2020

2020-5 Returns to Skill and the Evolution of Skills for Older Men

Qian Liu

Lance Lochner

Youngmin Park

Youngki Shin

Returns to Skill and the Evolution of Skills for Older Men

by

Qian Liu, Lance Lochner, Youngmin Park, and Youngki Shin

Working Paper #2020-5 **August 2020**

Centre for Human Capital and Productivity (CHCP)

Working Paper Series

Department of Economics
Social Science Centre
Western University
London, Ontario, N6A 5C2
Canada

Returns to Skill and the Evolution of Skills for Older Men

Qian Liu Lance Lochner Youngmin Park Youngki Shin

August 19, 2020

Abstract

This paper shows that repeated cross-section data with multiple skill measures (one continuous and repeated) available each period are sufficient to nonparametrically identify the evolution of skill returns and cross-sectional skill distributions. With panel data and the same available measurements, the dynamics of skills can also be identified. Our identification strategy motivates a multi-step nonparametric estimation strategy. We further show that if any continuous repeated measurement is shown to be linear in skills, a much simpler GMM estimator can be used.

Using HRS data on men ages 52+ from 1996-2016, we show that one of the available (continuous and repeated) skill measures is linear in skills and implement our GMM estimation approach. Our estimates suggest that the returns to skill were fairly stable from the mid-1990s to the Great Recession, rising thereafter. We document considerable differences in skills and lifecycle skill profiles over ages 52–70 across cohorts, with more recent cohorts possessing lower skills in their mid-50s but experiencing much weaker skill declines with age. We also document skill differences by education and race, which are stable across ages and explain roughly one-third and one-half, respectively, of the corresponding differences in wages. We observe substantial differences in skills for men in their mid-50s choosing to retire at different ages, but no clear evidence of sharp declines in skills surrounding retirement ages. Finally, we show that individual fixed effects account for more than a third of all skill variation at age 60, with considerable persistence in year-to-year skill innovations.

1 Introduction

Despite decades of research on the topic, there remains considerable interest in better understanding growing inequality in the U.S. and many other developed countries. An important focus of much of this research has been on the extent to which growing wage inequality is the result of rising returns to skill (often attributed to skill-biased technological change) vs. growth in the variance of skills across workers. Due to a lack of direct measures of skill, researchers are typically forced to make (strong) assumptions about the evolution of skill distributions or returns, with little external validation or evidence regarding those assumptions.

This paper establishes nonparametric identification of the returns to skills and cross-sectional distribution of skills over time given the availability of repeated cross-section data on wages and at least two skill measurements every period, with at least one continuous skill measure repeated each period. With longitudinal data, we show that it is also possible to identify the dynamics of skills (i.e. the distribution of skills in period t conditional on skills in period $t-1$). Our constructive identification strategy suggests a multi-stage estimation approach, which simplifies considerably if one of the repeated measurements is known to be linear in skills, something that is straightforward to verify. We use these methods and longitudinal data from the Health and Retirement Study (HRS) to estimate the evolution of skill returns and distributions, as well as the dynamics of skills, for men in the U.S. from 1996–2016.

Several distinct literatures in economics aim to distinguish interpersonal differences in skills from the market-level returns to those skills. For example, the primary objective of many empirical studies on discrimination is to determine the extent to which race or gender differences in wages, as well as the evolution of those gaps over time, can be explained by group differences in skill levels.[1] Similarly, researchers often attempt to decompose differences in the wage returns to schooling across countries (Leuven et al., 2004; Hanushek and Zhang, 2009) or over time (Heckman et al., 1998; Bowlus and Robinson, 2012) into differences in worker skill levels (deriving from, e.g., heterogeneous school quality or home

[1] There are vast literatures on race and gender wage differentials surveyed in Altonji et al. (2012). Among the most closely related studies on race, see Card and Lemieux (1996); Neal and Johnson (1996), and Chay and Lee (2000). See Blau and Kahn (1997, 2017) for closely related studies on gender wage gaps.

environments) and in the wage returns to those skills. Related research has framed the rapid rise in residual wage inequality (i.e., inequality within narrowly defined demographic groups) over the past several decades as a combination of changes in the distribution of unmeasured skills and in their labor market returns (e.g., Juhn et al., 1989; Katz and Murphy, 1992; Lemieux, 2006; Autor et al., 2008; Lochner et al., 2020).

Skill measurement is a critical challenge in all of these literatures. In most cases (e.g. studies using Census data or data from the Current Population Surveys, CPS), only a crude proxy or correlate of skill is available (e.g., educational attainment, per-pupil spending when young, age), especially when researchers are interested in studying inequality across long time periods. In these cases, skills are often equated with available measures like education or labor market experience. Other studies explicitly aim to estimate the role of unmeasured skills. To this end, Juhn et al. (1989) assume that the distribution of these skills remained constant over the period they study, attributing all growth in the variance of log wage residuals to an increase in the return to unobserved skill. Lemieux (2006) instead assumes that the variance of skills within narrowly defined observable groups (e.g., within age-education-race categories) remained unchanged over time, allowing for changes in the distribution of skills through changes in the composition of the workforce (by age, education, and race). He finds that a sizeable fraction of the growth in residual inequality can be traced to changes in the distribution of skills caused by the aging and growing education of the population. Using longitudinal data on wages from the Panel Study of Income Dynamics (PSID), Lochner et al. (2020) relax the assumption of invariant within-group distributions, assuming instead that variation in skill growth among older workers is idiosyncratic. Their estimates suggest declining returns to unobserved skill over the late 1980s and 1990s, while the growth in residual wage inequality is instead explained by growth in the variance of skills (due to growing variation in skill growth). Clearly, assumptions about the evolution of skill distributions have important implications for the conclusions one draws about driving forces underlying rising wage inequality.[2]

[2] In related research, Card and Lemieux (1996) and Chay and Lee (2000) use CPS data to study the extent to which changes in skill gaps and the returns to skill can explain the evolution of black – white wage differentials. Card and Lemieux (1996) consider a single skill model (composed of both observed and unobserved components) with restrictions on the evolution of skills over time, while Chay and Lee (2000) consider a model with differently priced observed and unobserved skills, placing restrictions on the distribution of unobserved skills within observable groups over time.

In some cases, researchers have used more specialized data sets like the National Longitudinal Surveys of Youth (NLSY), which contain cognitive test scores as direct measures of skill. Using the 1979 Cohort of the NLSY, Neal and Johnson (1996) demonstrate that differences in adolescent cognitive achievement (as measured by the Armed Forces Qualifying Test, AFQT) can, by themselves, explain the sizeable differences in wages between young black and white men in the U.S.[3] Several studies have also used the NLSY cohorts to study the evolution of inequality since the 1980s. For example, Herrnstein and Murray (1994) argue that the U.S. has become more meritocratic based on sharply increasing wage returns to AFQT. Others have used AFQT measures in an effort to disentangle whether the growing differences in earnings by educational attainment reflect rising returns to schooling or rising returns to cognitive ability (Heckman and Vytlacil, 2001; Taber, 2001; Castex and Dechter, 2014).[4] Deming (2017) exploits other non-cognitive measures in the NLSY, estimating that the returns to social skills have risen since the 1980s.

It is noteworthy that commonly used data sources with direct skill measures do not typically contain the same measures over time for the same individuals. For example, cohorts of the NLSY contain AFQT scores measured only once, during adolescence. Thus, studies using the NLSY estimate the effects of adolescent cognitive achievement, rather than contemporaneous skills, on wages later in life. Given the practical challenge of sorting out age and time effects from only a few birth-year cohorts, studies following individuals over time from one of the NLSY cohorts cannot determine whether growing wage inequality is driven by differential lifecycle growth in skills by AFQT or rising returns to skill (Heckman and Vytlacil, 2001).

Grogger and Eide (1995) and Murnane et al. (1995) address this issue by exploiting data on cognitive achievement and wages from two separate cohorts (National Longitudinal Study of the High School Class of 1972, NELS72, and High School and Beyond, HSB). Comparing the earnings of individuals at the same ages (roughly age 24), their estimates suggest that both the returns to schooling and cognitive skill rose between 1978 and 1986. However,

[3] To study differential returns to cognitive achievement across countries at a point in time, Leuven et al. (2004) and Hanushek and Zhang (2009) use international data from the International Adult Literacy Survey (IALS) while Hanushek et al. (2015) exploits data from an expanded set of countries from the Programme for International Assessment of Adult Competencies (PIAAC).

[4] In related work, Altonji et al. (2012) study changes in the distribution of skills (overall and by race and gender) for two NLSY cohorts using AFQT scores, education, and other individual and family characteristics.

a limitation of both studies is that the cognitive tests (taken during the last year of high school), while similar, were not the same across the two surveys. Thus, any differences in their mapping to true cognitive skills would be reflected in the estimated returns to skill over time. Castex and Dechter (2014) improve upon these studies by comparing the wages of men over ages 18-28 from the 1979 and 1997 Cohorts of the NLSY (born roughly 20 years apart), who took the same AFQT test during adolescence. Their estimates suggest that the returns to adolescent cognitive achievement declined while the returns to schooling increased between the 1980s and 2000s.[5] They also find that log wage differences by AFQT were quite similar upon labor market entry across the two cohorts, with log wage gaps by AFQT increasing with experience for the 1979 Cohort but not the 1997 Cohort. Thus, the difference in estimated returns to AFQT across cohorts only appeared after individuals had spent several years in the labor market.

While this cross-cohort approach to estimating changes in the returns to skills over time requires weaker assumptions than studies following a single cohort over time, it is not assumption-free. Because most respondents took the AFQT during adolescence, one can only interpret changes in wage returns to cognitive skills at older ages as changes in the returns to skill over time if the evolution of skills between the age of the test and the age at which wages are compared was the same across cohorts.[6] This would be violated if, for example, variation in early lifecycle skill growth changed. Indeed, the finding that log wage gaps by AFQT evolved differently with work experience for the two NLSY cohorts suggests that early skill growth may have differed between them. Alternatively, the returns

[5]In related work, Deming (2017) uses the 1979 and 1997 Cohorts of the NLSY to estimate changes in the wage returns to social skills over time. Because the same measures of social skills are not available for both NLSY cohorts, Deming (2017) works with normalized measures of social skills (measured during adolescence), effectively assuming identical distributions (and measurement quality) across the cohorts. Thus, the rising returns to social skills he estimates could also reflect greater variation in social skills (or more precise measurements) for the more recent cohort. Edin et al. (2017) exploit administrative data in Sweden that contain consistent cognitive and non-cognitive measures collected for men entering the military at ages 18-19 for cohorts born between 1951 and 1975. They estimate that the return to non-cognitive skill roughly doubled from 1992 to 2013, while the return to cognitive skill rose in the 1990s but fell in the 2000s.

[6]The same caveat applies to studies by Grogger and Eide (1995) and Murnane et al. (1995), who use data from NELS72 and HSB, and to Edin et al. (2017), who use administrative data from Sweden. Additionally, the fact that NLSY respondents took the AFQT test at different ages across the cohorts requires adjustments based on implicit assumptions regarding rank stability of cognitive scores across testing ages. Of course, test scores measured at age 22 for the oldest in the 1979 Cohort are likely to be much more strongly related to cognitive skills at ages 18-28 than are test scores measured at age 12 for the youngest in the 1997 Cohort.

to cognitive skill may have evolved differently in the 1980s vs. 2000s, which cannot be easily distinguished from differential lifecycle skill growth patterns for the two cohorts.

Despite any limitations, these cohort comparison studies provide some of the most convincing evidence on changes in the returns to cognitive skills over time. Yet, they only offer, at best, a few snapshots for the U.S.: Grogger and Eide (1995) and Murnane et al. (1995) find that the returns to cognitive ability rose between 1978 and 1986, while Castex and Dechter (2014) estimate that the returns declined between the 1980s and 2000s. These studies do not tell us anything about the 1960s and early 1970s or the most recent decade. Nor do they inform us as to when the sizeable decline in estimated returns occurred between the 1980s and 2000s. Yet, these are periods of considerable debate in the literature on log wage residual inequality and the evolution of returns to skill (Juhn et al., 1989; Katz and Murphy, 1992; Lemieux, 2006; Autor et al., 2008; Lochner et al., 2020).

