT.

throughout different ages, but much more modest, with estimates for some ages statistically indistinguishable from zero.

Next, I plot the mean wage levels by age for the NLSY79 sample and by task content of occupations. As shown in figure 2, the mean wage rises as respondents age.[16] A wage gap between the high cognitive/interpersonal intensity occupations and the high manual/routine intensity occupations starts at age 22 and gradually gets wider over time, resulting in very large wage differences after age 40. For example, at age 45, the mean wage level of high cognitive occupations ($42.78) is about 70% higher than the mean wage level of high routine occupations ($25.02).

These empirical facts provide strong motivation for a human capital accumulation model in which agents with multi-dimensional abilities dynamically make schooling decisions and subsequently sort into occupations based on the task content.

3 Model

In this section, I introduce a dynamic human capital accumulation model with a latent factor structure. The model combines features from the dynamic career choice model by Keane and Wolpin (1997) and factor models, developed in a series of papers (e.g., Carneiro et al. 2003; Heckman and Navarro 2007; Abbring and Heckman 2007; Heckman et al. 2016).

3.1 Overview of the Model

I model individuals as having four dimensions of latent ability factors, observed by individuals but unobservable to econometricians. The ability factors affect schooling, labor supply and occupation choices throughout the individual's lifetime. They do not change over time, nor are they affected by aforementioned decisions. Using i to denote the individual, I denote the vector of ability factors as $\mathbf{E}_i = (e_{i1}, e_{i2}, e_{i3}, e_{i4})$. In particular, e_{i1}, e_{i2}, e_{i3} and e_{i4} denote individual i's ability to perform *cognitive*, *socio-emotional*, *manual* and *routine* tasks, respectively.

Individuals, conditional on ability factors, solve a finite-horizon, discrete-time dynamic programming problem. In each period, each individual chooses one of the three alternative actions: work, remain unemployed or attend school.[17] If the individual chooses to work, he picks one occupation among the set of available occupations described in section 3.2. Utility

[16] Table A2 accompanies figure 2 to present wage summary statistics for selected ages.

[17] Keane and Wolpin (1997) label individuals who are neither working or enrolled in school as engaging in "home production". This paper labels them as "unemployed". Whether these individuals engage in home production or not while unemployed is outside the scope of the paper.

flows are choice-specific. Individuals maximize the present discounted value of expected lifetime utility.

3.2 Describing Occupations

I model occupations as bundles of discrete tasks along four dimensions that correspond to ability factors: *cognitive*, *interpersonal*, *manual* and *routine* tasks.[18] Let j denote a particular task and τ an indicator of whether the task is intensive for a given occupation. I use O*NET data to rate the task intensity of all occupations at the 3-digit census code level, as described in section 2.3. Using 1 to indicate high task intensity and 0 to indicate low task intensity, I describe all occupations as four-dimensional vectors with binary elements. The set of all potential occupations in the model can be expressed as

$$\mathcal{O} = \{(\tau_1, \tau_2, \tau_3, \tau_4) | \tau_j \in \{0, 1\}, j = 1, 2, 3, 4.\}. \tag{3.1}$$

As an example, a mathematician would be described as having the vector (1,1,0,0) as his occupation (high intensity in cognitive and interpersonal tasks, low intensity in manual and routine tasks). Note that the set of all possible occupations, $\mathcal{O}$, has 16 elements, and each element corresponds to a particular high/low intensity combination of the four task indicators.

3.3 Choice Set

At the beginning of each period, each agent can choose to work in a particular occupation (o), be unemployed (ne) or go to school (s). While working or unemployed, the individual can simultaneously choose to get the GED (g). I do not allow getting the GED to be a stand-alone choice.[19]

Let d_{it} denote the choice individual i makes in period t. Suppressing i, the set of choices for an individual entering period t is

$$\mathcal{D}_t = \{s, ne, (ne, g), [o, \forall o \in \mathcal{O}], [(o, g), \forall o \in \mathcal{O}]\}. \tag{3.2}$$

[18]The idea of modeling occupations as bundles of tasks is not new. See Gibbons and Waldman (2004), Acemoglu and Autor (2011) and Autor and Handel (2013) for conceptual frameworks.

[19]The final sample consists of 1,379 respondents, and 115 (8.3%) of them have ever obtained the GED. Hence, it is important to include the GED as a choice and the receipt of GED as a terminal schooling level. The decision to not model obtaining the GED as a stand-alone choice is based on the fact that the majority of respondents got their GED while working or unemployed. Sullivan (2010) models the choice to obtain the GED in a similar fashion.

Since there are 16 different occupations in the model, this choice set has 35 elements. Note that d_t could denote combination choices if getting the GED is part of the decision.[20]

Getting the GED drops out of the choice set after the individual graduates high school or obtains the GED certificate, so the choice set becomes

$$\mathcal{D}_t = \{s, ne, [o, \forall o \in \mathcal{O}]\} \tag{3.3}$$

with 18 elements.

If an individual has reached the highest level of schooling, then schooling drops out of the choice set, so the choice set becomes

$$\mathcal{D}_t = \{ne, [o, \forall o \in \mathcal{O}]\} \tag{3.4}$$

with 17 elements.[21]

3.4 Utility Function

The utility flow for individual i in period t depends on state variables ($\mathbf{S}_{it}$), decision (d_{it}), a set of observed exogenous characteristics ($\mathbf{X}_{it}$) and the unobserved individual-specific ability factors ($\mathbf{E}_i$). I model utility as having two components: wage income and non-pecuniary benefits,

$$U(d_{it}, \mathbf{S}_{it}, \mathbf{X}_{it}, \mathbf{E}_i) = w(d_{it}, \mathbf{S}_{it}, \mathbf{X}_{it}, \mathbf{E}_i) + H(d_{it}, \mathbf{S}_{it}, \mathbf{X}_{it}, \mathbf{E}_i), \tag{3.5}$$

where $w(\cdot)$ denotes the wage income component and $H(\cdot)$ denotes the non-pecuniary benefit component. The wage income component is set to zero if the individual does not work.

3.4.1 Wage

An individual i, conditional on working in period t, chooses an occupation $o_{it} \in \mathcal{O}$, and receives a wage w_{it}. I write the wage function as

$$w(o_{it}, \mathbf{S}_{it}, \mathbf{X}_{it}, \mathbf{E}_i) = \bar{w}(o_{it}, \mathbf{S}_{it}, \mathbf{X}_{it}, \mathbf{E}_i) + \epsilon^w_{iot}, \tag{3.6}$$

[20] In case of combination decisions, such as $d_t = (ne, g)$, I use $g \in d_t$ to denote the fact that getting the GED is part of the combination decision.

[21] The highest possible level of schooling allowed in this model is 20 years of education. This is because the highest graded completed variable is topcoded at 20 years in NLSY79.

where $\bar{w}(\cdot)$ is the deterministic portion of the wage equation and the wage error term (ϵ^w_{iot}) is individual-, occupation- and time-specific.

3.4.2 Non-Pecuniary Utility

I write the non-pecuniary benefit component of the utility function as

$$H(d_{it}, \mathbf{S}_{it}, \mathbf{X}_{it}, \mathbf{E}_i) = \bar{H}(d_{it}, \mathbf{S}_{it}, \mathbf{X}_{it}, \mathbf{E}_i) + \epsilon^u_{idt}, \tag{3.7}$$

where $\bar{H}(\cdot)$ is the choice-specific deterministic portion of the non-pecuniary utility flow and the random shock (ϵ^u_{idt}) to the utility flow is individual-, decision- and time-specific. Suppressing $\mathbf{S}_{it}, \mathbf{X}_{it}$ and $\mathbf{E}_i$, I further specify the components of the deterministic portion of the non-pecuniary utility as

$$\bar{H}(d_{it}) = H^s_{it} \cdot \mathbb{1}(d_{it} = s) + H^{ne}_{it} \cdot \mathbb{1}(ne \in d_{it}) + H^g_{it} \cdot \mathbb{1}(g \in d_{it}) + H^o_{it} \cdot \mathbb{1}(o \in d_{it}), \tag{3.8}$$

where H^s_{it}, H^{ne}_{it}, and H^g_{it} are the deterministic parts of utility flows of schooling, unemployment and getting the GED respectively, and H^o_{it} is the deterministic non-pecuniary utility flow of working in occupation $o \in \mathcal{O}$.

3.5 State Space

3.5.1 Schooling

Education levels affect wages. I allow the wage return to schooling to vary by level and by the task content of occupations. I distinguish between three different levels of schooling: years of high school (h_{it}), years of higher education (c_{it}) and an indicator for GED receipt (g_{it}).[22]

3.5.2 Human Capital

Aside from schooling, I also assume that individuals can obtain human capital from work experience, which in turn affects both wages and non-pecuniary utility flows. I keep track of five dimensions of human capital: general human capital (k_{i0t}) and four dimensions of task-specific human capital ($k_{ijt}, j = 1, 2, 3, 4.$). The inclusion of general human capital captures the idea that part of human capital accumulated on the job is general and not task-specific, such as rudimentary reading and writing, following basic instructions and being

[22]Higher education includes both college and graduate education.

punctual.[23]

I assume each dimension of human capital can take only M discrete values.[24] Arranging the values in ascending order, I specify the sets of possible values for general human capital and task-specific human capital as

$$k_{jt} \in \{K_j(1), ..., K_j(M)\}, j = 0, 1, 2, 3, 4.$$

I assume $M = 3$ in this paper to keep the size of the state space a reasonable number for computation.

After each year of work, regardless of occupation choice, the individual's general human capital may increase to the next higher level with probability p_0, or stay the same with probability $1 - p_0$. The task-specific human captial accumulation process is similar to the process of general human capital accumulation, except that the individual may accumulate only task-specific human capital along the task dimension that his current occupation requires high intensity of. For example, in the context of the model, if an individual works in occupation $o = (1, 1, 0, 0)$ for a period, then he may accumulate human capital only along task dimensions 1 and 2. I use p_j to denote the task-specific human capital growth probability for task j.

This specification of human capital accumulated on the job is very flexible. Since an individual can accumulate human capital along five dimensions, the model allows up to $3^5 = 243$ terminal human capital level combinations. Moreover, given a switch in occupation, the model allows the worker's human capital to be partially transferable if the new and old occupations have an overlap of tasks.[25]

3.5.3 State Variables

The endogenous state variables include the years of different types of schooling and total stock of human capital along the five task dimensions. Let $\boldsymbol{\epsilon_t^w}$ and $\boldsymbol{\epsilon_t^u}$ denote the vectors of occupation-specific wage errors and decision-specific utility shocks in period t. Dropping the subscript i and using a_t to denote age in period t, the vector of state variables is

$$\mathbf{S_t} = (a_t, h_t, c_t, g_t, k_{0t}, k_{1t}, k_{2t}, k_{3t}, k_{4t}, \boldsymbol{\epsilon_t^w}, \boldsymbol{\epsilon_t^u}).$$

[23] For ease of exposition, I reserve the use of the term "human capital" exclusively for human capital accumulated on the job henceforth.

[24] This method of modelling human capital accumulation was originally suggested by John Kennan to my advisor, Steve Stern.

[25] For example, if a programmer (occupation (1,0,0,0)) changes occupation to a manager (occupation (1,1,0,0)), then the human capital accumulated on cognitive tasks as a programmer directly transfers to the new occupation as a manager, because it also requires cognitive tasks.

3.6 Dynamic Programming Problem

An individual maximizes the present value of discounted lifetime expected utility from age 16 ($t = 1$) to a terminal age $t = T^{**}$. At the beginning, each individual takes ability factors ($\mathbf{E}$) as given and make decisions accordingly. State variables evolve as a result of the individual's decisions according to rules specified in section 3.5, and in turn affect his current utility flows. Let δ denote the discount factor. Suppressing the subscript for individual (i), the value function for choosing a decision (d) in period t is

$$V(d_t, \mathbf{X}_t, \mathbf{S}_t, \mathbf{E}) = U(d_t, \mathbf{X}_t, \mathbf{S}_t, \mathbf{E}) + \delta \mathbb{E} \max_{d_{t+1} \in \mathcal{D}_{t+1}} V(d_{t+1}, \mathbf{X}_{t+1}, \mathbf{S}_{t+1}, \mathbf{E}), \tag{3.9}$$

where the choice set for the next period, $\mathcal{D}_{t+1}$, depends on current state variables, $\mathbf{S}_t$, as defined in section 3.3. The expectation is taken over random shocks to wages, randomness in the human capital accumulation process and random shocks to utility levels.