Although we cannot comment on earlier time periods, we use biennial data from 1996-2016 HRS to estimate the evolution of returns to skill and skill distributions over this more recent period. Unlike previous studies, our data contain measures of wages and scores from the same cognitive tests repeated every other year. As a result, our approach requires no assumptions on skill distributions nor their dynamics. Instead, our key identification assumption is that the measurement function for at least one continuous repeated test (taken at ages 52+) is identical over time (i.e., the mapping between skills and expected individual test scores is time invariant). We further show that our use of panel data enables us to identify the lifecycle dynamics of skills under very general conditions.

Our estimates suggest that the returns to (cognitive) skill were relatively stable over the late 1990s and early 2000s but rose significantly after the Great Recession. While these estimates are noisy, they are roughly consistent with the patterns estimated by Lochner et al. (2020) using data from the PSID.

Our results also intersect with a broader literature studying cognitive skills late in life; although, this literature typically estimates latent skills (or factors) derived only from cognitive tests without linking (or anchoring) those skills to wages. By incorporating wages in our analysis, we are able to describe how skills, measured in log wage units, evolve for individuals ages 52-70. We document considerable variation in skills among individuals in their early- to mid-50s across cohorts (with much higher skills among those from earlier birth cohorts), but these differences are largely dissipated by the time individuals reach their early 60s as earlier cohorts experienced much faster declines in skill than did later cohorts. Indeed, there is little evidence of skill depreciation prior to age 63 (when last observed) among the cohort born 1954-1959.

Consistent with several studies using the HRS and similar data in other countries, we show that late-life cognitive performance differs significantly across education groups (Cagney and Lauderdale, 2002; Mazzonna and Peracchi, 2012) and racial groups (Zsembik and Peek, 2001; Karlamangla et al., 2009; Castora-Binkley et al., 2015). Our estimates suggest that, among older workers, cognitive skill differences are quite similar across ages and explain roughly one-third of the education gaps and nearly half of the race gaps in log wages.[7]

Prior research has also shown that retirement is negatively correlated with cognition (Adam et al., 2007); although, self-selection into retirement has posed challenges in estimating the causal relationship between retirement and cognition.[8] We make no effort to attribute causality; instead, we simply show lifecycle skill profiles from ages 52-70 for workers choosing to retire at different ages. Our results suggest that age 55 skill levels are increasing in retirement age, with those retiring at age 65 or older possessing roughly 15% more skills than those retiring prior to age 55. We see no evidence of sharp declines in cognitive skills surrounding retirement ages, with skills relatively constant through age 60 for those retiring prior to age 55 and skills declining almost linearly from ages 52 to 70 for those retiring over ages 55-64. To the extent that retirement does lead to cognitive decline, our results suggest that its impacts are relatively small or largely offset by other lifecycle forces.

Finally, we show that skills are quite persistent, with individual fixed effects accounting for more than a third of all skill variation at age 60. Year-to-year skill innovations are also persistent with an autocorrelation of 0.93.

This paper proceeds as follows. Section 2 describes our model of skill dynamics, the

[7]See Card (1999) and Heckman et al. (2006) for comprehensive surveys of wage differences by education. Neal and Johnson (1996) show that adolescent skill gaps can explain much of the early-career differences in wages by race.

[8]Several studies use exogenous policy variation, such as eligible retirement ages and pension policies, as instruments to identify the causal effects, producing mixed findings. Some of these studies estimate significant negative effects of retirement on cognition (Rohwedder and Willis, 2010; Bonsang et al., 2012; Mazzonna and Peracchi, 2012, 2017), while others find no causal effect (Coe and Zamarro, 2011; Coe et al., 2012). Still others estimate heterogeneous effects across different occupations (Mazzonna and Peracchi, 2017) or across individuals retiring early vs. at the statutory age (Celidoni et al., 2017).

relationship between skills and wages, and other skill measurement functions. We discuss identification and estimation of skill returns, measurement functions, skill distributions, and the dynamics of skills. Section 3 describes the HRS data we use in estimation, while Section 4 presents our estimation results. We offer concluding thoughts in Section 5.

2 Methodology

In this section, we provide a general model of the evolution of skills and log wages. Since we focus on older workers (ages 50+), we assume skills evolve exogenously, reflecting growth and/or depreciation. With panel data on both log wages and test-based measures of skills, we describe identification and estimation of the distribution of skills and their dynamics, the evolution of skill return functions over time, and the mapping between skills and their test-based measurements.

2.1 Model

Let $\ln W_{i,a,t}$ be the log wage for individual i at age a in year t and $T_{i,j,a,t}$ reflect skill test score measure $j = 1, \ldots, J$. We consider the following specification for log wages and skill measures:

$$\begin{aligned} \ln W_{i,a,t} &= \gamma_t + \lambda_t \theta_{i,a,t} + \varepsilon_{i,a,t}, \\ T_{i,1,a,t} &= \tau_1(\theta_{i,a,t}) + \eta_{i,1,a,t}, \\ T_{i,j,a,t} &= G_j(\tau_j(\theta_{i,a,t}) + \eta_{i,j,a,t}), \text{ for } j = 2, \ldots, J, \end{aligned} \tag{1}$$

where $\theta_{i,a,t}$ denotes unobserved skill, λ_t denotes "return" to skill in period t, $\varepsilon_{i,a,t}$ and $\eta_{i,j,a,t}$ are idiosyncratic non-skill shocks to wages and test measurement errors, respectively, and $\tau_j(\cdot)$ is a strictly increasing, age- and time-invariant measurement function that maps unobserved skill to cognitive measure j combined with a weakly increasing function $G_j(\cdot)$ for $j = 2, \ldots, J$. Notice that the model allows ordered discrete measures for $j = 2, \ldots, J$ if $T_{i,1,a,t}$ is continuous. Observations are i.i.d. over individual i for any (j, a, t). For any individual i, we assume that $(\theta_{i,a,t}, \varepsilon_{i,a,t}, \eta_{i,j,a,t})$ are mutually independent for all test measures j and that each of these variables is also independent of past and future realizations of the other two variables. The idiosyncratic measurement error $\eta_{i,j,a,t}$ is independent over j and t. Since a and t move together for each individual, $\eta_{i,j,a,t}$ is also independent over a, but it need not be identically distributed over ages or time. We normalize $\lambda_{t^*} = 1$ for some year t^*, which effectively measures skills in year t^* log wage units. We also normalize $E(\varepsilon_{i,a,t}) = E(\eta_{i,j,a,t}) = 0$ for all (j, a, t). Identification requires no assumptions about the serial dependence structure for log wage shocks $\varepsilon_{i,a,t}$.

Let $\alpha_{a,t} \equiv E(\theta_{i,a,t})$ and $\bar{\theta}_{i,a,t} := \theta_{i,a,t} - \alpha_{a,t}$ be the de-meaned skill value. Then, we can rewrite log wages as follows:

$$\begin{aligned}
\ln W_{i,a,t} &= \gamma_t + \lambda_t(\theta_{i,a,t} - \alpha_{a,t} + \alpha_{a,t}) + \varepsilon_{i,a,t} \\
&= \gamma_t + \lambda_t \alpha_{a,t} + \lambda_t \bar{\theta}_{i,a,t} + \varepsilon_{i,a,t} \\
&\equiv \widetilde{\gamma}_{a,t} + \lambda_t \bar{\theta}_{i,a,t} + \varepsilon_{i,a,t},
\end{aligned}$$

where $\widetilde{\gamma}_{a,t} := \gamma_t + \lambda_t \alpha_{a,t}$. Regressing log wages on interactions of age and time dummies yields consistent estimates of $\tilde{\gamma}_{a,t}$ and log wage residuals $w_{i,a,t} := \lambda_t \bar{\theta}_{i,a,t} + \varepsilon_{i,a,t}$. We work with these residuals below to discuss identification and estimation of returns to skill, λ_t, and the evolution of skill distributions.

2.2 Identification

Since the continuous measurement function $\tau_1(\cdot)$ is assumed to be time invariant, observing this same measurement along with at least one other measurement and log wage residuals in multiple periods, we can identify the returns to skill each period, nonparametric (age/cohort-specific) distributions of skills each period, and the dynamics of skills. (We also obtain nonparametric identification for the measurement function $\tau_1(\cdot)$ and the corresponding error distributions $f_{\eta_{1,a,t}}$ for all t.) The identification of skill dynamics is considered both under a general Markov structure and under the AR(1) structure with a fixed effect term. Notice that the latter requires identification of each component separately while the former focuses only on the conditional density of $\theta_{i,a,t}$ given $\theta_{i,a-1,t-1}$. We are careful to note which features of our model can be identified with repeated cross-section data alone and which require panel data.

2.2.1 Returns to Skill, Cross-Sectional Skill Distributions, and Measurements

We begin by discussing identification of returns to skill, cross-sectional skill distributions, test measurement functions, and the distribution of test score measurement errors.

Consider a normalized age and time pair (a^*, t^*), where $\lambda_{t^*} = 1$ and $\alpha_{a^*,t^*} = 0$. Let $c^* = t^* - a^*$ be the corresponding cohort. Applying Theorem 1 in Hu and Schennach (2008), we can identify the distributions $F_{\theta_{a^*,t^*}}(\cdot)$, $F_{\varepsilon_{a^*,t^*}}(\cdot)$, and $F_{\eta_{1,a^*,t^*}}(\cdot)$, as well as the measurement function $\tau_1(\cdot)$ from the joint density of $(w_{i,a^*,t^*}, T_{i,1,a^*,t^*}, ..., T_{i,J,a^*,t^*})$ for $J \geq 2$.[9] Appendix A provides details of the regularity conditions and the identification result.

Now, we consider an arbitrary age and time pair $(a, t) \neq (a^*, t^*)$, rewriting the model as

$$\begin{aligned} w_{i,a,t} &= \tilde{\theta}_{i,a,t} + \varepsilon_{i,a,t}, \\ T_{i,1,a,t} &= \tilde{\tau}_{1,a,t}(\tilde{\theta}_{i,a,t}) + \eta_{i,1,a,t} \\ T_{i,j,a,t} &= G_j\left(\tilde{\tau}_{j,a,t}(\tilde{\theta}_{i,a,t}) + \eta_{i,j,a,t}\right) \text{ for } j = 2, \ldots, J, \end{aligned} \tag{2}$$

where $\tilde{\theta}_{i,a,t} := \lambda_t \bar{\theta}_{i,a,t} = \lambda_t(\theta_{i,a,t} - \alpha_{a,t})$ and $\tilde{\tau}_{j,a,t}(x) := \tau_j(x/\lambda_t + \alpha_{a,t})$. Notice that $E(\tilde{\theta}_{i,a,t}) = 0$. Using the same arguments as above, we can identify $F_{\tilde{\theta}_{i,a,t}}(\cdot)$, $F_{\varepsilon_{i,a,t}}(\cdot)$, $F_{\eta_{i,1,a,t}}(\cdot)$, and $\tilde{\tau}_{1,a,t}(\cdot)$ from the joint density of $(w_{i,a,t}, T_{i,1,a,t}, ..., T_{i,J,a,t})$.

Knowledge of $\tau_1(\cdot)$ from (a^*, t^*) and $\tilde{\tau}_{1,a,t}(\cdot)$ from any other (a, t) identifies λ_t and $\alpha_{a,t}$.[10] To see this, consider two points θ_1 and θ_2 on the support of θ such that $\theta_1 < \theta_2$. Since $\tau_1(\cdot)$ is strictly increasing, $\theta_1 < \theta_2$ implies that $\tilde{\tau}_{1,a,t}(\theta_1) < \tilde{\tau}_{1,a,t}(\theta_2)$. By definition, $\tilde{\tau}_{1,a,t}(\theta_1) = \tau_1(\theta_1/\lambda_t + \alpha_{a,t})$ and $\tilde{\tau}_{1,a,t}(\theta_2) = \tau_1(\theta_2/\lambda_t + \alpha_{a,t})$. Solving this system of equations identifies

$$\begin{aligned} \lambda_t &= \frac{\theta_1 - \theta_2}{\tau_1^{-1}\left(\tilde{\tau}_{1,a,t}(\theta_1)\right) - \tau_1^{-1}\left(\tilde{\tau}_{1,a,t}(\theta_2)\right)} \\ \alpha_{a,t} &= \frac{\theta_2 \tau_1^{-1}\left(\tilde{\tau}_{1,a,t}(\theta_1)\right) - \theta_1 \tau_1^{-1}\left(\tilde{\tau}_{1,a,t}(\theta_2)\right)}{\theta_2 - \theta_1}. \end{aligned}$$

Having identified λ_t, $\alpha_{a,t}$, and $F_{\tilde{\theta}_{i,a,t}}(\cdot)$, we can then identify $F_{\theta_{i,a,t}}(\theta) = F_{\tilde{\theta}_{i,a,t}}(\lambda_t(\theta - \alpha_{a,t}))$.

We emphasize that none of the identification results thus far require panel data. Identification of the returns to skills and cross-sectional distributions of skills over time can be achieved with repeated cross-section data. While we have explicitly considered the case with $J \geq 2$ repeated measures each period, it is clear that only a single continuous measurement must be repeated every period. Other measurements can differ from period to period.[11]

[9] We make no effort to separately identify $\tau_j(\cdot)$ from $G_j(\cdot)$ for $j = 2, ..., J$, which may require additional assumptions for discrete measures and is not necessary given a single repeated continuous measure $T_{i,1,a,t}$.

[10] Note that identification of λ_t and $\alpha_{a,t}$ further identifies time effects in log wage equations, γ_t, from $\tilde{\gamma}_{a,t}$.

[11] Even more generally, as long as each period of data contains at least two independent measures, identification can be achieved with overlapping periods that contain a repeated continuous measurement (e.g., one continuous measurement over periods 1 and 2 with a different continuous measurement over periods 2 and 3, etc.). The same continuous measurement need not be available over the entire sample period.

2.2.2 Skill Processes: General Approach

Without any additional assumptions on $\theta_{i,a,t}$ and $\varepsilon_{i,a,t}$, we can identify their serial dependence structure with panel data on our continuous measure $T_{i,1,a,t}$ (given what we have already identified in Section 2.2.1). To see this, consider two time periods at (a^*, t^*) and (a, t). For example, $a = a^* + 1$ and $t = t^* + 1$. Let $\varphi_{\eta_{i,1,a^*,t^*},\eta_{i,1,a,t}}(x_1, x_2) := E[\exp(-i(x_1\eta_{i,1,a^*,t^*} + x_2\eta_{i,1,a,t}))]$ be a characteristic function of $(\eta_{i,1,a^*,t^*}, \eta_{i,1,a,t})$ and define characteristic functions for other random variables similarly. Note that

$$\begin{aligned}
\varphi_{T_{i,1,a^*,t^*},T_{i,1,a,t}}(x_1, x_2) &= \varphi_{\tau_1(\theta_{i,a^*,t^*}),\tau_1(\theta_{i,a,t})}(x_1, x_2) \cdot \varphi_{\eta_{i,1,a^*,t^*},\eta_{i,1,a,t}}(x_1, x_2) \\
\varphi_{\tau_1(\theta_{i,a^*,t^*}),\tau_1(\theta_{i,a,t})}(x_1, x_2) &= \frac{\varphi_{T_{i,1,a^*,t^*},T_{i,1,a,t}}(x_1, x_2)}{\varphi_{\eta_{i,1,a^*,t^*},\eta_{i,1,a,t}}(x_1, x_2)} \\
&= \frac{\varphi_{T_{i,1,a^*,t^*},T_{i,1,a,t}}(x_1, x_2)}{\varphi_{\eta_{i,1,a^*,t^*}}(x_1)\varphi_{\eta_{i,1,a,t}}(x_2)},
\end{aligned}$$

where the last equality holds by the time-independence of $\eta_{i,1,a,t}$ distributions. Since we know all distributions on the right hand side, we can identify the joint distribution of $(\tau_1(\theta_{i,a^*,t^*}), \tau_1(\theta_{i,a,t}))$. Then, the joint distribution of $(\theta_{i,a^*,t^*}, \theta_{i,a,t})$ is identified by

$$\begin{aligned}
F_{\theta_{i,a^*,t^*},\theta_{i,a,t}}(x_1, x_2) &= P(\theta_{i,a^*,t^*} \leq x_1, \theta_{i,a,t} \leq x_2) \\
&= P(\tau_1(\theta_{i,a^*,t^*}) \leq \tau_1(x_1), \tau_1(\theta_{i,a,t}) \leq \tau_1(x_2)),
\end{aligned}$$

where the final expression is identified since we know both $\tau_1(\cdot)$ and $\varphi_{\tau_1(\theta_{i,a^*,t^*}),\tau_1(\theta_{i,a,t})}(\cdot, \cdot)$. Therefore, we can construct the conditional density $f_{\theta_{i,a,t}|\theta_{i,a^*,t^*}}$ from the joint distribution and identify the serial dependence of $\theta_{i,a,t}$. Similarly, we can identify the joint distribution of $(\varepsilon_{i,a^*,t^*}, \varepsilon_{i,a,t})$ by noting that

$$\begin{aligned}
\varphi_{\varepsilon_{i,a^*,t^*},\varepsilon_{i,a,t}}(x_1, x_2) &= \frac{\varphi_{w_{i,a^*,t^*},w_{i,a,t}}(x_1, x_2)}{\varphi_{\theta_{i,a^*,t^*},\widetilde{\theta}_{i,a,t}}(x_1, x_2)} \\
&= \frac{\varphi_{w_{i,a^*,t^*},w_{i,a,t}}(x_1, x_2)}{\varphi_{\theta_{i,a^*,t^*},\theta_{i,a,t}}(x_1, \lambda_t x_2) \times \exp(-i\alpha_{a,t}\lambda_t x_2)},
\end{aligned}$$

where we already know $\alpha_{a,t}$, λ_t, and the two joint distributions on the right hand side. The serial dependence of $(\varepsilon_{i,a,^*,t^*}, \varepsilon_{i,a,t})$ follows immediately.

2.2.3 Skill Process: AR(1) and Fixed Effect

We now consider the identification problem for the skill process when skills are decomposed into the following three components: (i) a systematic lifecycle skill growth component, which can differ freely across cohorts, $\alpha_{a,t}$; (ii) an individual fixed effect ψ_i; and (iii) an AR(1) component $\phi_{i,a,t}$. Thus, the skill process can be written as follows:

$$\begin{aligned} \theta_{i,a,t} &= \alpha_{a,t} + \psi_i + \phi_{i,a,t}, \\ \phi_{i,a,t} &= \rho\phi_{i,a-1,t-1} + \nu_{i,a,t}, \end{aligned} \tag{3}$$

where $\nu_{i,a,t}$ is independent over t. We further assume that ψ_i is independent of $\phi_{i,a,t}$ and $\nu_{i,a,t}$ for all (a,t). We normalize $E(\psi_i) = E(\nu_{i,a,t}) = E(\phi_{i,a,t}) = 0$ for all (a,t), which implies $\alpha_{a,t} = E(\theta_{i,a,t})$ as before. We normalize $\alpha_{a^*,t^*} = 0$ for some (a^*, t^*).[12]

Notice that we have already identified the returns to skills and age/cohort- and time-specific skill distributions using only repeated cross-sections of log wages and skill measures (see Section 2.2.1). To identify the basic components in equation (3), we need panel data from (at least) three periods, t, $t+1$, and $t+2$

First, we can use similar arguments as in Section 2.2.2 to identify the joint distribution of $(\theta_{i,a,t}, \theta_{i,a+1,t+1}, \theta_{i,a+2,t+2})$. Then, we can construct the following moment conditions from equation (3):

$$\begin{aligned} Var(\theta_{i,a,t}) &= Var(\psi_i) + Var(\phi_{i,a,t}) \\ Cov(\theta_{i,a,t}, \theta_{i,a+1,t+1}) &= Var(\psi_i) + \rho Var(\phi_{i,a,t}) \\ Cov(\theta_{i,a,t}, \theta_{i,a+2,t+2}) &= Var(\psi_i) + \rho^2 Var(\phi_{i,a,t}). \end{aligned}$$

Solving this system of equations, we can identify ρ as

$$\rho = 1 - \frac{Var(\theta_{i,a,t}) - Cov(\theta_{i,a,t}, \theta_{i,a+2,t+2})}{Var(\theta_{i,a,t}) - Cov(\theta_{i,a,t}, \theta_{i,a+1,t+1})},$$

where moments on the right hand side are identified from the joint density of $(\theta_{i,a,t}, \theta_{i,a+1,t+1}, \theta_{i,a+2,t+2})$.

Next, we can identify cohort- and time-specific distributions for skill shocks and cohort-specific distributions for the fixed effects. We fix a cohort and let t_0 be the first year it is observed. First, since we already know $\alpha_{a,t}$ and the joint distribution $F_{\theta_{i,a+1,t+1},\theta_{i,a,t}}$ for all (a,t), we identify the left hand side of the following two equations:

$$\begin{aligned} \theta_{i,a,t_0} - \alpha_{a,t_0} &= \phi_{i,a,t_0} + \psi_i \\ \frac{(\theta_{i,a+1,t_0+1} - \alpha_{a+1,t_0+1}) - (\theta_{i,a,t_0} - \alpha_{a,t_0})}{\rho - 1} &= \phi_{i,a,t_0} + \frac{\nu_{i,a+1,t_0+1}}{\rho - 1}. \end{aligned}$$

[12]It is not necessary to use the same t^* here as used for normalizing $\lambda_{t^*} = 1$; however, we do so in this section to simplify the exposition.

Since ϕ_{i,a,t_0}, ψ_i, and $\nu_{i,a+1,t_0+1}$ are mutually independent, we can identify their distributions by applying Kotlarski's Lemma (Kotlarski, 1967). Second, we identify the distribution of $\phi_{i,a,t}$ sequentially for all $t \geq t_0+1$ from $\phi_{i,a+1,t+1} - \phi_{i,a,t} = (\theta_{i,a+1,t+1} - \alpha_{a+1,t+1}) - (\theta_{i,a,t} - \alpha_{a,t})$ by applying standard deconvolution arguments, since we know the distribution of the right hand side and that of ϕ_{i,a,t_0}. Finally, we identify the distribution of $\nu_{i,a,t}$ for $t \geq t_0 + 1$ from $\phi_{i,a,t} = \rho\phi_{i,a-1,t-1} + \nu_{i,a,t}$ by applying the deconvolution arguments again.

2.3 Estimation

We can estimate the full model nonparametrically, e.g. the sieve maximum likelihood estimator as in Hu and Schennach (2008). However, it is quite challenging in practice as the objective function involves multiple integration over many unobservables.

We mitigate this computational difficulty by developing a three-step estimation procedure based on the identification strategy outlined earlier. First, one can estimate the measurement functions $\tau_j(\cdot)$ and the distribution of skills θ_{i,a,t^*} from the cross-sectional observations of $\{w_{i,a,t^*}, T_{i,1,a,t^*}, \ldots, T_{i,J,a,t^*}\}$ for all ages a at time t^*. Second, repeated cross-sectional observations $\{w_{i,a,t}, T_{i,1,a,t}, \ldots, T_{i,J,a,t}\}$ at $(a,t) \neq (a^*, t^*)$ can be used along with the estimated $\widehat{\tau}_1(\cdot)$ (from Step 1) to estimate the skill return λ_t and skill distribution of $\theta_{i,a,t}$ for all (a,t). Finally, panel data (if available), can be used to estimate the dynamics of skill distributions.