4 Estimation

In this section, I describe a measurement system of latent abilities, specific parametric assumptions of utility flows and error structures of the model. Given these assumptions, I then describe the solution to individual's dynamic programming problem, the likelihood function and how to evaluate the likelihood function and implement the optimization algorithm. I conclude with a discussion on how model parameters are identified.

4.1 A Measurement System of Latent Ability Factors

Individuals' latent abilities, which correspond to the task content of occupations, are measured around the time of labor market entry. These measures are assumed to be fallible to account for reverse causality and potential measurement error.[26]

For each individual i, I observe a vector of test scores corresponding to the four task dimensions, denoted $\mathbf{R}_i = (R_{i1}, R_{i2}, R_{i3}, R_{i4})$, which are assumed to be linear functions of individual i's own and family background characteristics at the time tests are taken ($\mathbf{Z}_i$), his ability factors ($\mathbf{E}_i = (e_{i1}, e_{i2}, e_{i3}, e_{i4})$) and a vector of additively separable idiosyncratic

[26]Individuals' own education levels and family background at the time of test may affect test scores designed to measure abilities, raising concerns about reverse causality of education on test scores. For example, Neal and Johnson (1996) find that one additional year of education raises the AFQT score by 2 to 4 percentage points. This paper follows the method developed by Carneiro et al. (2003) and Hansen et al. (2004) to assume that a vector of latent ability factors generate both test scores and later life outcomes, including schooling choices. Heckman et al. (2006) and Cunha et al. (2010), among others, also employ this method to model test scores.

errors ($\boldsymbol{\epsilon}_i^e$),

$$\begin{cases} R_{i1} &= \mathbf{Z}_i\boldsymbol{\lambda}_1 + e_{i1} + \epsilon_{i1}^e \\ R_{i2} &= \mathbf{Z}_i\boldsymbol{\lambda}_2 + \eta_2 e_{i1} + e_{i2} + \epsilon_{i2}^e \\ R_{i3} &= \mathbf{Z}_i\boldsymbol{\lambda}_3 + \eta_3 e_{i1} + e_{i3} + \epsilon_{i3}^e \\ R_{i4} &= \mathbf{Z}_i\boldsymbol{\lambda}_4 + \eta_4 e_{i1} + e_{i4} + \epsilon_{i4}^e, \end{cases} \tag{4.1}$$

where e_{i1}, e_{i2}, e_{i3} and e_{i4} correspond to individual i's latent *cognitive*, *socio-emotional*, *manual* and *routine* abilities, respectively, and R_{i1}, R_{i2}, R_{i3} and R_{i4} correspond to the test scores designed to measure them. Following Prada and Urzúa (2017), I allow e_1 to enter the equation for all other scores to account for the positive correlations between these constructed scores and the cognitive score (see table 2).

4.2 Further Model Specifications

4.2.1 Parametric Assumptions of Utility Functions

Given occupation o_{it}, I specify the deterministic portion of the wage equation (eq. 3.6) as an augmented Mincer (1958, 1974) earnings equation,

$$\begin{aligned} \bar{w}_{it} = \beta_0^w + \beta_h^w h_{it} + \beta_c^w c_{it} + \beta_g^w g_{it} + \beta_{0k1}^w \mathbb{1}(k_{i0t} = K_0(1)) + \beta_{0k2}^w \mathbb{1}(k_{i0t} = K_0(2)) + \beta_{0k3}^w \mathbb{1}(k_{i0t} = K_0(3)) \\ + \sum_{j=1}^{4} \Big[\beta_{j0}^w + \beta_{jh}^w h_{it} + \beta_{jc}^w c_{it} + \beta_{jg}^w g_{it} + \beta_{je}^w e_{ij} + \beta_{jk1}^w \mathbb{1}(k_{ijt} = K_j(1)) + \beta_{jk2}^w \mathbb{1}(k_{ijt} = K_j(2)) \\ + \beta_{jk3}^w \mathbb{1}(k_{ijt} = K_j(3))\Big] \mathbb{1}(\tau_j = 1|o_{it}) + \sum_{1\leq m<n\leq 4} \beta_{mn}^w \mathbb{1}(\tau_m = 1 \cup \tau_n = 1|o_{it}). \end{aligned} \tag{4.2}$$

As a reminder, h_{it} denotes years of high school; c_{it} denotes years of higher education, including college and graduate school; g_{it} is a dummy variable indicating receipt of GED; $(k_{i0t}, k_{i1t}, k_{i2t}, k_{i3t}, k_{i4t})$ is the vector of human capital stock with k_{i0t} denoting general human capital and $k_{ijt}, j \in \{1,2,3,4\}$ denoting task-specific human capital; and $(e_{i1}, e_{i2}, e_{i3}, e_{i4})$ is the vector of individual endowment factors along the four task dimensions. The term $\mathbb{1}(\tau_j = 1|o_{it})$ is the indicator that occupation o_{it} requires high intensity along task dimension j, and the term $\mathbb{1}(\tau_m = 1 \cup \tau_n = 1|o_{it})$ is the indicator that occupation o_{it} requires high intensity of both tasks m and n.

It is worth discussing the assumptions about the wage equation. First, I allow the wage return of education to differ by both level and the task content of occupations. For example, advanced education may bring particularly high return for workers in occupations with high cognitive task intensity. Second, a worker i at any given period t has five dimensions of human capital accumulated on the job ($k_{ijt}, j = 0, 1, 2, 3, 4$), and for each dimension, the worker's human capital level can be low, intermediate or high. The worker always draws return from the general human capital conditional on working, but only draws return from the task-specific human capital conditional on working in an occupation that requires the corresponding task. Third, the worker draws additional return from performing specific tasks relative to the baseline occupation, (0,0,0,0). The parameters β^w_{j0} ($j = 1, 2, 3, 4$) quantify the return to tasks, and the parameters β^w_{mn} ($1 \leq m < n \leq 4$) quantify the return to performing any pair of task m and task n together in an occupation (pairwise task complementarity). Lastly, latent abilities also directly affect wage.

Next, I specify the parametric assumptions of the decision-specific non-pecuniary utility functions (eq. 3.8). Utility from attending school (s) is

$$H^s_{it} = \beta^s_0 + \beta^s_{a1} a_{it} + \beta^s_{a2}(a_{it} - 24) * \mathbb{1}(a_{it} \geq 25) + \sum_{j=1}^{4} \beta^s_j e_{ij}. \tag{4.3}$$

Note that schooling utility as a function of age is modeled as a linear spline with one node at age 25. This is based on the empirical fact that the observed rate of schooling choice steadily declines from age 16 to age 25 and remains at a very low level afterwards (see table 1).

The deterministic portion of the utility received from unemployment (u) is set to zero because non-percuniary utility can only be identified up to a base choice, so $H^u_{it} = 0$.[27]

The deterministic portion of the utility received from getting the GED (g) is

$$H^g_{it} = \beta^g_0 + \beta^g_{a1} \mathbb{1}(a_{it} \geq 25). \tag{4.4}$$

The reason for this parsimonious specification is that the choice of getting the GED is extremely rare and the data cannot support estimation of a model with a more sophisticated specification.[28]

[27]This applies to all unordered discrete choice models. See Train (2009) for standard identification issues regarding discrete choice models.

[28]Only 115 (0.26%) of the 44,759 "person-year" observations involve getting the GED, making it difficult to estimate the utility flow with more explanatory variables. In preliminary experiments, when I allowed the four latent abilities to enter the GED utility flow, the estimation program would not converge.

The deterministic portion of the non-pecuniary utility from working in occupation o is

$$H_{it}^o = \beta_0^o + \beta_{0k1}^o \mathbb{1}(k_{i0t} = K_0(1)) + \beta_{0k2}^o \mathbb{1}(k_{i0t} = K_0(2)) + \beta_{0k3}^o \mathbb{1}(k_{i0t} = K_0(3)) + \sum_{j=1}^{4} \Big[\beta_{jk1}^o \mathbb{1}(\beta_j^o + k_{ijt} = K_l(1)) + \beta_{jk2}^o \mathbb{1}(k_{ijt} = K_l(2)) + \beta_{jk3}^o \mathbb{1}(k_{ijt} = K_l(3))\Big] \mathbb{1}(\tau_j = 1 | o_{it}). \tag{4.5}$$

Non-pecuniary utility from work is assumed to vary by task content of occupations and task-specific human capital levels.

4.2.2 Distributional Assumptions

I assume the random shocks in wages to be i.i.d normal with zero mean: $\epsilon_{iot}^w \overset{i.i.d}{\sim} N(0, \sigma_{\epsilon^w}^2)$. The probabilities of general and task-specific human capital growth are specified as standard logit: $p_j = \dfrac{\exp(\alpha_j)}{1+\exp(\alpha_j)}$, $j = 0, 1, 2, 3, 4$. I also assume the ability factors and idiosyncratic errors in the test scores to be i.i.d normal with zero mean: $e_{ij} \overset{i.i.d}{\sim} N(0, \sigma_{e_j}^2)$ and $\epsilon_{ij}^e \overset{i.i.d}{\sim} N(0, \sigma_{\epsilon_j^e}^2)$, $j = 1, 2, 3, 4$.

To close the model, I assume the decision-specific utility shock, ϵ_{idt}^u, to be i.i.d extreme value with mean $\sigma_{\epsilon^u}\gamma$ and variance $\pi^2\sigma_{\epsilon^u}^2/6$, where γ denotes Euler's constant. Lastly, the discount factor is fixed at 0.95.[29]

4.3 Solution to Dynamic Programming Problem

I use the standard backwards induction technique to solve the finite horizon dynamic programming problem. I assume that agents stop making decisions at age T^* and die at age T^{**}.[30] I also assume the vector of state variables, $\mathbf{S}_{it}$, to stay constant after T^*. The value

[29] In this model, identification of the discount factor relies on functional form assumptions. See Magnac and Thesmar (2002) for identification issues in finite horizon dynamic discrete choice models. In practice, Rust and Phelan (1997) find the value of the likelihood function to be very flat with respect to the discount factor in a model of dynamic retirement behavior. Existing literature has largely estimated the discount factor to be between 0.93 and 0.97, such as Keane and Wolpin (1997) (0.936), Gourinchas and Parker (2002) (0.966) and French and Jones (2011) (0.945; median type). In this paper, I follow a long list of dynamic labor supply studies to fix the discount factor at 0.95 (e.g., Berkovec and Stern 1991; Sullivan 2010; Pavan 2011).

[30] In this paper, I set $T^* = 60$ and $T^{**} = 65$. The choice for T^{**} is easy to justify because 65 is the full retirement age. As for T^*, let T_i denote the age I stop observing choices made by individual i. In my dataset, I can observe individuals up to around 50 years of age. The Turnpike Theorem requires picking a T^* large enough such that $\delta^{T^*-T_i}$ is small enough to ensure that, for any individual i, deviations in optimal decisions after T^* do not affect optimal decisions up to time T_i (Neumann, 1945; McKenzie, 1976; Yano, 1984).

function for individual i at age T^* is

$$V_{T^*}(d_{T^*}, \mathbf{S}_{iT^*}, \mathbf{E}_i) = U_{T^*}(d_{T^*}, \mathbf{S}_{iT^*}, \mathbf{E}_i) + \mathbb{E} \sum_{t=T^*+1}^{T^{**}} \delta^{t-T^*} U_t(d_{T^*}, \mathbf{S}_{iT^*}, \mathbf{E}_i), \tag{4.6}$$

where the expectation is taken over two sources of randomness: random wage shocks ϵ_{it}^w and random utility shocks ϵ_{it}^u, for $t = T^*+1, T^*+2, ..., T^{**}$. Then $V_t(d_t, \mathbf{S}_{it}, \mathbf{E}_i)$ can be evaluated recursively for all $t < T^*$. The solution to the dynamic programming problem provides the choice-specific value functions used in the construction of the likelihood function.

The $\mathbb{E}$max expressions in the dynamic programming program (eq. 3.6) do not have closed-form solutions, so they are approximated using simulation methods. Because of the assumption that the choice-specific utility shock (ϵ^u) is distributed extreme value, conditional on the other random components ($\boldsymbol{\epsilon}_{it}^w$), the vector of ability factors ($\mathbf{E}_i$) and the unobserved part of the state variables ($\mathbf{S}_{it}^2$), the expected maximum has the following closed-form solution:

$$\begin{aligned}\mathbb{E} \max_{d_{it} \in \mathcal{D}_{it}} \{V(d_{it} | \boldsymbol{\epsilon}_{it}^w, \mathbf{S}_{it}^2, \mathbf{E}_i)\} &= \sigma_{\epsilon^u}\{\gamma + \ln[\sum_{d_{it} \in \mathcal{D}_{it}} \exp(\frac{\bar{V}(d_{it} | \boldsymbol{\epsilon}_{it}^w, \mathbf{S}_{it}^2, \mathbf{E}_i)}{\sigma_{\epsilon^u}})]\} && (4.7)\\ &\equiv \Psi(\boldsymbol{\epsilon}_{it}^w, \mathbf{S}_{it}^2, \mathbf{E}_i), && (4.8)\end{aligned}$$

where $\bar{V}(d_{it}) = V(d_{it}) - \epsilon_{idt}^u$. Integration over the distributions of $\boldsymbol{\epsilon}_{it}^w$ and $\mathbf{S}_{it}^2$ provides the unconditional expected maximum for an individual with ability vector $\mathbf{E}_i$. Let $f(\cdot)$ denote the probability density function.

$$\mathbb{E} \max_{d_{it} \in \mathcal{D}_{it}} \{V(d_{it} | \mathbf{E}_i)\} = \int [\int \Psi(\boldsymbol{\epsilon}_{it}^w, \mathbf{S}_{it}^2, \mathbf{E}_i) f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_{it}^w) d\boldsymbol{\epsilon}_{it}^w] f_{S^2}(\mathbf{S}_{it}^2 | \mathbf{S}_{i,t-1}^2, d_{i,t-1}, \mathbf{E}_i) d\mathbf{S}_{it}^2. \tag{4.9}$$

The integral over $\boldsymbol{\epsilon}_{it}^w$ in equation 4.9 does not have an analytical solution, so it is approximated by simulation. Specifically, the integral is approximated with the average value of the integrand, given Q draws from the joint density $f_{\boldsymbol{\epsilon}^w}(\cdot)$. The integral over $\mathbf{S}_{it}^2$ is in fact a weighted sum because the distribution of $\mathbf{S}_{it}^2$ is discrete. Let q index simulation draws. Let $supp(\mathbf{S}_t^2)$ denote the support of of the probability density function of $\mathbf{S}_{it}^2$. As a result,

$$\mathbb{E} \max_{d_{it} \in \mathcal{D}_{it}} \{V(d_{it} | \mathbf{E}_i)\} \approx \frac{1}{Q} \sum_{q=1}^{Q} [\sum_{\mathbf{S} \in supp(\mathbf{S}_t^2)} \Pr(\mathbf{S}_{it}^2 = \mathbf{S} | \mathbf{S}_{i,t-1}^2, d_{i,t-1}, \mathbf{E}_i) \cdot \Psi(\boldsymbol{\epsilon}_{it}^{w,q}, \mathbf{S}_{it}^2 = \mathbf{S}, \mathbf{E}_i)]. \tag{4.10}$$

4.4 The Likelihood Function

Each individual contributes two parts to the likelihood function. The first part comes from the solution to the dynamic programming problem. Individual i is observed from period 1 to period T_i and provides data $\{\tilde{d}_{it}, w_{it}(\tilde{d}_{it})\}_{t=1}^{T_i}$, where $\tilde{d}_{it}$ is the observed choice in period t, and $w_{it}(\tilde{d}_{it})$ is the wage associated with choice $\tilde{d}_{it}$. Note that $w_{it}(\tilde{d}_{it})$ may be unobserved. Let $\mathbf{\Gamma}_{iT}$ denote a particular realization of $\{\mathbf{S}_{it}^2\}_{t=1}^{T}$. Denoting the set of parameters to be estimated as $\mathbf{\Theta}$, the first part of the likelihood contribution from individual i, conditional on his unobserved abilities and a particular realization of $\mathbf{\Gamma}_{iT_i}$, is

$$L_{i,e,S^2}^1(\mathbf{\Theta}|\mathbf{E}_i, \mathbf{\Gamma}_{iT_i}) = \prod_{t=1}^{T_i} \Pr(\tilde{d}_{it}|w_{it}(\tilde{d}_{it}), \mathbf{S}_{it}^1, \mathbf{S}_{it}^2, \mathbf{E}_i) \Pr(w_{it}(\tilde{d}_{it})|\mathbf{S}_{it}^1, \mathbf{S}_{it}^2, \mathbf{E}_i). \tag{4.11}$$

Given that ϵ^u is distributed i.i.d standard extreme value, if $w_{it}(\tilde{d}_{it})$ is observed, then

$$\begin{aligned}
&\Pr(\tilde{d}_{it}|w_{it}(\tilde{d}_{it}), \mathbf{S}_{it}^1, \mathbf{S}_{it}^2, \mathbf{E}_i) \Pr(w_{it}(\tilde{d}_{it})|\mathbf{S}_{it}^1, \mathbf{S}_{it}^2, \mathbf{E}_i) \\
=&\Big[\int \frac{\exp(\bar{V}(\tilde{d}_{it}|w_{it}(\tilde{d}_{it}), \mathbf{S}_{it}^1, \mathbf{S}_{it}^2, \mathbf{E}_i)/\sigma_{\epsilon^u})}{\sum_{d_{it}\in\mathcal{D}_t} \exp(\bar{V}(d_{it}|w_{it}(\tilde{d}_{it}), \mathbf{S}_{it}^1, \mathbf{S}_{it}^2, \mathbf{E}_i)/\sigma_{\epsilon^u})} dF_{\boldsymbol{\epsilon}^w}(\boldsymbol{\epsilon}_{it}^w \backslash \epsilon_{it}^w(\tilde{d}_{it})|\epsilon_{it}^w(\tilde{d}_{it}))\Big] f_{\epsilon^w}(\epsilon_{it}^w(\tilde{d}_{it})|w_{it}(\tilde{d}_{it}), \mathbf{S}_{it}^1, \mathbf{S}_{it}^2, \mathbf{E}_i),
\end{aligned}$$

where $\epsilon_{it}^w(d_{it})$ denotes the wage error associated with choice d_{it}, and F denotes the cumulative distribution function. If $w_{it}(\tilde{d}_{it})$ is not observed, then

$$\begin{aligned}
&\Pr(\tilde{d}_{it}|w_{it}(\tilde{d}_{it}), \mathbf{S}_{it}^1, \mathbf{S}_{it}^2, \mathbf{E}_i) \Pr(w_{it}(\tilde{d}_{it})|\mathbf{S}_{it}^1, \mathbf{S}_{it}^2, \mathbf{E}_i) \\
=&\int \frac{\exp(\bar{V}(\tilde{d}_{it}|w_{it}(\tilde{d}_{it}), \mathbf{S}_{it}^1, \mathbf{S}_{it}^2, \mathbf{E}_i)/\sigma_{\epsilon^u})}{\sum_{d_{it}\in\mathcal{D}_t} \exp(\bar{V}(d_{it}|w_{it}(\tilde{d}_{it}), \mathbf{S}_{it}^1, \mathbf{S}_{it}^2, \mathbf{E}_i)/\sigma_{\epsilon^u})} dF_{\boldsymbol{\epsilon}^w}(\boldsymbol{\epsilon}_{it}^w).
\end{aligned}$$

Next, note that $\mathbf{S}_{it}^2$ is unobserved for all periods. I need to integrate over the joint distribution of all elements of $\mathbf{S}_{it}^2$ over the entire T_i periods that person i is in the sample, conditional on the observed sequence of choices up to period T_i, so

$$L_{i,e}^1(\mathbf{\Theta}|\mathbf{E}_i) = \int_{supp(\mathbf{\Gamma}_{iT_i})} \Big[\prod_{t=1}^{T_i} \Pr(\tilde{d}_{it}|w_{it}(\tilde{d}_{it}), \mathbf{S}_{it}^1, \mathbf{S}_{it}^2, \mathbf{E}_i) \Pr(w_{it}(\tilde{d}_{it})|\mathbf{S}_{it}^1, \mathbf{S}_{it}^2, \mathbf{E}_i)\Big] dF_{\mathbf{\Gamma}_{iT_i}}(\mathbf{\Gamma}_{iT_i}|d_{i1}, \cdots, d_{i,T_i-1}, \mathbf{E}_i), \tag{4.12}$$

where $supp(\mathbf{\Gamma})$ denotes the support of the probability density function of $\mathbf{\Gamma}$.

The second part of an individual's likelihood contribution comes from the test scores ($\mathbf{R}_i$), which provide noisy measures of the individual's ability factors.

$$L_{i,e}^2(\mathbf{\Theta}|\mathbf{E}_i) = f_{\boldsymbol{\epsilon}^e}(\boldsymbol{\epsilon}^e|\mathbf{R}_i, \mathbf{E}_i). \tag{4.13}$$

Finally, given a sample of N individuals, the likelihood function is

$$L(\mathbf{\Theta}) = \prod_{i=1}^{N}[\int L_i^1(\mathbf{\Theta}|\mathbf{E}_i)L_i^2(\mathbf{\Theta}|\mathbf{E}_i)dF_E(\mathbf{E})]. \tag{4.14}$$

4.5 Likelihood Function Evaluation and Estimation Strategy

4.5.1 Likelihood Contribution from the Dynamic Programming Problem

Two issues related to computing the likelihood contribution from the dynamic programming problem merit discussion. First, the summation in equation 4.12 cannot be computed with frequency simulators because human capital levels are unobserved, and there are an impractically large number of possible realizations of the sequence of human capital levels. The solution is to use a modified GHK simulator[31] to directly simulate entire paths of human capital levels multiple times and take the average.[32]

Second, about 3.2% of the person-year observations are missing, creating "holes" in career history spells. This causes problems because sequentially simulating the unobserved part of the state space depends on observed decisions. The solution is to simulate the unobserved decisions jointly with the unobserved part of the state space.

Let individual i have endowment $\mathbf{E}_i$, and denote the value of the expression that needs to be evaluated $I(\mathbf{E}_i)$ (left-hand side of equation 4.12). For individual i observed up to T_i periods, the steps to calculate $I(\mathbf{E}_i)$ are:

1. For a particular draw, indexed q, from the beginning, set I^q=1.

2. For any $t \in \{2, ..., \bar{T}_i\}$, if $d_{i,t-1}$ is observed, draw $\mathbf{S}_{it}^2|\mathbf{S}_{i,t-1}^2, d_{i,t-1}, \mathbf{E}_i$; otherwise, simulate $\boldsymbol{\epsilon}_{i,t-1}^w$ and $\boldsymbol{\epsilon}_{i,t-1}^u$, and pick $d_{i,t-1} = \text{argmax } V(d_{i,t-1}|\mathbf{S}_{i,t-1}^2, \mathbf{E}_i, \boldsymbol{\epsilon}_{i,t-1}^w, \boldsymbol{\epsilon}_{i,t-1}^u)$, then draw $\mathbf{S}_{it}^2|\mathbf{S}_{i,t-1}^2, d_{i,t-1}, \mathbf{E}_i$.

3. Update $I^q = I^q * \Pr(d_{it}|w_{it}(d_{it}), \mathbf{S}_{it}^2, \mathbf{E}_i) * \Pr(w_{it}(d_{it})|\mathbf{S}_{it}^2, \mathbf{E}_i)$ if and only if d_{it} is observed.

4. For $t = \bar{T}_i + 1$, stop. Go to the next step.

[31] See Geweke (1989, 1991); Hajivassiliou and McFadden (1998); Keane (1994).

[32] Brien et al. (2006) and Sullivan (2010) have used this technique to estimate dynamic models.

5. Repeat these steps for a total of Q times. $I(\mathbf{E}_i) \approx \frac{1}{Q}\sum_{q=1}^{Q} I^q$.

The benefit of the algorithm is that it is easy to simulate $\mathbf{S}^2_{it}$ conditional on $\mathbf{S}^2_{i,t-1}$, $d_{i,t-1}$ and $\mathbf{E}_i$. Comparing to drawing entire sequences of realizations of human capital levels, this method is straightforward and avoids overwhelming computational burden.