The estimation of skill dynamics (and other features of the model) becomes much simpler when any continous measurement function $\tau_j(\cdot)$ estimated in Step 1 turns out to be linear. We discuss this simpler estimation approach at the end of this section.

2.3.1 Step 1: Estimating $\tau_j(\cdot)$ and $f_{\theta_{a,t^*}}(\cdot)$

We can estimate $\tau_j(\cdot)$ functions and the skill distributions $f_{\theta_{a,t^*}}(\cdot)$ for all ages using cross-sectional data at time t^*. Normalizing $\lambda_{t^*} = 1$ and $\alpha_{a^*,t^*} \equiv E(\theta_{i,a^*,t^*}) = 0$, we consider a nonparametric maximum likelihood estimation (NPMLE) approach by using flexible functional form and distributional assumptions (e.g. polynomials for measurement functions, mixtures of normal distributions for densities, or sieve estimation using Hermite polynomials). The complexity of the underlying parameter space can be adjusted depending on the model structure and the sample size.

Let $f_{w_{a^*,t^*},T_{1,,a^*,t^*},\ldots,T_{J,a^*,t^*}}$ be the joint density function of $(w_{a^*,t^*},, T_{1,a^*,t^*}, \ldots, T_{J,a^*,t^*})$.

Since observations are i.i.d. over individuals, we drop the subscript i unless it causes any confusion. For simplicity, we assume that $G(\cdot)$ is an identity function and all T_j are continuous.[13] The independence assumption among $(\theta_{a^*,t^*}, \varepsilon_{a^*,t^*}, \eta_{1,a^*,t^*}, \ldots, \eta_{J,a^*,t^*})$ implies that

$$\begin{aligned}
& f_{w_{a^*,t^*},T_{1,a^*,t^*},\ldots,T_{J,a^*,t^*}}(w, T_1, \ldots, T_J) \\
&= \int_\Theta f_{\varepsilon_{a^*,t^*}}\left(w - \theta; \boldsymbol{\beta}_{\varepsilon_{a^*,t^*}}\right) \\
&\quad \times f_{\eta_{1,a^*,t^*}}\left(T_1 - \tau_1(\theta; \boldsymbol{\beta}_{\tau_1}); \boldsymbol{\beta}_{\eta_{1,a^*,t^*}}\right) \times \cdots \times f_{\eta_{J,a^*,t^*}}\left(T_J - \tau_J(\theta; \boldsymbol{\beta}_{\tau_J}); \boldsymbol{\beta}_{\eta_{J,a^*,t^*}}\right) \\
&\quad \times f_{\theta_{a^*,t^*}}\left(\theta; \boldsymbol{\beta}_{\theta_{a^*,t^*}}\right) d\theta,
\end{aligned} \tag{4}$$

where $\boldsymbol{\beta}_x$ for a generic x denotes a vector of the parameters or the polynomial coefficients for the unknown distribution or function. Recall that all density functions in (4) should satisfy the mean zero restriction. The above density function can be used to form the log-likelihood function. Let $\boldsymbol{\beta}_{a^*,t^*} \equiv (\{\boldsymbol{\beta}_{\tau_j}, \boldsymbol{\beta}_{\eta_{j,a^*,t^*}}\}_{j=1}^J, \boldsymbol{\beta}_{\varepsilon_{a^*,t^*}}, \boldsymbol{\beta}_{\theta_{a^*,t^*}})$ be the stacked vector of all unknown parameters. It can be estimated by

$$\widehat{\boldsymbol{\beta}}_{a^*,t^*} = \underset{\beta_{a^*,t^*} \in \mathcal{B}}{\arg\max} \frac{1}{|\mathcal{I}_{c^*}|} \sum_{i \in \mathcal{I}_{c^*}} \log f_{w_{a^*,t^*},T_{1,a^*,t^*},\ldots,T_{J,a^*,t^*}}(w_{i,a^*,t^*}, , T_{i,1,a^*,t^*}, \ldots, T_{i,J,a^*,t^*}; \boldsymbol{\beta}_{a^*,t^*}), \tag{5}$$

where $\mathcal{I}_{c^*}$ is the subset of individuals who belong to cohort $c^* = t^* - a^*$ and $|\mathcal{I}|$ is the number of elements in set $\mathcal{I}$. Of particular interest to us are the estimates $\{\widehat{\boldsymbol{\beta}}_{\tau_j}\}_{j=1}^J$ and $\boldsymbol{\beta}_{\theta_{a^*,t^*}}$, which give us the estimated measurement functions $\{\widehat{\tau}_j(\cdot)\}_{j=1}^J$ and skill distributions $\widehat{f}_{\theta_{a^*,t^*}}(\cdot)$, respectively.[14] We can repeat the estimation procedure in (5) for each cohort (a, t^*) and estimate the skill distribution $f_{\theta_{a,t^*}}$ of cohort $c = t^* - a$ at time t^*.

We can also increase the estimation efficiency of $\widehat{\tau}_j(\cdot)$ by including all cohorts in a single optimization procedure. For any cohort $c = a - t^*$, define the cohort specific objective function:

$$Q(\boldsymbol{\beta}_{a,t^*}) \equiv \frac{1}{|\mathcal{I}_c|} \sum_{i \in \mathcal{I}_c} \log f_{w_{a,t^*},T_{1,a,t^*},\ldots,T_{J,a,t^*}}(w_{i,a,t^*}, , T_{i,1,a,t^*}, \ldots, T_{i,J,a,t^*}; \boldsymbol{\beta}_{a,t^*}). \tag{6}$$

[13] When T_j for $j \geq 2$ include a discrete measure, we can replace $f_{\eta_{j,a^*,t^*}}$ with a proper discrete probability mass function. For example, see Appendix B

[14] Appendix B provides expressions for likelihoods assuming mixtures of normal distributions for errors and discusses the case of both discrete and continuous measurements, T_{j,a^*,t^*}. Details on estimation of standard errors for $\widehat{\boldsymbol{\beta}}_{a^*,t^*}$ are also provided in Appendix B.

Let $\boldsymbol{\beta}_{t^*} \equiv \{\boldsymbol{\beta}_{a,t^*}\}_{a\in\mathcal{A}}$ be the stacked parameter vector, where $\mathcal{A}$ is an index set of all different ages (cohorts) at time t^*. Then, the parameter of interest as well as some nuisance parameters can be estimated by

$$\widehat{\boldsymbol{\beta}}_{t^*} = \underset{\boldsymbol{\beta}_{t^*}\in\mathcal{B}^{|\mathcal{A}|}}{\arg\max} \sum_{a\in\mathcal{A}} Q(\boldsymbol{\beta}_{a,t^*})$$

Notice that we normalize the mean of the skill distribution only in cohort c^*, so the density function $f_{\theta_{a,t^*}}$ for $a \neq a^*$ is allowed to have a non-zero mean. This estimation approach will be more efficient if the measurement errors $\{\varepsilon_{a,t^*}\}_{a\in\mathcal{A}}$ and $\{\eta_{a,t^*}\}_{a\in\mathcal{A}}$ have identical distributions across different cohorts.

Finally, we note that if one is simply interested in determining whether any of the $\tau_j(\cdot)$ functions is linear (or only the density of skills in period t^* across all ages/cohorts is desired), then the likelihood in equation (4) can alternatively be written in terms of $f_{\theta_{t^*}}$ (with parameters $\boldsymbol{\beta}_{\theta_{t^*}}$), normalizing this density to be mean zero. Indeed, this is the approach we take below in determining that one of our measures is linear in skills.

2.3.2 Step 2: Estimating λ_t and $f_{\theta_{a,t}}(\cdot)$

Now, we discuss estimation of skill returns, λ_t, and the skill distributions, $f_{\theta_{a,t}}(\cdot)$, with additional cross-sectional data at time $t \neq t^*$. Embedding the estimated measurement function $\widehat{\tau}_1(\cdot)$ and the unknown skill return λ_t, we can write the density function at time t as

$$\begin{aligned}
&f_{w_{a,t},T_{1,a,t},\ldots,T_{J,a,t}}(w, T_1, \ldots, T_J)\\
&= \int_\Theta f_{\varepsilon_{a,t}}\left(w - \lambda_t\theta; \boldsymbol{\beta}_{\varepsilon_{a,t}}\right)\\
&\times f_{\eta_{1,a,t}}\left(T_1 - \widehat{\tau}_1(\theta); \boldsymbol{\beta}_{\eta_{1,a,t}}\right) \times f_{\eta_{2,a,t}}\left(T_2 - \tau_2(\theta; \boldsymbol{\beta}_{\tau_2}); \boldsymbol{\beta}_{\eta_{2,a,t}}\right) \times \cdots \times f_{\eta_{J,a,t}}\left(T_J - \tau_J(\theta; \boldsymbol{\beta}_{\tau_J}); \boldsymbol{\beta}_{\eta_{J,a,t}}\right)\\
&\times f_{\theta_{a,t}}\left(\theta; \boldsymbol{\beta}_{\theta_{a,t}}\right) d\theta.
\end{aligned} \tag{7}$$

Define $Q(\lambda_t, \tilde{\boldsymbol{\beta}}_{a,t})$ as in (6) by adding the skill return parameter λ_t and let $\tilde{\boldsymbol{\beta}}_t \equiv \{\tilde{\boldsymbol{\beta}}_{a,t}\}_{a\in\mathcal{A}}$, where we drop $\boldsymbol{\beta}_{\tau_1}$ from each $\boldsymbol{\beta}_{a,t}$ as we already plugged in the estimate from Step 1. If any measure observed at time t^* is repeated at time t, we can replace it with $\widehat{\tau}_j(\cdot)$ and drop the relevant parameters. The parameter set is further simplified when $\varepsilon_{a,t}$ and $\eta_{j,a,t}$ are age/time-invariant since we can plug-in the corresponding estimates from Step 1. Then, we can estimate the skill return and other underlying parameters at time t by

$$\left(\widehat{\lambda}_t, \widehat{\boldsymbol{\beta}}_t\right) = \underset{(\lambda_t, \tilde{\boldsymbol{\beta}}_t) \in \Lambda \times \tilde{\mathcal{B}}^{|\mathcal{A}|}}{\arg\max} \sum_{a \in \mathcal{A}} Q(\lambda_t, \tilde{\boldsymbol{\beta}}_{a,t}).$$

Once we obtain estimates $\widehat{\boldsymbol{\beta}}_{\theta_{a,t}}$ for skill distributions $f_{\theta_{a,t}}(\cdot)$, we can estimate $\alpha_{a,t}$ for all $a \in \mathcal{A}$ at time t. In addition, we can estimate time effects in the wage equation by $\widehat{\gamma}_t = \widetilde{\gamma} - \widehat{\lambda}_t \widehat{\alpha}_{a,t}$, where we have already estimated all the components on the right hand side.