4.5.2 Integration over Latent Endowment Factors

The likelihood contribution of both the dynamic programming problem and the test scores are conditional on the individual-specific unobserved endowment factors $\mathbf{E}_i$. Integration over $\mathbf{E}_i$ for each individual delivers the unconditional likelihood. The integral (eq. 4.14) is evaluated by simulation. Specifically, for each individual i, I draw $\mathbf{E}_i$ P times. Using p to index each draw, equation 4.14 can be approximated with the following simulated likelihood:

$$L^s(\mathbf{\Theta}) = \prod_{i=1}^{N}\{\frac{1}{P}\sum_{p=1}^{P}[L_i^1(\mathbf{\Theta}|\mathbf{E}_i^p)L_i^2(\mathbf{\Theta}|\mathbf{E}_i^p)]\}. \tag{4.15}$$

4.5.3 Estimation Strategy

The estimator $\hat{\mathbf{\Theta}}$ is the vector of parameter values that maximizes the simulated likelihood function L^s.[33,34] The simulated likelihood is a function of the 117 parameters in the fully-specified model and is maximized using the *BHHH* optimization algorithm (Berndt et al. 1974).[35]

The main challenge of estimation is the computational burden associated with solving for the value functions from the dynamic programming problem, given the large state space size (5.35 million) and the large number of parameters (117). Since the optimization algorithm is an iterative convergent process, the value functions need to be evaluated repeatedly until convergence. Previous literature has dealt with the computational burden by interpolating

[33] Although the Maximum Simulated Likelihood (MSL) estimator is inconsistent in theory, (Börsch-Supan and Hajivassiliou (1993) show that, in practice, the MSL estimator has negligible bias even with a modest number of simulations. In all their examples with multi-dimensional correlated unobservables, 20 replications (without antithetic acceleration) were sufficient to produce estimators with negligible bias. Geweke (1988) shows that antithetic acceleration significantly reduces the number of simulations needed to produce consistent estimators. Specifically, as the sample size increases, the number of replications required with antithetic acceleration relative to the number required with random sampling is inversely proportional to sample size. Drawing on these results, for all simulation procedures described in sections 4.3–4.5, I perform the simulation 10 times with antithetic acceleration.

[34] The Maximum Likelihood method is frequently employed in estimation of such dynamic structural models. Examples include Keane and Wolpin (1997), Rust and Phelan (1997), Brien et al. (2006), Blau and Gilleskie (2008), Sullivan (2010), Pavan (2011) and Chan (2013). See Keane et al. (2011) and Stern (1997) for further review.

[35] Calculation of the covariance matrix of the maximum likelihood estimator follows naturally from the BHHH algorithm.

the value functions[36] or making further simplifying assumptions in estimation to reduce state space size.[37] The computational challenge in this paper, however, is overcome by parallel computing.[38] This method does not require further restrictive assumptions of the model. Also, by calculating the value functions directly, I avoid potential errors introduced by the interpolation procedure.

4.6 Identification

The model in this paper is a fully-specified dynamic discrete choice model with associated outcomes and a set of latent factors. The goal is to estimate the full set of parameters and the distribution of unobservables. Identification for dynamic factor models is established in a series of papers, including Heckman and Navarro (2007), Hu and Schennach (2008), Blevins (2014), Heckman et al. (2016) and Freyberger (2018).[39] In particular, Freyberger (2018) establishes sufficient conditions for point identification of parameters in a model with a set of latent factors entering unknown structural outcome functions non-additively. Heckman et al. (2016) build on Freyberger's results and establish identification for dynamic discrete choice models with associated outcomes and a factor structure. The formal identification result of my model follows Heckman et al. (2016).[40]

It is still valuable to provide basic intuition on the identification of model parameters. To begin, I aggregate model parameters into broad categories for ease of exposition. Let $\boldsymbol{\beta^{w,1}}$ denote all parameters from the log wage equation (eq. 4.2), except for parameters associated with human capital levels, denoted $\boldsymbol{\beta^{w,2}}$, and ability factors, denoted $\boldsymbol{\beta^{w,3}}$. Similarly, let $\boldsymbol{\beta^{s,1}}$ denote all parameters from the schooling equation (eq. 4.3), except for parameters associated with ability factors, denoted $\boldsymbol{\beta^{s,2}}$. Let $\boldsymbol{\beta^{g}}$ and $\boldsymbol{\beta^{o}}$ denote parameters from the utility flow of getting the GED (eq. 4.4) and the non-pecuniary utility flow (eq. 4.5), respectively.

[36]A prominent example is Keane and Wolpin (1997). See Keane and Wolpin (1994) for a detailed description of the interpolation method.

[37]For example, in Sullivan (2010), he only tracks human capital accumulated from the most recent occupation. This assumption helps reduce state space size, but forces each agent in the model to forfeit human capital accumulated in old occupations when he changes to a new occupation, which is a departure from empirical reality.

[38]The estimation calculations were performed on University of Virginia's high-performance computing cluster using a total of 1,379 cores. The estimation of this model is well suited for the type of parallelism utilized because it requires a large amount of numerical computation within each process, but very little communication among the processes. The real time efficiency gain is on the order of 0.9^*N with N cores.

[39]See Cunha et al. (2010), Heckman et al. (2013) and Heckman et al. (2018) for empirical applications.

[40]My model differs from the models discussed in Heckman et al. (2016) in that I explicitly model agents' preferences and the choice sets at each decision period, whereas Heckman et al. (2016) do not impose these assumptions. However, as shown in Heckman et al. (2016), the identification result of economically meaningful parameters, such as the average treatment effect of an additional year of schooling on wage return, does not depend on the functional form assumptions of preferences or explicit assumptions about the choice sets.

Lastly, let $\boldsymbol{\lambda}^{\mathbf{1}}$ denote parameters from the test score equations (eq. 4.1), except for those associated with ability factors, denoted $\boldsymbol{\lambda}^{\mathbf{2}}$; $\boldsymbol{\sigma_e}$ be the vector of dispersion parameters of the ability factors; $\boldsymbol{\sigma_{\epsilon^e}}$ be the vector of standard deviations of errors in the tests; and $\boldsymbol{\alpha}$ be the vectors of parameters governing the human capital accumulation process (section 4.2.2). As a reminder, σ_{ϵ^w} denotes the standard deviation of wage errors, and σ_{ϵ^u} denotes scale parameter of the extreme value error.

Since increase in human capital is modeled as discrete (see section 3.5.2), the rates at which observed wages of occupations with different task composition discretely jump identify $\boldsymbol{\alpha}$. Similar to linear models, co-variations between wages and observables, such as task indicators and education levels, identify $\boldsymbol{\beta^{w,1}}$. Co-variations between simulated human capital levels and wages identify $\boldsymbol{\beta^{w,2}}$. Identification of parameters of non-pecuniary benefit utility functions, including $\boldsymbol{\beta^{s,1}}$, $\boldsymbol{\beta^{g}}$ and $\boldsymbol{\beta^{o}}$, rests on the fact that the present value of discounted wage differentials do not fully explain career choices. Conditional on wages, the extent to which observables correlate with choices identifies these parameters. The size of variation in the residual wage terms identifies σ_{ϵ^w}. The extent to which wage drives choices identifies the relative importance of wage to non-pecuniary utility, which is equivalent to identifying the scale parameter, σ_{ϵ^u}.[41]

The co-variations bewteen test scores and individual's own and family characteristics identify $\boldsymbol{\lambda}^{\mathbf{1}}$. The remaining parameters include ability factor loadings in the log wage equation ($\boldsymbol{\beta^{w,3}}$), the schooling utility flow ($\boldsymbol{\beta^{s,2}}$) and the test score equations ($\boldsymbol{\lambda}^{\mathbf{2}}$), the dispersion parameters of abilities ($\boldsymbol{\sigma_e}$) and the standard deviations of test score errors ($\boldsymbol{\sigma_{\epsilon^e}}$). A generalized matching method and conditional independence assumptions provide identification of these parameters. See Carneiro et al. (2003) for a straightforward discussion.

It is worth noting that this identification approach differs significantly from the instrumental variable (IV) approach, employed frequently in non-structural studies. The IV estimator is mainly applied to static settings and seeks to minimize spurious correlation between latent factors and outcomes, whereas the approach taken in this paper considers a dynamic process where latent factors and a series of decisions jointly affect outcomes and seeks to quantify not only the effects of observables but also the effects of latent variables on outcomes. A drawback of the IV estimator is the difficulty to find suitable instruments with multi-dimensional latent factors that may have different effects on the outcomes. Moreover, Carneiro et al. (2011) and Heckman et al. (2016) show that, in dynamic discrete choice models, IV estimates are not economically interpretable or policy relevant, unless policy variables are instruments.

[41]See section 3.2 of Train (2009) for a detailed discussion on the identification and interpretation of the scale parameter in logit models.

5 Results

Tables 4 to 9 present estimates of all model parameters and the associated standard errors. Next, I discuss selected parameter estimates.

5.1 Return to Education

Table 4 presents all parameter estimates from the log wage equation (eq. 4.2). I start by discussing the estimates of the return to education, which are presented in the top panel.

The first column of the schooling panel presents the wage returns to an additional year of high school, an additional year of higher education and the return to GED for the baseline occupation that does not have any special task requirement, i.e., occupation (0,0,0,0).[42] The next four columns present the *changes* to these returns if the occupation requires high intensity along each of the four task dimensions. It is easier to interpret these parameters with concrete examples. Table 5 presents the return to education estimates for the two task-bundle-defined occupations that account for around 60% of the labor force: (1,1,0,0) and (0,0,1,1). For example, the return to an additional year of college education for occupation (1,1,0,0) is about 0.107 log wage points, which is calculated as the sum of the college wage return parameters from table 4 for the baseline occupation (-0.008) and the two associated tasks, cognitive (0.095) and interpersonal (0.020).

This paper's approach allows the return to education to vary not only by level but also based on the task content of occupations. Estimates confirm that returns to education varies significantly across tasks, and hence occupations.[43]

Much of the existing literature does not distinguish returns to education by occupations.[44] For the purpose of comparison, I compute the weighted averages of return to an additional year of high school and college based on the respondents' actual occupation choices at each age from 25 to 50. Results are presented in figure 3. The average return to an additional year of high school grows modestly and almost linearly from 2.55% at age 25 to 3.06% at age 50. In contrast, the average return to an additional year of college grows steadily

[42] As a reminder, higher education includes college and graduate education. I proceed to use college education henceforth with the understanding that it refers to higher education in general, including graduate education.

[43] Results show negative returns to certain levels of education for certain occupations. For example, as shown in table 5, an additional year of college education leads to a decrease of 0.037 log wage points for a typical factory worker (occupation (0,0,1,1)). Other studies have reported negative returns to education for manual tasks, using models that account for multi-dimensional abilities (Willis and Rosen, 1979; Yamaguchi, 2012; Prada and Urzúa, 2017).

[44] An exception is Sullivan (2010). He does not model occupations as bundles of tasks but includes five occupations classified by the first digit of the census occupation code. He also finds that the returns to education vary significantly across occupations.

from 2.80% at age 25 to 5.08% at age 40, and stabilizes at around 4.80% afterwards. The large change in return to college by age reflects the fact that, as workers with at least some college education age, they gradually move to cognitive- and interpersonal-task-intensive occupations. The average education return estimates from this paper are lower compared with most reduced-form estimates,[45] but in line with the relatively low average return to schooling estimated from earlier dynamic structural models (e.g., Belzil and Hansen 2002; Sullivan 2010) and very close to recent estimates from factor models that account for dynamic forward-looking behavior (Heckman et al., 2016).

Average returns to GED by age are also calculated and presented in figure 4. The average return to GED is modest for all ages examined, never exceeding the average return to two years of high school. This confirms previous findings that the GED is not equivalent to a high school diploma and brings much more modest wage returns (e.g., Cameron and Heckman 1993; Heckman and Rubinstein 2001).

5.2 Other Parameters from the Log Wage Equation

Aside from returns to education, the parameters from the log wage equation also quantify the returns to human capital accumulated on the job, latent abilities and the task content of occupations.

The estimates of return to human capital are presented in the second panel of table 4, and they measure the returns of accruing an additional unit of general human capital (column 1) and task-specific human capital (columns 2–5). For example, moving from the lowest (beginner) level to the second (intermediate) level increases the return to performing general tasks by 0.131 log wage points. The return to human capital varies widely across tasks. For cognitive, interpersonal, routine and manual tasks, increasing from the lowest human capital level to the highest brings a wage premium of 0.494, 0.221, 0.271 and 0.086 log wage points, respectively. The return to human capital for the routine tasks is very low, especially considering that this is a long panel tracking workers into their early 50s. Estimates of the probability of human capital transition are presented in the top panel in table 6. For each year, conditional on working, the probability of an increase in general human capital is 12.0%. Task-specific human capital increases with lower probability, ranging from 6.4% for cognitive tasks to 8.7% for routine tasks.

The third panel of table 4 presents returns to latent ability by task. Abilities are modeled as continuous, and each estimate corresponds to the log wage change resulting from a one unit increase in latent ability. Except for socio-emotional ability, the rest of the

[45]See Card (1999) for a review.

abilities are assumed to only affect the wage return to performing the corresponding tasks, and the estimates confirm that higher abilities bring positive wage returns to performing the corresponding tasks, which confirms multi-dimensional sorting based on the match between abilities and task content of occupations. Specifically, a one standard deviation increase in cognitive, socio-emotional, manual and routine ability brings a wage premium of 0.026, 0.015, 0.017 and 0.016 log wage points, respectively, conditional on the worker performing the corresponding tasks.[46] Higher socio-emotional ability also brings large positive wage premiums to performing cognitive and routine tasks, but incurs a wage penalty to performing manual tasks.

The bottom panel of table 4 presents estimates of returns to task content and pairwise task complementarity. Compared with the baseline occupation, which has no special task requirement, performing either interpersonal or manual tasks brings very large wage returns (0.856 and 0.939 log wage points, respectively). The returns to performing cognitive and routine tasks are positive but smaller in magnitude (0.023 and 0.260 log wage points, respectively). Moreover, the cognitive-interpersonal tasks pair and the manual-routine tasks pair are complementary, and performing either pair in one occupation brings an additional return of around 0.6 log wage points. Any other pair of tasks incurs a wage penalty. These estimates not only confirm the importance of task content in explaining observed wage variation, but also demonstrate the importance of pairwise task complementarity, which has largely been overlooked in the existing literature.[47] The result is intuitive given the observed pattern that occupations (1,1,0,0) and (0,0,1,1) account for 60% of the labor supply (see table 3).

5.3 Non-Pecuniary Utility Functions

Table 7 presents the parameter estimates for the non-pecuniary utility function from work. The constant and task indicators show that workers with entry levels of general and task-specific human capital view work as a "bad", net of wage income.[48] However, as workers accrue higher levels of human capital, they become more "attached" to work, drawing much higher levels of non-pecuniary utility.

[46] These estimates are calculated as the products of the estimated effects of a one unit increase in abilities (displayed in table 4) and the estimated standard deviations of the ability factors (displayed in the lower panel of table 9).

[47] The exception is Deming (2017), but he focuses on the complementarity between cognitive and interpersonal tasks.

[48] For example, the disutility for a worker in occupation (1,1,0,0) with entry levels of human capital is the sum of the disutility from work (-4.464) and the disutility from performing cognitive and interpersional tasks (-1.058 and -1.333, respectively), equaling -6.855 units. The scale parameter of the extreme value error (0.443) in table 6 is the conversion rate between the unit of utility and log wage points. Applying this conversion rate, the total disutility for this worker is equivalent to -3.037 log wage points.

Table 8 presents the parameter estimates for the non-pecuniary utility from schooling. Results confirm the fact that different abilities have different effects on educational attainment. While higher abilities to perform cognitive and routine tasks increase the utility from schooling, socio-emotional ability has no direct effect on schooling utility, and higher manual ability decreases schooling utility.

5.4 Test Score Functions

The top panel of table 9 presents the effects of observed characteristics and latent abilities on test scores. The effects of observed characteristics all have the expected signs. For examples, more years of education completed at the time of test boost tests scores across the board. The effects of ability factors on test scores are identified up to a normalization, so the effects of each ability on the corresponding test score is set to one. Increases in cognitive ability also result in higher test scores designed to measure the other three abilities. The bottom panel of table 9 presents the estimates of the size of signal and the size of noise from each test. In future work, I will perform a variance decomposition to quantify the contribution of observed characteristics, latent abilities and measurement errors to the observed variance in test scores.

6 Conclusion and Future Work

This paper builds and estimates a dynamic model of schooling, labor supply and occupational choices with multi-dimensional latent abilities (*cognitive*, *socio-emotional*, *manual* and *routine*). The model combines features from the dynamic career choice framework (Keane and Wolpin, 1997), factor models (e.g., Carneiro et al. 2003; Heckman et al. 2006, 2016) and the insight that occupations consist of bundles of tasks (Gibbons and Waldman, 2004; Autor and Handel, 2013). The key feature is that workers self-select into occupations based on not only level of education and human capital accumulated on the job, but also the match between their abilities and the task content of occupations.

The model is estimated with 35 years of observations on a sample of white men from NLSY79. Results provide thorough estimates on how latent abilities affect schooling decisions, and subsequently, how abilities, level of education and human capital accumulated on the job affect occupation choices and wages. While higher abilities to perform cognitive and routine tasks increase the utility from schooling, socio-emotional ability has no direct effect on schooling utility, and higher manual ability decreases schooling utility. The return to schooling varies by both level and task content of occupations. In particular, college education yields a large wage premium for cognitive tasks, a small wage premium for interpersonal tasks, but slightly negative return for manual and routine tasks. The average return to college education increases with age, reflecting the fact that workers with college education gradually sort into occupations with better-paying tasks (*cognitive* and *interpersonal*).

This paper is still a work in progress. Future work includes two parts. First, I will perform simulations to examine the fit of the model. In particular, I will compare simulated mean wage levels and choice proportions of schooling, labor supply and tasks to observed data by age. Second, using the structural estimates, I will conduct counterfactual simulations to evaluate various human capital development policies proposed to promote welfare (e.g., universal 4-year college subsidies). Simulation results quantify the income gain throughout workers' careers and how much of the income gain is due to increase in education leading workers to sort into occupations with better-paying tasks (e.g., *cognitive*). I will also quantify the net impact on welfare for each proposed policy.

References

Abbring, J. H. and J. J. Heckman, "Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation," *Handbook of econometrics* 6 (2007), 5145–5303.

Acemoglu, D. and D. Autor, "Skills, tasks and technologies: Implications for employment and earnings," in *Handbook of Labor Economics*volume 4 (Elsevier, 2011), 1043–1171.

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

Autor, D. H. and D. Dorn, "The growth of low-skill service jobs and the polarization of the US labor market," *American Economic Review* 103 (2013), 1553–97.

Autor, D. H., D. Dorn and G. H. Hanson, "The China syndrome: Local labor market effects of import competition in the United States," *American Economic Review* 103 (2013), 2121–68.

Autor, D. H. and M. J. Handel, "Putting tasks to the test: Human capital, job tasks, and wages," *Journal of Labor Economics* 31 (2013), S59–S96.

Autor, D. H., L. F. Katz and M. S. Kearney, "The polarization of the US labor market," *American Economic Review* 96 (2006), 189–194.

———, "Trends in US wage inequality: Revising the revisionists," *The Review of Economics and Statistics* 90 (2008), 300–323.

Autor, D. H., F. Levy and R. J. Murnane, "The skill content of recent technological change: An empirical exploration," *The Quarterly journal of economics* 118 (2003), 1279–1333.

Becker, G. S., "Human capital theory," *Columbia, New York* 1964 (1964).

Belzil, C. and J. Hansen, "Unobserved ability and the return to schooling," *Econometrica* 70 (2002), 2075–2091.

Ben-Porath, Y., "The production of human capital and the life cycle of earnings," *Journal of political economy* 75 (1967), 352–365.

Berkovec, J. and S. Stern, "Job exit behavior of older men," *Econometrica* (1991), 189–210.

Berndt, E. R., B. H. Hall, R. E. Hall and J. A. Hausman, "Estimation and inference in nonlinear structural models," *Annals of Economic and Social Measurement* 3 (1974), 4.

Björklund, A. and R. Moffitt, "The estimation of wage gains and welfare gains in self-selection models," *The Review of Economics and Statistics* (1987), 42–49.

Blau, D. M. and D. B. Gilleskie, "The role of retiree health insurance in the employment behavior of older men," *International Economic Review* 49 (2008), 475–514.

Blevins, J. R., "Nonparametric identification of dynamic decision processes with discrete and continuous choices," *Quantitative Economics* 5 (2014), 531–554.

Blinder, A. S. and Y. Weiss, "Human capital and labor supply: A synthesis," *Journal of Political Economy* 84 (1976), 449–472.

Börsch-Supan, A. and V. A. Hajivassiliou, "Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models," *Journal of Econometrics* 58 (1993), 347–368.

Brien, M. J., L. A. Lillard and S. Stern, "Cohabitation, marriage, and divorce in a model of match quality," *International Economic Review* 47 (2006), 451–494.

Cameron, S. V. and J. J. Heckman, "The nonequivalence of high school equivalents," *Journal of labor economics* 11 (1993), 1–47.

Card, D., "The causal effect of education on earnings," in *Handbook of labor economics*volume 3 (Elsevier, 1999), 1801–1863.

Carneiro, P., K. Hansen and J. Heckman, "Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice," *International Economic Review* 44 (2003), 361–422.

Carneiro, P., J. J. Heckman and E. J. Vytlacil, "Estimating marginal returns to education," *American Economic Review* 101 (2011), 2754–81.

Carneiro, P. and S. Lee, "Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality," *Journal of Econometrics* 149 (2009), 191–208.

Chan, M. K., "A dynamic model of welfare reform," *Econometrica* 81 (2013), 941–1001.

Cunha, F., J. J. Heckman and S. M. Schennach, "Estimating the technology of cognitive and noncognitive skill formation," *Econometrica* 78 (2010), 883–931.

Deming, D. J., "The growing importance of social skills in the labor market," *The Quarterly Journal of Economics* 132 (2017), 1593–1640.

French, E. and J. B. Jones, "The effects of health insurance and self-insurance on retirement behavior," *Econometrica* 79 (2011), 693–732.

Freyberger, J., "Non-parametric Panel Data Models with Interactive Fixed Effects," *The Review of Economic Studies* 85 (2018), 1824–1851.

Geweke, J., "Antithetic acceleration of Monte Carlo integration in Bayesian inference," *Journal of Econometrics* 38 (1988), 73–89.

———, "Bayesian inference in econometric models using Monte Carlo integration," *Econometrica* (1989), 1317–1339.

———, "Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities," (1991).

Gibbons, R. and M. Waldman, "Task-specific human capital," *American Economic Review* 94 (2004), 203–207.

Gould, E. D., "Rising wage inequality, comparative advantage, and the growing importance of general skills in the United States," *Journal of Labor Economics* 20 (2002), 105–147.

Gourinchas, P.-O. and J. A. Parker, "Consumption over the life cycle," *Econometrica* 70 (2002), 47–89.

Guvenen, F., B. Kuruscu, S. Tanaka and D. Wiczer, "Multidimensional skill mismatch," *American Economic Journal: Macroeconomics* (2018), under revision.

Hajivassiliou, V. A. and D. L. McFadden, "The method of simulated scores for the estimation of LDV models," *Econometrica* (1998), 863–896.

Hansen, K. T., J. J. Heckman and K. J. Mullen, "The effect of schooling and ability on achievement test scores," *Journal of Econometrics* 121 (2004), 39–98.

Heckman, J., R. Pinto and P. Savelyev, "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes," *American Economic Review* 103 (2013), 2052–86.

Heckman, J. J., "A life-cycle model of earnings, learning, and consumption," *Journal of political economy* 84 (1976), S9–S44.

Heckman, J. J., J. E. Humphries and G. Veramendi, "Dynamic treatment effects," *Journal of Econometrics* 191 (2016), 276–292.

———, "Returns to education: The causal effects of education on earnings, health, and smoking," *Journal of Political Economy* 126 (2018), S197–S246.

Heckman, J. J. and S. Navarro, "Dynamic discrete choice and dynamic treatment effects," *Journal of Econometrics* 136 (2007), 341–396.

Heckman, J. J. and Y. Rubinstein, "The importance of noncognitive skills: Lessons from the GED testing program," *American Economic Review* 91 (2001), 145–149.

Heckman, J. J. and G. Sedlacek, "Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market," *Journal of political Economy* 93 (1985), 1077–1125.

Heckman, J. J., J. Stixrud and S. Urzua, "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior," *Journal of Labor Economics* 24 (2006), 411–482.