2.3.3 Step 3: Estimating Skill Dynamics

We discuss estimation of skill dynamics for two different cases, both using panel data. First, for a general Markov skill process, we can apply the same idea as above to estimate its dynamics using any repeated continuous measure of skills. Given the estimated elements of the model, we can write the joint density function of repeated continuous measure j at time t and $t+1$ as follows:

$$\begin{aligned} f_{T_{j,a,t}, T_{j,a+1,t+1}}(T_t, T_{t+1}; \boldsymbol{\beta}_{\theta_{a,t}, \theta_{a+1,t+1}}) = \iint_{\Theta \times \Theta} & \widehat{f}_{\eta_{j,a,t}}(T_t - \widehat{\tau}_j(\theta_t)) \widehat{f}_{\eta_{j,a+1,t+1}}(T_{t+1} - \widehat{\tau}_j(\theta_{t+1})) \\ & \times f_{\theta_{a,t}, \theta_{a+1,t+1}}(\theta_t, \theta_{t+1}; \boldsymbol{\beta}_{\theta_{a,t}, \theta_{a+1,t+1}}) d\theta_t d\theta_{t+1}, \end{aligned} \tag{8}$$

where $f_{\theta_{a,t}, \theta_{a+1,t+1}}(\theta_t, \theta_{t+1}; \boldsymbol{\beta}_{\theta_{a,t}, \theta_{a+1,t+1}})$ is the joint density function of $(\theta_{a,t}, \theta_{a+1,t+1})$. The measure j specific objective function can be defined as

$$Q_j(\boldsymbol{\beta}_{\theta_{a,t}, \theta_{a+1,t+1}}) = \frac{1}{|\mathcal{I}_c|} \sum_{i \in \mathcal{I}_c} \log f_{T_{j,a,t}, T_{j,a+1,t+1}}(T_{i,j,a,t}, T_{i,j,a+1,t+1}; \boldsymbol{\beta}_{\theta_{a,t}, \theta_{a+1,t+1}}).$$

Then, the parameters for the joint density function are estimated by

$$\widehat{\boldsymbol{\beta}}_{\theta_{a,t}, \theta_{a+1,t+1}} = \underset{\boldsymbol{\beta}_{\theta_{a,t}, \theta_{a+1,t+1}} \in \mathcal{B}_\theta}{\arg\max} Q_j(\boldsymbol{\beta}_{\theta_{a,t}, \theta_{a+1,t+1}}).$$

Once we have estimated the joint density function, the dynamics of the skill process follow immediately from the conditional density function.

Second, if the skill process follows the AR(1) with fixed effect structure as in (3), the parameters for this process can be estimated following a similar strategy as above using a modified version of equation (8) that incorporates an additional time period to estimate the joint density $f_{\theta_t, \theta_{t+1}, \theta_{t+2}}$ where the parameters for this density, $\boldsymbol{\beta}_{\theta_{a,t}, \theta_{a+1,t+1}, \theta_{a+2,t+2}}$ include the relevant cohort $c = t - a$ distribution for ψ_i and parameters of the AR(1) process (ρ and parameters determining distributions for $\phi_{i,a,t}$, $\nu_{a+1,t+1}$, and $\nu_{a+2,t+2}$).

2.3.4 Estimation of Skill Distributions, Skill Dynamics, and Returns to Skill when a Linear Measurement is Available

In our empirical context, one of the measurements, say $T_{1,a,t}$, is determined to be linear in skills from the estimation procedure described in Section 2.3.1. We use this information to facilitate estimation of the returns to skill over time and the evolution and dynamics of skills assuming the special case where $\theta_{i,a,t}$ follows the AR(1) plus fixed effect process described in equation (3).

Using the known linear measurement $T_{i,1,a,t} = \beta_{1,0} + \beta_{1,1}\theta_{i,a,t} + \eta_{i,1,a,t}$, our model for log wage residuals and the skill measurement can be written in terms of de-meaned skills:

$$\begin{aligned}
w_{i,a,t} &= \lambda_t \bar{\theta}_{i,a,t} + \varepsilon_{i,a,t}, \\
\bar{\theta}_{i,a,t} &= \psi_i + \phi_{i,a,t}, \\
\phi_{i,a,t} &= \rho\phi_{i,a-1,t-1} + \nu_{i,a,t}, \\
T_{i,1,a,t} &= (\beta_{1,0} + \beta_{1,1}\alpha_{a,t}) + \beta_{1,1}\bar{\theta}_{i,a,t} + \eta_{i,1,a,t}.
\end{aligned}$$

These imply the following covariances for (a, t):

$$\begin{aligned}
Cov(w_{a,t}, T_{1,a+k,t+k}) &= \lambda_t \beta_{1,1} \left[Var(\psi|t-a) + \rho^k Var(\phi_{a,t})\right], \qquad \text{for } k \geq 0 \\
Cov(T_{1,a,t}, T_{1,a+k,t+k}) &= \beta_{1,1}^2 \left[Var(\psi|t-a) + \rho^k Var(\phi_{a,t})\right], \qquad \text{for } k \geq 1 \\
Cov(T_{1,a,t}, w_{a+k,t+k}) &= \lambda_{t+k} \beta_{1,1} \left[Var(\psi|t-a) + \rho^k Var(\phi_{a,t})\right], \qquad \text{for } k \geq 1.
\end{aligned}$$

Assuming the distribution of skill shocks depends only on time, we define $\sigma_{\nu_t}^2 \equiv Var(\nu_{a,t})$ for all (a, t) and can write

$$Var(\phi_{a,t}) = \rho^{2(t-t_1)} Var(\phi_{a-(t-t_1),t_1}) + \sum_{s=t_1+1}^{t} \rho^{2(t-s)} \sigma_{\nu_s}^2, \qquad \forall t \geq t_1 + 1,$$

where t_1 is the initial period of observation. As discussed earlier, we normalize $\lambda_{t^*} = 1$ and $\alpha_{a^*,t^*} = 0$. With these assumptions, the generalized methods of moments (GMM) can be used to jointly estimate the time-varying returns to skill (λ_t), autocorrelation for skill shocks (ρ), variances of initial skills by cohort ($Var(\psi|t-a)$ for all observed cohorts), initial variances of the persistent skill shock ($Var(\phi_{a,t_1})$ for cohorts observed in initial period t_1 and $Var(\phi_{a_1,t})$ for cohorts entering the sample at age a_1 at later dates), time-varying skill shock variances ($\sigma_{\nu_t}^2$), and the measurement function parameters ($\beta_{1,0}, \beta_{1,1}$). Further details on estimation and calculation of the standard errors are provided in Appendix C.

3 HRS Data

We use data from the Health and Retirement Study (HRS), a national U.S. panel survey of individuals over age 50 and their spouses.[15] It consists of seven cohorts with the initial cohort first interviewed in 1992. New cohorts of individuals were added in 1993, 1998, 2004, 2010, and 2016.[16] The survey has been fielded every two years since 1992 and it provides information about demographics, income, and cognition, making it ideal data for the purpose of our study. Because one of the cognitive tests (word recall) in 1992 and 1994 differs from the later years, we use data collected from 1996 to 2016.[17]

The HRS records the respondent's and spouse's wage rates if they are working at the time of the interview. We use the hourly wage rate, deflating nominal values to 1996 dollars using the Consumer Price Index.[18] The HRS also provides various cognitive functioning measures. We use four measures in our estimation: word recall, serial 7's, quantitative reasoning, and retrieval fluency. Table 1 provides a brief summary of these measures. The word recall test evaluates the memory of the respondents by reading a list of 10 words and asking them to recall immediately (immediate recall) and after a delay of about 5 minutes (delayed recall). We sum up the number of words the respondent recalled in the two tasks and obtain a score of 21 different values. The serial 7's test asks the respondent to subtract 7 from the previous number, starting with 100 for five trials. This test score is the number of trials that the respondent answered correctly, and it has 6 different values. Quantitative reasoning consists of three simple arithmetic questions assessing the numeracy of the respondent. We construct a test score based on the answers and the resulting score ranges from 0 to 4. The retrieval

[15]More precisely, the sample does include some individuals age 50. For example, someone from the original cohort (born in 1931-1941) who was born late in 1941 may have been age 50 at the date of their first interview in 1992 if they were interviewed earlier in the calendar year.

[16]The HRS sample was built up over time. The initial cohort consisted of persons born between 1931 and 1941 (aged 51 to 61 at first interview in 1992). The Asset and Health Dynamics Among the Oldest Old (AHEAD) cohort, born before 1924 was added in 1993 and interviewed in 1993, 1995, and biennially from 1998 forward. In 1998, two new cohorts were enrolled: the Children of the Depression (CODA) cohort, born 1924 to 1930, and the War Baby (WB) cohort, born 1942 to 1947. Early Baby Boomer (EBB, born 1948 to 1953) cohort was added in 2004, Mid Baby Boomer (MBB, born 1954 to 1959) cohort was added in 2010, and Late Baby Boomer (LBB, born 1960 to 1965) cohort was added in 2016. In addition to respondents from eligible birth years, the survey interviewed the spouses of married respondents or the partner of a respondent, regardless of age.

[17]The word recall test contains a list of 20 words in 1992 and 1994, while it has been reduced to 10 words in later years.

[18]https://www.bls.gov/cpi/research-series/home.htm#CPI-U-RS20Data

fluency test asks the respondents to name as many animals as they can in 60 seconds. The test score is the total number of correct answers, ranging from 0 to 90. Additional details about the measures and the construction of other key variables are provided in Appendix D.

Our sample is restricted to age-eligible (i.e. born in eligible years when first interviewed) men. We use observations when men are ages 50-70 if their potential labor market experience is between 30 and 50 years.[19] We trim the top and bottom 1% of all wages within year by college- vs. non-college-educated status and 10-year experience cells. In estimation, we use non-imputed wages and cognitive measures only. The sample contains 9,848 individuals and 37,518 person-year observations.

Our sample consists of 64% white, 18% black, 13% Hispanic, and 5% other races with an average age of 60 years. We create five education categories based on years of education: 0-11 years (less than high school graduate), 12 years (high school graduate), 13-15 years (some college), 16 years (college graduate), and 17 or more years (above college). In our sample, 20% had less than 12 years of schooling, 30% had 12 years of schooling, 24% had some college, 14% completed college, and 13% had more than 16 years of schooling. Table 2 shows the mean and the standard deviation of the cognitive scores and log hourly wage, along with the correlation between these variables. The correlations between test scores range from nearly 0.3 to 0.5, with the highest correlation between Serial 7's and quantitative reasoning. Retrieval fluency has the lowest correlation with log wages (0.214), while quantitative reasoning has the highest (0.303).

4 Estimation Results

4.1 Cross-Sectional Results for Measurement Functions

As discussed in Section 2.3.1, we use data from a single year, $t^* = 2010$, to estimate measurement functions $\tau_j(\cdot)$, as well as $F_{\theta_{t^*}}(\cdot)$, $F_{\varepsilon_{t^*}}(\cdot)$, and $F_{\eta_{j,t^*}}(\cdot)$. We use data from 2010, because this is the only year that all four cognitive measures we consider were recorded for every respondent.[20] We further restrict the sample to have non-missing wages and at least one non-missing cognitive measure in 2010. The sample size for this 2010 analysis is 1,980.

[19]Potential experience is defined as age minus 6 minus years of schooling.

[20]In other waves, either one or more of the cognitive tests were not administered or some of the testes were only administered to new interviewees and/or re-interviewees ages 65 or older.

We estimate the parameters by maximum likelihood as described in Section 2.3.1, normalizing $\lambda_{t^*} = 1$. We treat word recall (T_1) and retrieval fluency (T_2) as continuous measures, assuming both $\tau_1(\cdot)$ and $\tau_2(\cdot)$ are polynomial functions. In practice, we use likelihood ratio tests to determine the polynomial degree for each measure. We treat the serial 7's (T_3) and quantitative reasoning (T_4) scores as discrete measurements generated from ordered probits with latent index functions $\tau_j(\theta_{i,t}) + \eta_{i,j,t}$.[21]

We estimate the model for three cases. In Case 1, we assume that log wage shocks ε_t, continuous measurement errors (η_1, η_2), and unobserved skills θ_t are all normally distributed. Case 2 assumes that skill θ_t is distributed as a mixture of two normal distributions, while ε_t and (η_1, η_2) are all normally distributed. Finally, Case 3 is most general, assuming both skill θ_t and wage shocks ε_t are distributed as mixtures of two normal distributions, while the measurement errors (η_1, η_2) are normally distributed. While we assume the distributions of measurement errors are time invariant, we allow the distributions for skills and log wage shocks to vary freely over time.

For each case, we estimate different specifications by increasing the degree of polynomials for τ_1 and τ_2, starting from a linear specification until the model cannot be improved further, as determined by likelihood ratio tests. Then, the "best" specifications from each of the three cases are compared to determine the "best" overall specification, again using the likelihood ratio test. Table 3 reports the log-likelihood associated with several specifications and the three cases, along with likelihood ratio test statistics and p-values. Based on the likelihood ratio tests, the "best" overall specification allows both θ_t and ε_t to be distributed as mixtures of normal distributions. Furthermore, $\tau_1(\cdot)$ (word recall) is linear in skill, while $\tau_2(\cdot)$ (retrieval fluency) is a polynomial of degree 7 in skill. Table 4 reports parameter estimates and standard errors for this preferred specification.

4.2 Results for Skill Distributions, Skill Dynamics, and Returns to Skill

The fact that word recall scores are linear in skill is convenient and enables us to use the relatively simple approach described in Section 2.3.4 to estimate the returns to skill over

[21]For each of the discrete measures, we have $K_j = 5$ or 6 choices, and we estimate linear terms (intercept terms are normalized to zero) for the $\tau_j(\cdot)$ functions along with $K_j - 1$ cutoff parameters. See Appendix B for details.

time, evolution of skill distributions over time, and the dynamics of skills. To do so, we use the panel nature of the HRS from 1996-2016 (11 biennial surveys), further restricting the sample to include men ages 52-65. There are very few observations outside of that age range for the years we examine.

Since we only observe wages for those who are working at the time of the survey, there are natural concerns about the implications of selection for any covariance moments that include log wage residuals. (Fortunately, test scores are available regardless of work status.) To explore the potential implications of selection, we consider four different sampling schemes for this analysis:

1. "Full" Sample: This includes covariances for log wage residuals and/or test scores whenever they are available. Therefore, covariances using log wage residuals are only calculated for workers, while covariances based only on test scores are calculated for both workers and non-workers.

2. "Worker" Sample: This eliminates covariances for test score measures unless an individual is working in both periods.

3. "Wage 50-60" Sample: This only includes covariances with log wage residuals for years when the worker is ages 50-60.

4. "Exp 30-40" Sample: This only includes covariances with log wage residuals for years when the worker has 30-40 years of potential experience.

The "Full" Sample raises the most concern about selection due to early retirement. For example, if some (e.g. lower skill) workers retire early when they experience a low wage shock, ε_t, while other workers (e.g. higher skill) do not, this can distort the covariance between log wage residuals and test score measures. Yet, the covariance between test scores is unaffected by this sort of selection. The "Worker" Sample does not address the selection problems, but it would provide more direct estimates that apply to the selected sample of workers. The "Wage 50-60" and "Exp 30-40" Samples address concerns about selective retirement to the extent that the vast majority of men are still working throughout their 50s or prior to reaching 40 years of potential experience (e.g. age 58 for a high school graduate). Even those retiring "early" typically work during these years.

Using data available for even-numbered years, we use the GMM approach of Section 2.3.4 (with identity weighting matrix) to estimate $\beta_{1,1}$ (for word recall), returns to skill λ_t, and the evolution of skill distributions. Because we allow for age/cohort variation in the distribution of "initial" $\phi_{a,t}$ skill shocks, we assume a cohort-invariant distribution of skill fixed effects (i.e. $Var(\psi|c) = Var(\psi)$ for all cohorts c). Regarding the skill distributions, we estimate the two-year autocorrelation for persistent skills ρ^2, variance of fixed effects $Var(\psi)$, variances of skill shocks $\sigma^2_{\nu_t}$, and variances of "initial" $\phi_{a,t}$ skill shocks when individuals enter the sample.[22]

Table 5 reports the estimates and standard errors for $\beta_{1,1}$, ρ^2, and $Var(\psi)$ for all four samples. The estimates are fairly similar across samples; although, the estimated $\beta_{1,1}$ mapping skills into word recall scores ranges from 9.0 for the "Worker" Sample to 11.8 for the "Exp 30-40" Sample. Of greater interest are the skill fixed effects variance estimates, which range from 0.014 to 0.025. Based on the "Full" Sample estimates, these suggest that variation in these permanent skill differences accounts for 38% of the variation in skills and 5% of the variation in log wages at age 60 in 2002. Our estimates for ρ^2, which reflect the dynamics of skill shocks, are also similar across samples at 0.86 to 0.87. These imply values for ρ of about 0.93, within the typical range of autocorrelations for log earnings innovations in the earnings dynamics literature (Meghir and Pistaferri, 2011).

The estimated time patterns for $\sigma^2_{\nu_t}$ are shown in Figure 1, while estimates for $Var(\phi_{a,t})$ are shown in Figure 2 (estimates for all ages in 1996) and Figure 3 (estimates for ages 52 and 53 for years 1998-2016). These estimates suggest considerable stability in the process for persistent skill shocks over time and across cohorts.

Finally, Figure 4 plots estimated returns to skill, λ_t, over time for all samples, along with their 95% confidence intervals. Estimated profiles for all four samples suggest relative stability in skill returns until the Great Recession, after which they appear to rise steadily through the end of our sample period. Point estimates for the "Full" Sample suggest that the returns to skill rose from a low of 0.82 in 2008 to a high of 1.21 in 2016.[23] The qualitative

[22]We estimate λ_t for $t = 1996$ to 2016, normalizing $\lambda_{t^*} = 1$ for $t^* = 2010$. We estimate $\sigma^2_{\nu_t}$ for years $t = 1998$ to 2016, normalizing $\sigma^2_{\nu_{1996}} = 0$. We estimate variances of initial AR(1) skill shocks for men first observed in 1996, $Var(\phi_{a,1996})$ for ages $a = 52$ to 65, as well as for men first observed at ages 52 and 53 in other years, i.e. $Var(\phi_{52,t})$ and $Var(\phi_{53,t})$ for $t = 1998$ to 2016.

[23]Using a Wald test, we reject that the return does not change from 2008 to 2016 at 1% significance level.

pattern of relatively stable returns to skill in the late 1990s and early 2000s, followed by a rise after 2008, is consistent with the estimates of Lochner et al. (2020); however, standard errors for $\hat{\lambda}_t$ are large, making it is difficult to say how much returns actually rose after 2008, or to evaluate year-to-year changes in returns, with any confidence.

4.3 Average Skill Profiles

We now explore average skill profiles by age and time for various subpopulations of interest. Because we observe cognitive test scores for individuals whether they work or not, we can examine the evolution of skills through and after retirement. Indeed, we report average lifecycle skill profiles over ages 50-70. We begin by showing profiles for different birth cohorts in the HRS, then consider different skill profiles by education and race. Finally, we explore differences in skill profiles for workers who retire at different ages.

Since $T_{i,1,a,t} = \beta_{1,0} + \beta_{1,1}\theta_{i,a,t} + \eta_{i,1,a,t}$ for word recall, we can write

$$\theta_{i,a,t} = \frac{T_{i,1,a,t} - \beta_{1,0}}{\beta_{1,1}} - \frac{\eta_{i,1,a,t}}{\beta_{1,1}},$$

where $\eta_{i,1,a,t}$ is mean zero. Linearity of the test score function implies that actual skills are simply a re-scaled measure of test scores plus idiosyncratic noise. While test score measures only allow us to obtain very noisy estimates for any specific individual's skill level, we can obtain much more precise estimates of average skill levels.

Since we normalize $\alpha_{54,1996} = 0$, we have $\beta_{1,0} = E(T_{1,54,1996})$, which can easily be estimated using the sample mean for word recall scores among individuals age 54 in 1996: $\hat{\beta}_{1,0} = \bar{T}_{1,54,1996}$. Using this along with our estimate of $\hat{\beta}_{1,1}$ from Table 5, we estimate average skills as

$$\widehat{\alpha}_{a,t} = \frac{\bar{T}_{1,a,t} - \hat{\beta}_{1,0}}{\hat{\beta}_{1,1}}, \tag{9}$$

where $\bar{T}_{1,a,t}$ reflects average $T_{1,a,t}$ for all individuals age a in year t. We can similarly obtain estimates for average skills (by age and time) conditional on any personal characteristics as long as those characteristics are independent of cognitive achievement measurement errors. In this case, we would simply use average word recall scores for the subpopulation of interest in equation (9). It is also worth noting that we can calculate average skills for ages outside the range we used in estimation, assuming that the measurement function mapping skills to test scores is age-invariant.

We begin by estimating average skills by age and time, $\widehat{\alpha}_{a,t}$. Recall that skills are measured in log wage units (as of 2010), so differences in skill translate roughly into percentage differences. Rather than show all of these estimates, we first regress $\widehat{\alpha}_{a,t}$ on age and year indicators (weighting by the number of observations in each age-year cell) to explore the extent to which these vary with age and time. In Figure 5, panel (a) plots the regression coefficients on age indicators for all four estimation samples, while panel (b) plots the regression coefficients on year dummies. The age patterns suggest relative stability, except for a roughly 3 percentage point jump up in average skills from age 52 (our base group in the regression) to age 53 and a similar sized drop between ages 61 and 63. Panel (b) shows relative stability in average skills over time with a sharp drop between 2008 and 2010 with the introduction of the Mid-Baby Boomer cohort.

Because cohorts may differ in their skills, the introduction of new cohorts to the HRS sample can produce jumps up and down in average skills like those seen Figure 5. We next look at the age profiles (ages 52-70) for different cohorts, which should be representative of average skills for those cohorts. Because each of these cohorts faced different educational, social, and economic conditions throughout their lives, one might expect to observe differences in their accumulated skills as of age 52 and beyond. Indeed, we see sizeable differences as documented in Figure 6. From age 52 to 62, average skill levels are highest for men born during World War II (WWII) and earlier, followed by the Early Baby Boomer cohorts (born 1948-1953), and then subsequent cohorts. At age 55, cohorts born before the War had average skill levels that were about 11 percentage points higher than men born between 1954 and 1959. By age 60, this gap had shrunk to about 3 percentage points, disappearing by age 63. Beyond age 63, cohort differences are small, even reversing with the earliest cohort exhibiting a much more rapid decline in skills with age compared to the War Babies and Early Baby Boomers. While average skill levels began to decline with age for men in their mid-50s for the cohorts born before, during, and immediately after WWII (by 8-13 percentage points), skill profiles remained relatively flat for the Mid-Baby Boomer cohorts (born 1954-59) throughout their late-50s and early-60s. Unfortunately, few men from the most recent cohort (born 1960-65) are observed beyond age 55, so it is difficult to say whether the apparent flattening of lifecycle skill profiles among men ages 55+ will continue.

Next, we explore whether lifecycle skill profiles among older workers differ systematically by education or race. To remove the influence of any cohort differences, we regress $\hat{\theta}_{i,a,t} = (T_{i,1,a,t} - \hat{\beta}_{1,0})/\hat{\beta}_{1,1}$ on HRS cohort indicators and interactions between annual age indicators and educational attainment indicators (less than high school, high school graduate, some college, college graduate, post-graduate) or race indicators (white vs. non-white).

Not surprisingly, Figure 7 shows sizeable and statistically significant differences in skills across education groups, with college graduates possessing about 15-20% higher skill levels than high school graduates over ages 56-70.[24] High school dropouts have 10-17% lower skill levels than high school graduates. What is, perhaps, most noteworthy about this figure is the apparent parallelism in skills, even through typical retirement and post-retirement ages. Skills are systematically declining beyond age 55 with similar rates of decline for all education groups.[25]

Figure 8 shows the estimated average skill profiles by race. Consistent with lower wages among non-whites, we see that average skill levels are about 10-20 percentage points lower for non-white men over ages 52-57. (These gaps are statistically significant at all ages.) As with education, we see similar lifecycle profiles for both whites and non-whites.[26]

Table 6 reports average estimated skill levels and log wage residuals, which net out age and time effects, by race and education for individuals ages 55-60. Column 1 reports average skills for the full sample, while column 2 reports average skills for the sample of workers (i.e. respondents reporting wages during the same periods). Average skill levels by education and race are larger for the sample of workers, but the differences are modest and vary little across education and race groups. This suggests that selection into retirement has quite modest effects on average skill levels in the workforce. As already evident in Figures 7 and 8, the average skill gap between college and high school graduates is quite similar to the skill gap between whites and non-whites, about 15%. Column (3) shows that the corresponding gaps in average log wage residuals (also at ages 55-60) are much larger — college graduates have 44% higher wages than high school graduates, while the race gap in wages is about 35%.