Hu, Y. and S. M. Schennach, "Instrumental variable treatment of nonclassical measurement error models," *Econometrica* 76 (2008), 195–216.

Keane, M. P., "A computationally practical simulation estimator for panel data," *Econometrica* (1994), 95–116.

Keane, M. P., P. E. Todd and K. I. Wolpin, "The structural estimation of behavioral models: Discrete choice dynamic programming methods and applications," in *Handbook of labor economics*volume 4 (Elsevier, 2011), 331–461.

Keane, M. P. and K. I. Wolpin, "The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte Carlo evidence," *The Review of Economics and Statistics* (1994), 648–672.

———, "The career decisions of young men," *Journal of Political Economy* 105 (1997), 473–522.

Lin, D., R. Lutter and C. J. Ruhm, "Cognitive performance and labour market outcomes," *Labour Economics* 51 (2018), 121–135.

Lindqvist, E. and R. Vestman, “The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment,” *American Economic Journal: Applied Economics* 3 (2011), 101–28.

Magnac, T. and D. Thesmar, “Identifying dynamic discrete decision processes,” *Econometrica* 70 (2002), 801–816.

Marianne, B., “New perspectives on gender,” in *Handbook of Labor Economics*volume 4 (Elsevier, 2011), 1543–1590.

McKenzie, L. W., “Turnpike theory,” *Econometrica* (1976), 841–865.

Mincer, J., “Investment in human capital and personal income distribution,” *Journal of Political Economy* 66 (1958), 281–302.

———, “Schooling, experience, and earnings,” *NBER Books* (1974).

Murnane, R. J., J. B. Willett, M. J. Braatz and Y. Duhaldeborde, “Do different dimensions of male high school students’ skills predict labor market success a decade later? Evidence from the NLSY,” *Economics of Education Review* 20 (2001), 311–320.

Neal, D. A. and W. R. Johnson, “The role of premarket factors in black-white wage differences,” *Journal of Political Economy* 104 (1996), 869–895.

Neumann, J. v., “A model of general economic equilibrium,” *Review of Economic Studies* 13 (1945), 1–9.

Nybom, M., “The distribution of lifetime earnings returns to college,” *Journal of Labor Economics* 35 (2017), 903–952.

Pavan, R., “Career choice and wage growth,” *Journal of Labor Economics* 29 (2011), 549–587.

Poletaev, M. and C. Robinson, “Human capital specificity: evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys, 1984–2000,” *Journal of Labor Economics* 26 (2008), 387–420.

Prada, M. F. and S. Urzúa, “One size does not fit all: Multiple dimensions of ability, college attendance, and earnings,” *Journal of Labor Economics* 35 (2017), 953–991.

Robinson, C., “Occupational mobility, occupation distance, and specific human capital,” *Journal of Human Resources* 53 (2018), 513–551.

Roy, A. D., “Some thoughts on the distribution of earnings,” *Oxford economic papers* 3 (1951), 135–146.

Rust, J. and C. Phelan, “How social security and medicare affect retirement behavior in a world of incomplete markets,” *Econometrica* (1997), 781–831.

Speer, J. D., “Pre-market skills, occupational choice, and career progression,” *Journal of Human Resources* 52 (2017), 187–246.

Stern, S., “Simulation-based estimation,” *Journal of Economic Literature* 35 (1997), 2006–2039.

Sullivan, P., “A dynamic analysis of educational attainment, occupational choices, and job search,” *International Economic Review* 51 (2010), 289–317.

Train, K. E., *Discrete choice methods with simulation* (Cambridge university press, 2009).

Willis, R. J. and S. Rosen, “Education and self-selection,” *Journal of political Economy* 87 (1979), S7–S36.

Yamaguchi, S., “Tasks and heterogeneous human capital,” *Journal of Labor Economics* 30 (2012), 1–53.

———, “Changes in returns to task-specific skills and gender wage gap,” *Journal of Human Resources* 53 (2018), 32–70.

Yano, M., “The turnpike of dynamic general equilibrium paths and its insensitivity to initial conditions,” *Journal of Mathematical Economics* 13 (1984), 235–254.

Tables and Figures

Table 1: Schooling and Labor Supply Decisions by Age

Age	Schooling	Work	Unemployed	Missing	N
16	88.8%	1.5%	3.5%	6.3%	1379
17	71.1%	10.8%	10.7%	7.3%	1379
18	41.0%	25.7%	13.4%	19.9%	1379
19	35.5%	36.7%	13.4%	14.4%	1379
20	31.1%	46.7%	13.5%	8.7%	1379
21	22.6%	60.0%	16.5%	0.9%	1377
22	13.8%	69.8%	16.0%	0.3%	1373
23	9.1%	76.9%	13.8%	0.2%	1371
24	6.9%	82.1%	10.7%	0.4%	1371
25	4.8%	85.1%	10.1%	0.1%	1363
⋮	⋮	⋮	⋮	⋮	⋮
30	2.3%	91.1%	6.3%	0.2%	1327
⋮	⋮	⋮	⋮	⋮	⋮
40	0.4%	89.8%	7.4%	2.4%	1208
⋮	⋮	⋮	⋮	⋮	⋮
50	0.0%	85.5%	11.5%	3.0%	627

Note: This table displays age-specific choice percentages for the final sample of white men from the NLSY79. The last column shows the total number of observations by age.

Table 2: Correlations Between Constructed Composite Test Scores

	Cognitive	Socio-emotional	Manual	Routine
Cognitive	1.00			
Socio-emotional	0.35	1.00		
Manual	0.68	0.26	1.00	
Routine	0.71	0.31	0.47	1.00

Note: This table shows the correlations between constructed composite test scores for the 1,379 NLSY79 male respondents included in the final sample.

Table 3: Distribution of Occupations from the 2000 Census and the NLSY79 Sample

Occupation	Total Labor Supply in 2000		Final NLSY79 Sample	
	N	Percentage	*N*	Percentage
(1,1,1,1)	7,650,573	5.13%	1,112	3.24%
(1,1,1,0)	5,053,429	3.39%	1,291	3.76%
(1,1,0,1)	3,630,031	2.44%	1,092	3.18%
(1,0,1,1)	5,377,633	3.61%	1,594	4.64%
(0,1,1,1)	2,120,852	1.42%	521	1.52%
(1,1,0,0)	45,083,292	30.25%	10,495	30.54%
(1,0,1,0)	1,370,186	0.92%	564	1.64%
(1,0,0,1)	5,052,946	3.39%	649	1.89%
(0,1,1,0)	3,707,461	2.49%	508	1.48%
(0,1,0,1)	2,968,260	1.99%	325	0.95%
(0,0,1,1)	40,949,372	27.48%	12,110	35.24%
(1,0,0,0)	1,086,375	0.73%	286	0.83%
(0,1,0,0)	3,789,357	2.54%	1,330	3.87%
(0,0,1,0)	7,290,161	4.89%	1,795	5.22%
(0,0,0,1)	6,742,985	4.52%	137	0.40%
(0,0,0,0)	7,156,915	4.80%	556	1.62%
Total	149,029,826	100.00%	34,365	100.00%

Note: This table shows the distribution of occupations from the 2000 census and the NLSY79 final sample using occupations denoted as task bundles. For each occupation, the task components are arranged in the order of *cognitive*, *interpersonal*, *manual* and *routine*, with 1 denoting high task intensity and 0 denoting low task intensity.

Table 4: Estimated Coefficients from the Log Wage Equation

	General	Cognitive	Interpersonal	Manual	Routine
Schooling					
High School	0.227**	-0.048**	-0.102**	-0.154**	-0.075**
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
College	-0.008**	0.095**	0.020**	-0.021**	-0.008**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
GED	0.216**	-0.033**	-0.036**	-0.124**	-0.086**
	(0.013)	(0.012)	(0.012)	(0.011)	(0.012)
Human Capital					
Level 1	baseline	baseline	baseline	baseline	baseline
	(-)	(-)	(-)	(-)	(-)
Level 2	0.315**	0.439**	0.131**	0.110**	0.076**
	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)
Level 3	0.446**	0.494**	0.221**	0.271**	0.086**
	(0.004)	(0.007)	(0.006)	(0.005)	(0.006)
Ability Factors					
Cognitive		0.041**			
		(0.003)			
Socio-emotional		0.144**	0.085**	-0.042**	0.091**
		(0.024)	(0.016)	(0.014)	(0.017)
Manual				0.143**	
				(0.040)	
Routine					0.117**
					(0.025)
Constant					
	0.538**				
	(0.008)				
Task Indicators					
		0.023*	0.856**	0.939**	0.260**
		(0.010)	(0.009)	(0.008)	(0.008)
Task Complementarity					
Cognitive					
Interpersonal		0.601**			
		(0.004)			
Manual		-0.467**	-0.703**		
		(0.005)	(0.005)		
Routine		-0.048**	-0.626**	0.593**	
		(0.004)	(0.005)	(0.004)	

Note: Standard errors are in parentheses below estimates. ** p<0.01, * p<0.05.

Table 5: Estimated Returns to Education for Selected Occupations

Occupation	(1,1,0,0)	(0,0,1,1)
Example	Typical Managers	Typical Factory Workers
High school	0.076**	-0.003
	(0.001)	(0.008)
College	0.107**	-0.037**
	(0.001)	(0.002)
GED	0.147**	0.006
	(0.011)	(0.011)

Note: Returns to high school and college are annual. Standard errors are in parentheses below estimates. ** p<0.01, * p<0.05.

Table 6: Human Capital Transition Probability and Dispersion Parameters

Human Captial Transition	Parameters	Implied Transition Probability
General (α_0)	-1.997** (0.003)	12.0%
Cognitive (α_1)	-2.681** (0.002)	6.4%
Interpersonal (α_2)	-2.385** (0.002)	8.4%
Manual (α_3)	-2.517** (0.002)	7.5%
Routine (α_4)	-2.349** (0.003)	8.7%
Dispersion Parameters of Time-Varying Errors		
standard deviation of wage error (σ_{ϵ^w})		0.521** (0.001)
scale parameter of extreme value error (σ_{ϵ^u})		0.443** (0.002)

Note: Standard errors are in parentheses below estimates. ** p<0.01, * p<0.05.

Table 7: Parameter Estimates for the Non-Pecuniary Utility Function from Work

	General	Cognitive	Interpersonal	Manual	Routine
Constant					
	-4.464** (0.007)				
Task Indicators					
		-1.058** (0.010)	-1.333** (0.004)	-0.532** (0.010)	-0.632** (0.009)
Human Capital					
Level 1	baseline (-)	baseline (-)	baseline (-)	baseline (-)	baseline (-)
Level 2	4.190** (0.027)	1.525** (0.024)	1.817** (0.018)	0.180** (0.015)	0.425** (0.016)
Level 3	1.707** (0.012)	1.597** (0.017)	1.833** (0.013)	1.571** (0.014)	1.473** (0.013)

Note: Standard errors are in parentheses below estimates. ** p<0.01, * p<0.05.