[24] F-tests for equality of skills across any education comparison at any age from 52 to 70 yield p-values less than 0.05 for all but four comparisons.

[25] Using F-tests for equality of average changes in skill (from age a to $a+1$ for all available a shown in the figure) across education groups, we cannot reject parallelism in age profiles (at 5% significance level) for any education comparison. We also cannot reject parallelism for any education groups over subperiods, including ages 55-60, 60-65, and 65-70.

[26] Based on F-tests, we only reject parallelism over ages 52-55. We cannot reject parallelism over ages 55-60, 60-65, and 65-70.

Thus, the cognitive skills captured by our measures explain an important share of education and race wage gaps, but other factors also play an important role.

The results presented so far suggest a systematic decline in skills for men that begins when they are in their mid-50s (or earlier). Is this explained by a gradual increase in rates of retirement with sharp declines in skills for those who retire, or does it reflect more gradual declines for all workers regardless of when they decide to retire? The patterns presented in Figure 9 favor the latter explanation. Panel (a) shows lifecycle average skill profiles separately for workers who retire at ages 50-54, 55-59, 60-64, and 65+, while panel (b) removes cohort effects by regressing $\hat{\theta}_{i,a,t}$ on cohort indicators and interactions of age indicators with retirement age indicators (as with education and race above). In neither case do we see evidence of steep drops over the ages when individuals retire or in the time immediately following retirement.

Still, the skill levels and lifecycle patterns notably differ for those who retire early and those who retire late compared to those retiring between ages 55 and 64. Those who retire before age 55 have much lower skill levels in their 50s compared to those who retire later; however, their skills continue to grow until age 61, while the skills of men retiring at ages 55-64 decline over most of these ages. Differences between very early retirees and those retiring at ages 55-64 are largely eliminated by age 60.

Those who retire at ages 65 or older possess the highest skill levels; however, their lifecycle profile over ages 52-70 looks more like that of very early retirees than those retiring in their late 50s and early 60s. These late retirees experience modest skill growth until their mid-50s, stable skill levels to about age 60, and strong declines thereafter. By age 65, about half of the difference in skills between them and those retiring over ages 55-64 is eliminated.

These patterns imply a complex relationship between retirement and skills. There is clear evidence that those with high skills in their mid-50s choose to retire late while those with low skills (a difference of more than 10 percentage points) choose to retire quite early. But, there is little evidence to suggest that retirement itself is strongly associated with a decline in skills.[27] Among those retiring very young, skills continue to increase for years after they retire, several years after they have already started declining for those retiring in their late

[27]Both Rohwedder and Willis (2010) and Bonsang et al. (2012) estimate significant negative causal effects of retirement on cognition using the HRS; however, Coe et al. (2012) does not.

50s or early 60s. Over the same ages, skills are also increasing or stable for men who retire after age 65.

5 Conclusions

With multiple skill measures and wages each period, we have shown that if at least one measure is continuous and repeated, it is possible to nonparametrically identify the evolution of skill prices and cross-sectional skill distributions over time without any assumptions on the distributions or dynamics of skills. With panel data, the same measurements and wages can identify skill dynamics as well. Our constructive identification analysis motivates a very general multi-step estimation approach. We also show that if any of the continuous measurements is found to be linear in skills (in the first estimation step), a simple GMM approach can be used to estimate skill returns, the means and variances of skill distributions over time, and a flexible dynamic process for skills with a fixed effect and AR(1) stochastic process.

Using data from the 1996-2016 HRS, we estimate the evolution of skill returns, skill distributions, and skill dynamics for American men ages 52+ over that period. We first show that one of the repeated continuous test measures we observe is linear in skills and then use our simpler GMM estimation approach. Our estimates suggest that the returns to (cognitive) skills were fairly stable from the mid-1990s through the early 2000s, but then began to rise significantly after the Great Recession. This pattern is broadly consistent with that of Lochner et al. (2020).

We document considerable differences in average skill levels and lifecycle profiles across cohorts. More recent cohorts of men had significantly lower average skill levels in their mid-50s than did earlier cohorts when they were the same ages. However, earlier cohorts experienced much faster declines in skill with age, such that the earlier skill differences had largely disappeared by the time cohorts had reached their 60s. For the latest cohort we observe (men born in 1954–1959), we see no discernable decline in average skills prior to age 63 when they are last observed. Distinguishing individuals by education and race, we find that average skills are monotonically increasing in education and are higher for whites than non-whites. These education and race gaps are quite similar across ages and explain about one-third of the education differences and nearly half of the race differences in log wages.

We also consider the interaction of skills and retirement, showing that those who retire at older ages have substantially higher skills in their mid-50s. While skills generally decline with age for men, at least after reaching age 60, we see no sharp declines around the time men retire. To the extent that retirement does lead to cognitive decline, our results suggest that the effects are relatively modest or largely offset by other lifecycle forces.

Finally, we show that individual fixed effects account for more than a third of all skill variation at age 60. Year-to-year fluctuations are also persistent (though not a random walk) with an autocorrelation of 0.93.

In future work with the HRS data, we plan to test the validity of previous assumptions in the literature regarding the evolution of skill differences or skill growth over the lifecycle. If some of these assumptions are shown to be valid, it would provide additional credibility to previous studies and further justification for those assumptions when using other data sources without direct skill measures. We can also make more use of additional test measures (even those that are non-linear in skills), using our GMM approach to obtain more precise estimates of skill returns and skill distributions over time. In addition to measuring differences in average skill levels over time, it is straightforward to estimate changes in the distributions of skills. For example, we can estimate the distributions of skills for workers choosing to retire at different ages to better understand selection into retirement. Finally, it is possible to allow log wage equations to differ by education and/or race, accounting for the fact that other factors or skills (besides those measured by the cognitive tests in HRS) might play an important role in wage determination.

References

Adam, Stéphane, Eric Bonsang, Sophie Germain, and Sergio Perelman (2007) "Retirement and Cognitive Reserve: A Stochastic Frontier Approach Applied to Survey Data," CREPP Working Paper 2007/04.

Altonji, Joseph G, Prashant Bharadwaj, and Fabian Lange (2012) "Changes in the Characteristics of American Youth: Implications for Adult Outcomes," *Journal of Labor Economics*, Vol. 30, pp. 783–828.

Autor, David H, Lawrence F Katz, and Melissa S Kearney (2008) "Trends in US Wage Inequality: Revising the Revisionists," *The Review of Economics and Statistics*, Vol. 90, pp. 300–323.

Blau, Francine D and Lawrence M Kahn (1997) "Swimming Upstream: TRends in the Gender Wage Differential in the 1980s," *Journal of labor Economics*, Vol. 15, pp. 1–42.

——— (2017) "The Gender Wage Gap: EXtent, Trends, and Explanations," *Journal of Economic Literature*, Vol. 55, pp. 789–865.

Bonsang, Eric, Stéphane Adam, and Sergio Perelman (2012) "Does Retirement Affect Cognitive Functioning?" *Journal of Health Economics*, Vol. 31, pp. 490–501.

Bowlus, Audra J and Chris Robinson (2012) "Human Capital Prices, Productivity, and Growth," *American Economic Review*, Vol. 102, pp. 3483–3515.

Cagney, Kathleen A and Diane S Lauderdale (2002) "Education, Wealth, and Cognitive Function in Later Life," *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, Vol. 57, pp. P163–P172.

Card, David (1999) "The Causal Effect of Education on Earnings," in *Handbook of Labor Economics*, Vol. 3: Elsevier, pp. 1801–1863.

Card, David and Thomas Lemieux (1996) "Wage Dispersion, Returns to Skill, and Black-White Wage Differentials," *Journal of Econometrics*, Vol. 74, pp. 319–361.

Castex, Gonzalo and Evgenia Kogan Dechter (2014) "The Changing Roles of Education and Ability in Wage Determination," *Journal of Labor Economics*, Vol. 32, pp. 685–710.

Castora-Binkley, Melissa, Carol L Peronto, Jerri D Edwards, and Brent J Small (2015) "A Longitudinal Analysis of the Influence of Race on Cognitive Performance," *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, Vol. 70, pp. 512–518.

Celidoni, Martina, Chiara Dal Bianco, and Guglielmo Weber (2017) "Retirement and Cognitive Decline. a Longitudinal Analysis Using SHARE Data," *Journal of Health Economics*, Vol. 56, pp. 113–125.

Chay, Kenneth Y and David S Lee (2000) "Changes in Relative Wages in the 1980s Returns to Observed and Unobserved Skills and Black–White Wage Differentials," *Journal of Econometrics*, Vol. 99, pp. 1–38.

Coe, Norma B, Hans-Martin von Gaudecker, Maarten Lindeboom, and Jürgen Maurer (2012) "The Effect of Retirement on Cognitive Functioning," *Health Economics*, Vol. 21, pp. 913–927.

Coe, Norma B and Gema Zamarro (2011) "Retirement Effects on Health in Europe," *Journal of Health Economics*, Vol. 30, pp. 77–86.

Deming, David J (2017) "The Growing Importance of Social Skills in the Labor Market," *The Quarterly Journal of Economics*, Vol. 132, pp. 1593–1640.

Edin, Per-Anders, Peter Fredriksson, Martin Nybom, and Bjorn Ockert (2017) "The Rising Return to Non-Cognitive Skill," Working Paper.

Grogger, Jeff and Eric Eide (1995) "Changes in College Skills and the Rise in the College Wage Premium," *Journal of Human Resources*, pp. 280–310.

Hanushek, Eric A, Guido Schwerdt, Simon Wiederhold, and Ludger Woessmann (2015) "Returns to Skills Around the World: EVidence From PIAAC," *European Economic Review*, Vol. 73, pp. 103–130.

Hanushek, Eric A and Lei Zhang (2009) "Quality-Consistent Estimates of International Schooling and Skill Gradients," *Journal of Human Capital*, Vol. 3, pp. 107–143.

Heckman, James J, Lance J Lochner, and Petra E Todd (2006) "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond," *Handbook of the Economics of Education*, Vol. 1, pp. 307–458.

Heckman, James J, Lance Lochner, and Christopher Taber (1998) "Explaining Rising Wage Inequality: Explorations With a Dynamic General Equilibrium Model of Labor Earnings With Heterogeneous Agents," *Review of Economic Dynamics*, Vol. 1, pp. 1–58.

Heckman, James and Edward Vytlacil (2001) "Identifying the Role of Cognitive Ability in Explaining the Level of and Change in the Return to Schooling," *Review of Economics and Statistics*, Vol. 83, pp. 1–12.

Herrnstein, Richard J and Charles A Murray (1994) *The Bell Curve: Reshaping of American Life by Differences in Intelligence*: Free Press.

Hu, Yingyao and Susanne M Schennach (2008) "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, Vol. 76, pp. 195–216.

Juhn, Chinhui, Kevin Murphy, and Brooks Pierce (1989) "Wage Inequality and the Returns to Skill," *Journal of Political Economy.*

Karlamangla, Arun S, Dana Miller-Martinez, Carol S Aneshensel, Teresa E Seeman, Richard G Wight, and Joshua Chodosh (2009) "Trajectories of Cognitive Function in Late Life in the United States: Demographic and Socioeconomic Predictors," *American Journal of Epidemiology*, Vol. 170, pp. 331–342.

Katz, Lawrence F and Kevin M Murphy (1992) "Changes in Relative Wages, 1963–1987: Supply and Demand Factors," *The Quarterly Journal of Economics*, Vol. 107, pp. 35–78.

Kotlarski, Ignacy (1967) "On characterizing the gamma and the normal distribution," *Pacific Journal of Mathematics*, Vol. 20, pp. 69–76.