Table 8: Parameter Estimates for the Non-Pecuniary Utility Function from Schooling

	Utility Flow
Schooling Utility	
Constant	1.412**
	(0.022)
Age	-0.342**
	(0.004)
(Age-24)*$\mathbb{1}$(Age$\geq$25)	0.341**
	(0.005)
Cognitive	0.365**
	(0.017)
Socio-emotional	0.069
	(0.058)
Manual	-0.356**
	(0.128)
Routine	0.549**
	(0.130)
GED	
Constant	-1.983**
	(0.086)
$\mathbb{1}$(Age$\geq$25)	-0.561**
	(0.109)

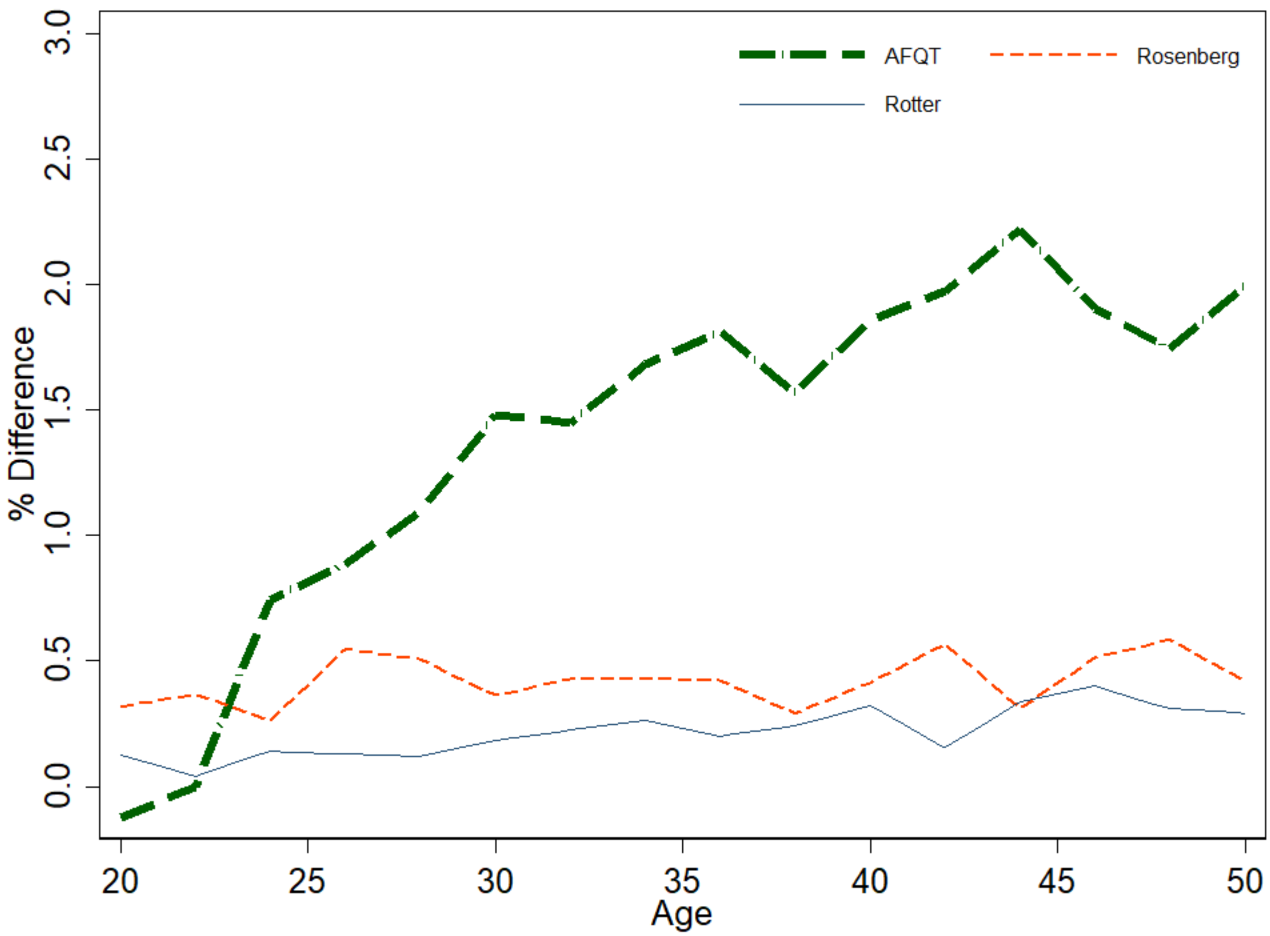
Note: Standard errors are in parentheses below estimates. ** p<0.01, * p<0.05.

Table 9: Parameter Estimates for Test Score Functions

	Cognitive	Socio-emotional	Manual	Routine
Individual and Family Background				
Constant	-2.266**	-1.752**	-1.296**	-2.491**
	(0.251)	(0.220)	(0.142)	(0.123)
Years of Education at Time of Test	0.149**	0.115**	0.156**	0.161**
	(0.016)	(0.015)	(0.010)	(0.008)
Whether GED at Time of Test	0.452*	0.297	0.722**	0.259**
	(0.205)	(0.160)	(0.125)	(0.098)
Mom's Years of Education	0.043**	0.030*	0.034**	0.038**
	(0.015)	(0.015)	(0.009)	(0.007)
Dad's Years of Education	0.037**	0.016	0.007	0.020**
	(0.010)	(0.010)	(0.006)	(0.005)
Whether Urban at 14	-0.061	0.040	-0.186**	0.050
	(0.061)	(0.063)	(0.042)	(0.032)
Whether South at 14	-0.114	-0.064	-0.231	-0.077
	(0.064)	(0.059)	(0.041)	(0.033)
Number of Siblings	-0.037	-0.015	-0.040**	-0.025**
	(0.019)	(0.018)	(0.012)	(0.010)
Ability Factors				
Cognitive	1	0.289**	0.732**	0.746**
	(-)	(0.028)	(0.023)	(0.018)
Socio-emotional		1		
		(-)		
Manual			1	
			(-)	
Routine				1
				(-)
Estimated Std. Dev. of Ability Factors				
	0.639**	0.177**	0.116**	0.133**
	(0.013)	(0.025)	(0.029)	(0.025)
Estimated Std. Dev. of Errors in Tests				
	0.362**	0.856**	0.660**	0.643**
	(0.022)	(0.018)	(0.013)	(0.008)

Note: Standard errors are in parentheses below estimates. ** $p<0.01$, * $p<0.05$.

Figure 1: Wage Gains Associated with Increases in AFQT, Rosenberg and Rotter Scores, by Age

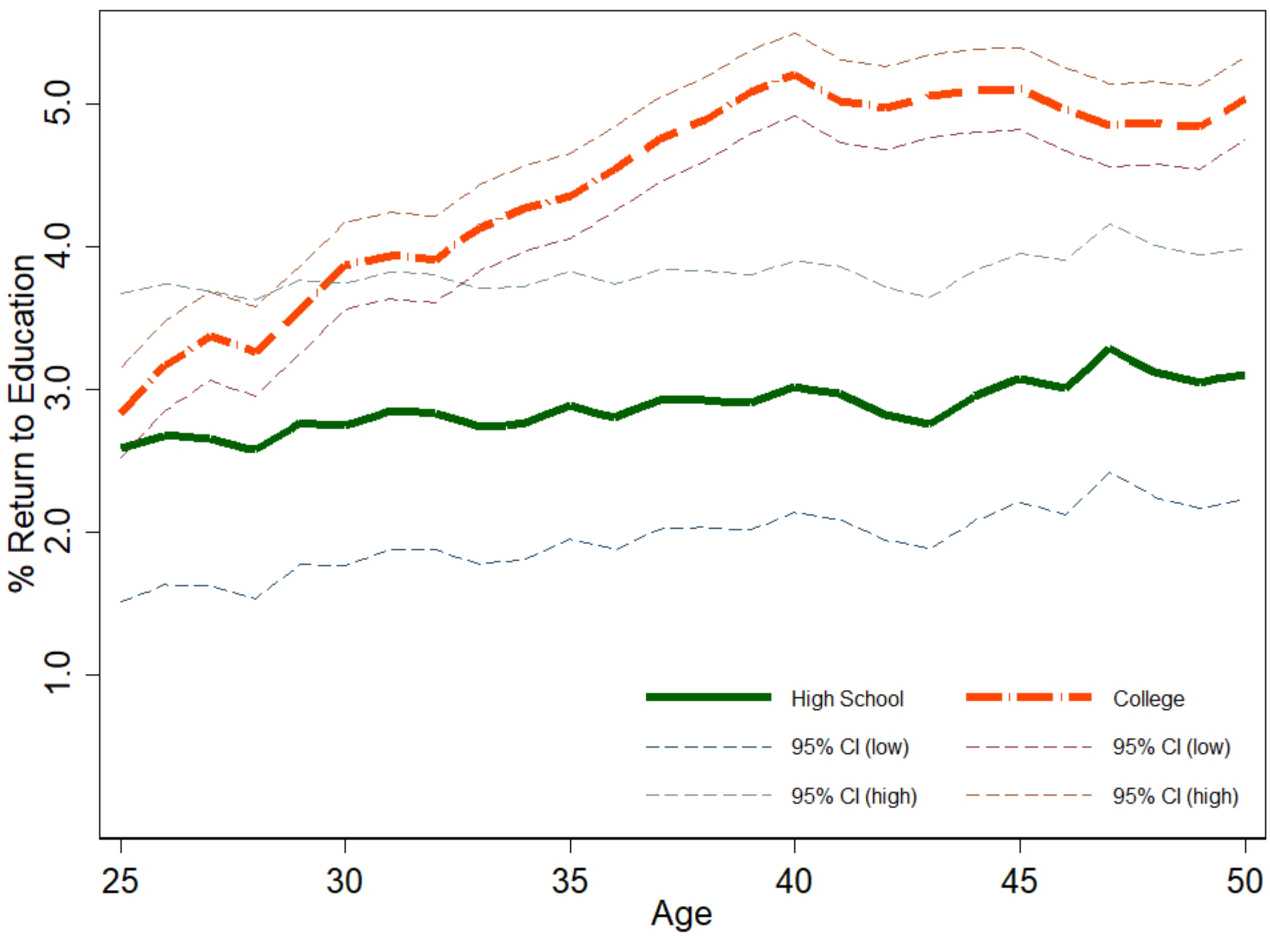


Note: This figure shows gains in wages associated with a 0.1 standard deviation increase in AFQT, the Rosenberg self-esteem and the (reverse) Rotter Locus of Control scores, respectively, for men from the representative sample of NLSY79. The horizontal axis shows two-year age groupings beginning with the age listed (e.g., 30 refers to 30-31 year-olds). The vertical axis shows percentage difference in wage. The 2006 renormed version of AFQT percentile scores are converted to standardized z-scores assuming normality. Information on construction of age-adjusted AFQT percentile scores is available at: https://www.nlsinfo.org/content/cohorts/nlsy79/topical-guide/education/aptitude-achievement-intelligence-scores. Rosenberg and Rotter scores are also age-adjusted and converted to standardized z-scores. Estimates are obtained from OLS regressions with log(wage) as the outcome variable, and adjusted z-scores of AFQT, Rosenberg, and Rotter scores as main explanatory variables. Other explanatory variables include: self-reported sociability (extremely shy to extremely outgoing); race/ethnicity (white, black and Hispanic); magazines, library cards and newspapers in home at age 14 (3 variables); urban residence at 14; whether respondent and respondents' parents are foreign born (3 variables); mother and father's education (below high school, high school graduate, some college and college graduate); number of siblings (0, 1, 2, 3, 4, $\geq$5), older siblings (0, 1, 2, 3, $\geq$4); and survey year dummy variables. Two-year age groupings are used because NLSY79 changed from annual surveys to biennial surveys in 1994, and wage data are available only biennially for the later ages. The number of observations is between 1,655 (age group 20) and 2,344 (age group 26). The difference in number of observations by age is due to a) some older respondents starting the survey at ages 22 and 23 with wages unreported for earlier ages, and b) attrition over time. See table A1 for estimates at selected ages.

Figure 2: Wage Growth Between Age 20 and Age 50

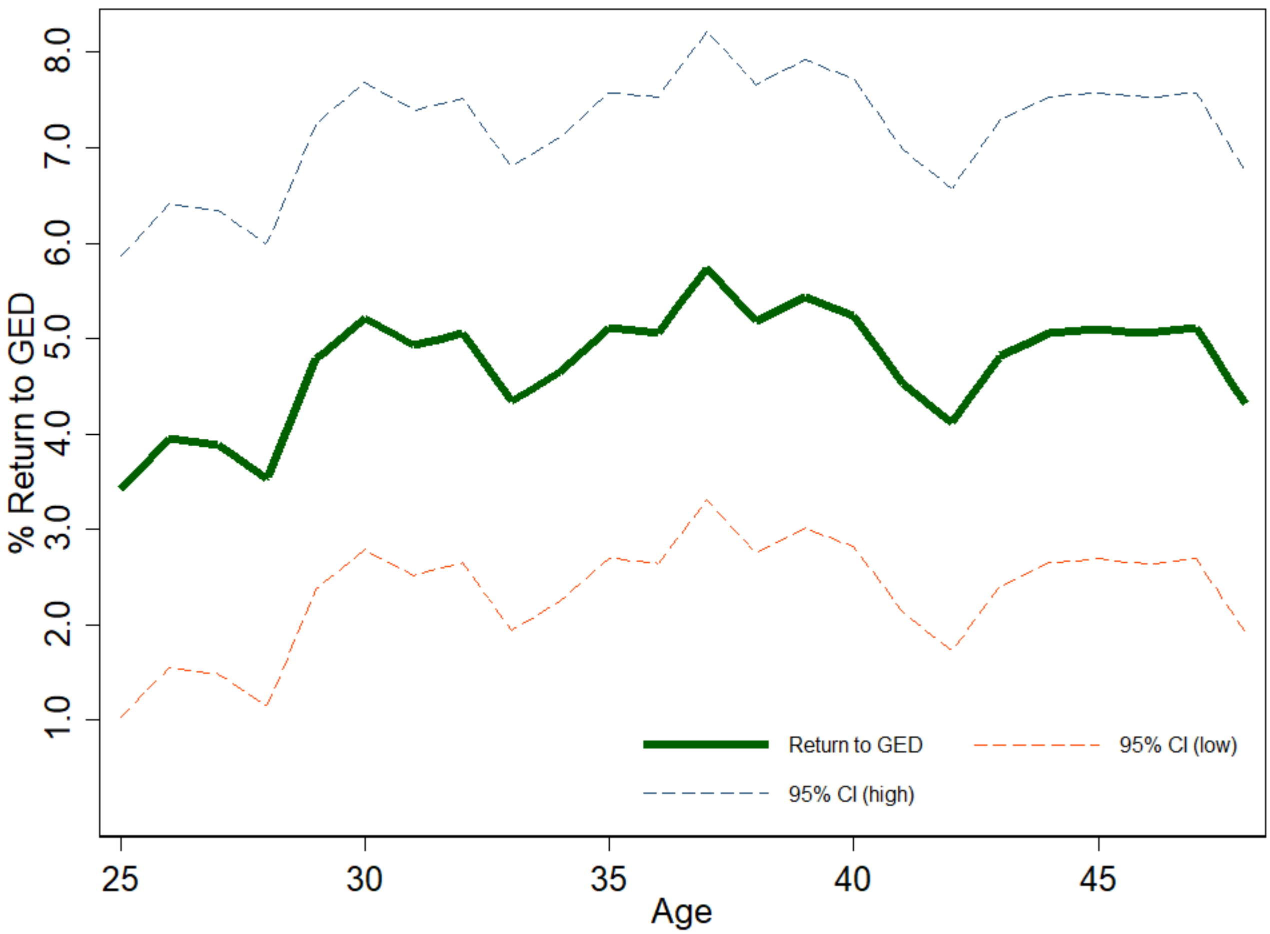
Note: This figure plots the mean wage levels for workers in the full final sample and by task intensity. The "High Cognitive" line plots the mean wage levels for occupations with high cognitive task intensity, rated using the composite task intensity scores generated from O*NET descriptors, as described in section 2.3. The "High Interpersonal," "High Manual" and "High Routine" lines have similar interpretations. See table A2 for summary statistics at selected ages.

Figure 3: Average Return to an Additional Year of High School or College, Ages 25-50



Note: This figure presents the weighted averages of return to an additional year of high school and college, respectively, for ages 25 to 50. For each age, estimates are calculated using parameter estimates on the return to schooling from table 4 and applying actual occupation choice percentages by respondents in the sample as weights. Standard errors are calculated using the delta method. See table A4 for estimates at selected ages.

Figure 4: Average Return to the GED, Ages 25-50



Note: This figure presents the weighted averages of return to have received the GED for ages 25 to 50. For each age, estimates are calculated using parameter estimates on the return to GED from table 4 and applying actual occupation choice percentages by respondents in the sample as weights. Average returns are not calculated for ages 49 and 50, because fewer than 50 GED recipients remain in the sample (n=43 at age 49; n=27 at age 50). Standard errors are calculated using the delta method. See table A4 for estimates at selected ages.

Appendices

A Supplemental Tables and Figures

Table A1: Wage Gains Associated with Increases in Test Scores at Selected Ages

Age	AFQT	Rosenberg	Rotter	N
24-25	0.743**	0.263*	0.141	2,333
	(0.128)	(0.112)	(0.114)	
30-31	1.479**	0.362**	0.184	2,285
	(0.133)	(0.119)	(0.106)	
34-35	1.681**	0.428**	0.263*	2,226
	(0.147)	(0.122)	(0.113)	
40-41	1.854**	0.414**	0.320*	2,032
	(0.161)	(0.150)	(0.140)	
44-45	2.219**	0.310*	0.338*	1,883
	(0.171)	(0.156)	(0.148)	
50-51	1.994**	0.422*	0.292	1,745
	(0.206)	(0.171)	(0.168)	

Note: Standard errors are in parentheses below estimates. ** $p<0.01$, * $p<0.05$. Estimates are percentages in wage gains associated with a 0.1 standard deviation increase in the three test scores. This is a supplemental table to figure 1. See note to figure 1 for more details.

Table A2: Descriptive Statistics for Wage Levels at Selected Ages

Age	Cognitive	Interpersonal	Manual	Routine	Full Sample	N
25	\$22.89 (\$26.27)	\$24.81 (\$50.38)	\$18.93 (\$12.47)	\$19.65 (\$17.71)	\$21.37 (\$34.06)	1,363
30	\$29.20 (\$39.57)	\$27.98 (\$29.93)	\$22.69 (\$47.16)	\$23.64 (\$49.11)	\$25.26 (\$41.01)	1,327
35	\$37.36 (\$50.73)	\$37.62 (\$51.86)	\$29.40 (\$71.17)	\$27.12 (\$54.30)	\$32.05 (\$53.70)	1,271
40	\$39.73 (\$34.44)	\$40.04 (\$35.16)	\$23.58 (\$17.46)	\$23.79 (\$13.47)	\$32.10 (\$28.79)	1,208
45	\$42.78 (\$32.12)	\$41.55 (\$31.63)	\$25.72 (\$23.46)	\$25.02 (\$17.46)	\$33.87 (\$27.64)	1,145
50	\$41.79 (\$31.11)	\$42.01 (\$31.78)	\$26.06 (\$21.50)	\$25.55 (\$18.85)	\$33.34 (\$26.94)	627

Note: This is a supplemental table to figure 2. See note to figure 2.

Table A3: Top- and Bottom-Ranked Occupations by Task Intensity

	Cognitive	Interpersonal	Manual	Routine
Top Five	Physicists and astronomers	Legislators	Truck, delivery and tractor drivers	Slicing and cutting machine operators
	Actuaries	Clergy and religious workers	Bus drivers	Programmers of numerically controlled machine tools
	Operations and systems researchers and analysts	Sales engineers	Millwrights	Shoemaking machine operators
	Chemical engineers	Managers in education and related fields	Excavating and loading machine operators	Printing machine operators
	Statistical clerks	Athletes, sports instructors and officials	Airplane pilots and navigators	Crane, derrick, winch and hoist operators
Bottom Five	Graders and sorters of agricultural products	Graders and sorters of agricultural products	Operations and systems researchers and analysts	Door-to-door sales, street sales and news vendors
	Housekeepers, maids, butlers, stewards and lodging quarters cleaners	Pressing machine operators (clothing)	Actuaries	Guides
	Packers and packagers by hand	Shoe repairers	Financial managers	Occupational therapists
	Janitors	Mail and paper handlers	Lawyers	Child care workers
	Pressing machine operators (clothing)	Crossing guards and bridge tenders	Legislators	Clergy and religious workers

Note: For the curious readers of this paper, among the 326 occupations rated, economists rank 8th for cognitive skill intensity, 70th for interpersonal skill intensity, 318th for manual skill intensity and 291th for routine skill intensity.

Table A4: Average Return to Education at Selected Ages

Age	High School	College	GED
25	2.552**	2.795**	3.360**
	(0.538)	(0.157)	(1.192)
30	2.709**	3.793**	5.077**
	(0.492)	(0.150)	(1.187)
35	2.844**	4.264**	4.984**
	(0.465)	(0.146)	(1.185)
40	2.971**	5.075**	5.109**
	(0.437)	(0.141)	(1.186)
45	3.029**	4.980**	4.979**
	(0.431)	(0.140)	(1.185)
50	3.056**	4.917**	-
	(0.433)	(0.140)	-

Note: This table supplements figure 3 and figure 4 to present the average percentage effects of an additional year of high school, college, and having received the GED, respectively, on wages. See note to figure 3 for more details on calculating the average return to high school and college, and note to figure 4 for more details on calculating the average return to GED.

Figure A1: Sample Questions from Numerical Operation and Coding Speed Tests

Numerical Operations

The Numerical Operations subtest consists of simple mathematical computations. This is a speed test, so work as fast as you can to determine your answers without making mistakes.

DOD Example:

3 x 4=

A. I

B. 7

C. 12

D. 14

(C is the correct answer.)

Coding Speed

The Coding Speed subtest contains questions that test how quickly and accurately you can find a number in a table. At the top of each section is a number table or "key." The key is a group of words with a code number for each word. Each item in the test is a word taken from the key at the top of that page. From among the possible answers listed for each item, find the one that is the correct code number for that word. This is a speed test, so work as fast as you can to determine your answers without making mistakes.

DOD Example:

Key

green	2715	man	3451	salt	4586
hat	1413	room	2864	tree	5972

Questions	**Answers**				
	A	B	C	D	E
1. room	1413	2715	2864	3451	4586
2. green	2715	2864	3451	4586	5972
3. tree	1413	2715	3451	4586	5972
4. hat	1413	2715	3451	4586	5972
5. room	1413	2864	3451	4586	5972

(The correct answers are 1C, 2A, 3E, 4A, 513.)

B O*NET Descriptors

This appendix lists O*NET descriptors used to construct task intensity measures by occupation. The selection of these descriptors follows Acemoglu and Autor (2011) and Deming (2017).

Cognitive

4.A.2.a.4 Analyzing data/information

4.A.2.b.2 Thinking creatively

4.A.4.a.1 Interpreting information for others

1.A.1.b.4 Deductive reasoning

1.A.1.b.5 Inductive reasoning

1.A.1.c.1 Mathematical reasoning

1.A.1.c.2 Number facility

Interpersonal

4.A.4.a.4 Establishing and maintaining personal relationships

4.A.4.b.4 Guiding, directing and motivating subordinates

4.A.4.b.5 Coaching/developing others

2.B.1.a Social perceptiveness

2.B.1.b Coordination

2.B.1.c Persuasion

2.B.1.d Negotiation

Manual

4.A.3.a.4 Operating vehicles, mechanized devices, or equipment

4.C.2.d.1.g Spend time using hands to handle, control or feel objects, tools or controls

1.A.2.a.2 Manual dexterity

1.A.1.f.1 Spatial orientation

Routine

4.C.3.b.7 Importance of repeating the same tasks

4.C.3.b.4 Importance of being exact or accurate

4.C.3.b.8 Structured v. unstructured work (reverse)

4.C.3.d.3 Pace determined by speed of equipment

4.A.3.a.3 Controlling machines and processes

4.C.2.d.1.i Spend time making repetitive motions

C Data Aggregation and Sample Selection

In this appendix, I describe how I aggregate individual's decisions up to the annual level. The method of data aggregation largely follows Keane and Wolpin (1997) and Sullivan (2010).

In the model, the decision period is a school year, from September to August. Enrollment status is determined by a monthly school attendance array and degree attainment by the end of August each year. If an individual reports to be enrolled, but does not advance a grade by the end of the school year, following Keane and Wolpin (1997) and Sullivan (2010), I code the individual as unemployed for that year. For more than half of the sample, respondents are left-censored, because they started the survey after the age of 16. In this case, I assume these respondents to be in the tenth grade at age 16, and they continue to go to school in consecutive years until they reach their reported grade in the first wave that they join the survey. If there still are years in between that are unaccounted for, I code these annual decisions as missing.

Annual employment status is determined using the job history arrays. For any given year, the NLSY79 keeps track of up to five jobs each individual has ever held and records weekly employment status for those jobs. For each employment spell, respondents also report the occupation at the 3-digit census code level. For the purpose of aggregation, I only consider jobs with more than 20 hours worked each week. I then tally the employment status by week. The activity that an individual engages in for the majority of the weeks during the year, including unemployment, is assigned to be his choice for that year.

Respondents are dropped if their schooling history is inconsistent or if they ever served in the military. Note that it is common for respondents to miss certain waves of interview and reappear in later waves. I retain all observations for each respondent, including missing waves, until either the latest survey wave available or the wave the respondent appears in the survey for the last time. This approach preserves the maximum amount of information available from the data. Simulation techniques, introduced in section 4.5.1 in the main text, are employed to overcome this missing data problem in estimation.