Lemieux, Thomas (2006) "Postsecondary Education and Increasing Wage Inequality," *American Economic Review*, Vol. 96, pp. 195–199.

Leuven, Edwin, Hessel Oosterbeek, and Hans Van Ophem (2004) "Explaining International Differences in Male Skill Wage Differentials by Differences in Demand and Supply of Skill," *Economic Journal*, Vol. 114, pp. 466–486.

Lochner, Lance, Youngmin Park, and Youngki Shin (2020) "Wage Dynamics and Returns to Unobserved Skill," Working Paper.

Mazzonna, Fabrizio and Franco Peracchi (2012) "Ageing, Cognitive Abilities and Retirement," *European Economic Review*, Vol. 56, pp. 691–710.

——— (2017) "Unhealthy Retirement?" *Journal of Human Resources*, Vol. 52, pp. 128–151.

Meghir, Costas and Luigi Pistaferri (2011) "Earnings, Consumption and Life Cycle Choices," in David Card and Orley Ashenfelter eds. *Handbook of labor economics*, Vol. 4B: Elsevier, Chap. 9, pp. 773–854.

Murnane, Richard J, John B Willett, Frank Levy et al. (1995) "The Growing Importance of Cognitive Skills in Wage Determination," *The Review of Economics and Statistics*, Vol. 77, pp. 251–266.

Neal, Derek A and William R Johnson (1996) "The Role of Premarket Factors in Black-White Wage Differences," *Journal of Political Economy*, Vol. 104, pp. 869–895.

Rohwedder, Susann and Robert J Willis (2010) "Mental Retirement," *Journal of Economic Perspectives*, Vol. 24, pp. 119–38.

Taber, Christopher R (2001) "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?" *The Review of Economic Studies*, Vol. 68, pp. 665–691.

Zsembik, Barbara A and M Kristen Peek (2001) "Race Differences in Cognitive Functioning Among Older Adults," *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, Vol. 56, pp. S266–S274.

Table 1: Description of Cognitive Measures

	Meant to measure	Number of values	Available years
Word recall	Memory	21 (0-20)	1996-2016
Serial 7's	Numeracy	6 (0-5)	1996-2016
Quantitative reasoning	Numeracy	5 (0-4)	2002-2016
Retrieval fluency	Fluency	91 (0-90)	2010-2016

Table 2: Mean, standard deviation (S.D.), and correlations of cognitive scores and log wages

	Number of obs	Mean	S.D.	Correlations				
Word recall	35,747	10.31	3.14	1.000				
Serial 7's	35,859	3.94	1.45	0.311	1.000			
Quantitative reasoning	11,789	2.07	1.25	0.365	0.494	1.000		
Retrieval fluency	6,800	18.46	7.20	0.295	0.285	0.335	1.000	
Log hourly wage	20,796	2.73	0.69	0.225	0.220	0.303	0.214	1.000

Table 3: First Step Estimation (Selected Specifications)

Model	$\tau_1(\cdot)$	$\tau_2(\cdot)$	Log-likelihood	Compared	p-value
A. Both ε_t and θ_t are normal					
1	linear	linear	-18546.93		
2	linear	7th	-18500.87	2 vs. 1	0.0000
3	linear	8th	-18500.46	3 vs. 2	0.3615
4	quadratic	7th	-18498.71	4 vs. 2	0.0374
5	cubic	7th	-18497.37	5 vs. 4	0.1021
6	quadratic	8th	-18498.15	6 vs. 4	0.2914
B. ε_t is normal. θ_t is mixture of two normals.					
7	linear	linear	-18535.97		
8	linear	7th	-18487.42	8 vs. 7	0.0000
9	linear	8th	-18486.80	9 vs. 8	0.2680
10	quadratic	7th	-18487.25	10 vs. 8	0.5652
C. Both ε_t and θ_t are mixture of two normals					
11	linear	linear	-18504.07		
12	linear	7th	-18455.07	12 vs. 11	0.0000
13	linear	8th	-18454.50	13 vs. 12	0.2871
14	quadratic	7th	-18454.92	14 vs. 12	0.5891
Comparing A vs. B vs. C					
				8 vs. 4	0.0000
				12 vs. 8	0.0000

Note: η_1 and η_2 are normally distributed.

Table 4: First Step Estimated Parameters

Description	Symbol	Value	Standard Error
Skill Function			
Word recall			
	$\beta_{1,0}$	10.38	0.06
	$\beta_{1,1}$	5.50	0.49
	$\sigma^2_{\eta_1}$	6.57	0.26
Retrieval fluency			
	$\beta_{2,0}$	18.02	0.56
	$\beta_{2,1}$	16.94	4.51
	$\beta_{2,2}$	38.30	35.75
	$\beta_{2,3}$	-58.50	73.66
	$\beta_{2,4}$	-312.69	296.94
	$\beta_{2,5}$	62.05	314.11
	$\beta_{2,6}$	728.30	693.87
	$\beta_{2,7}$	426.80	513.29
	$\sigma^2_{\eta_2}$	36.43	1.66
Serial 7's			
	$\beta_{3,1}$	3.60	0.34
	$\chi_{3,1}$	-2.77	0.11
	$\chi_{3,2}$	-1.95	0.14
	$\chi_{3,3}$	-1.40	0.15
	$\chi_{3,4}$	-0.78	0.15
	$\chi_{3,5}$	-0.05	0.15
Quantitative reasoning			
	$\beta_{4,1}$	4.63	0.45
	$\chi_{4,1}$	-1.90	0.09
	$\chi_{4,2}$	-0.70	0.12
	$\chi_{4,3}$	0.53	0.13
	$\chi_{4,4}$	1.52	0.14
Skill			
	$p_{\theta_t,1}$	0.87	0.17
	$p_{\theta_t,2}$	0.13	0.17
	$\mu_{\theta_t,1}$	0.06	0.05
	$\mu_{\theta_t,2}$	-0.42	0.63
	$\sigma^2_{\theta_t,1}$	0.04	0.01
	$\sigma^2_{\theta_t,2}$	0.04	0.06
Wage shocks			
	$p_{\varepsilon_t,1}$	0.12	0.07
	$p_{\varepsilon_t,2}$	0.88	0.07
	$\mu_{\varepsilon_t,1}$	0.41	0.13
	$\mu_{\varepsilon_t,2}$	-0.05	0.04
	$\sigma^2_{\varepsilon_t,1}$	0.83	0.28
	$\sigma^2_{\varepsilon_t,2}$	0.27	0.02

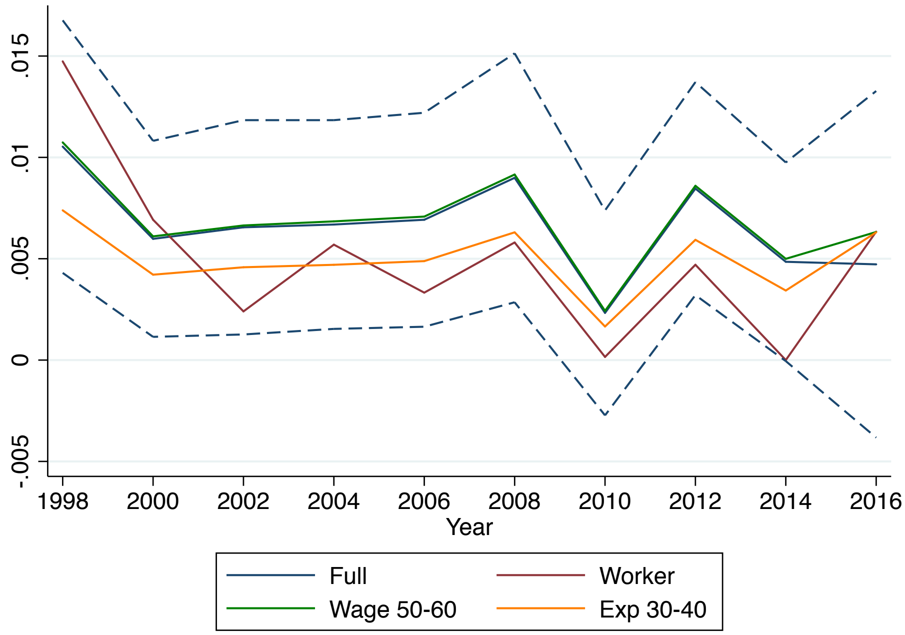
Table 5: GMM Estimation Results (Selected Parameters)

Parameter	Full	Worker	Wage 50-60	Exp 30-40
$\beta_{1,1}$	9.936	9.032	9.834	11.844
	(0.743)	(0.724)	(0.802)	(1.169)
ρ^2	0.861	0.867	0.861	0.861
	(0.037)	(0.050)	(0.037)	(0.037)
$Var(\psi)$	0.021	0.025	0.021	0.014
	(0.007)	(0.011)	(0.007)	(0.005)

Table 6: Average Estimated Skill and Log Wage Residuals by Education/Race (Ages 55-60)

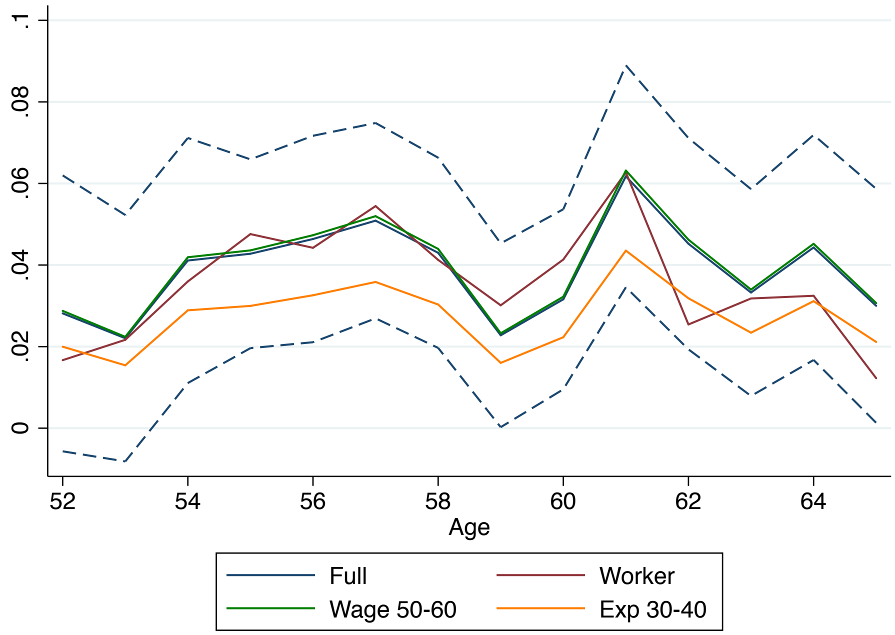
	Estimated Skill (Full)	Estimated Skill (Workers)	Log Wage Residual (Workers)
A. By Education			
Less than HS	-0.287	-0.241	-0.414
	(0.006)	(0.008)	(0.015)
HS grad	-0.147	-0.121	-0.186
	(0.005)	(0.006)	(0.010)
Some college	-0.085	-0.053	-0.035
	(0.005)	(0.006)	(0.012)
College grad	0.002	0.022	0.249
	(0.007)	(0.007)	(0.017)
Above college	0.083	0.089	0.483
	(0.006)	(0.007)	(0.017)
B. By Race			
White	-0.051	-0.022	0.105
	(0.003)	(0.004)	(0.008)
Non-white	-0.203	-0.160	-0.244
	(0.005)	(0.006)	(0.011)
Num. of obs.	14,047	9,064	9,064

Notes: Standard errors in parentheses.



Note: Dashed lines represent the 95% confidence intervals for the Full sample.

Figure 1: Estimated Skill Shocks $\sigma^2_{\nu_t} \equiv Var(\nu_t)$ by Year



Note: Dashed lines represent the 95% confidence intervals for the Full sample.

Figure 2: Estimated $Var(\phi_{a,t})$ by Age for $t = 1996$

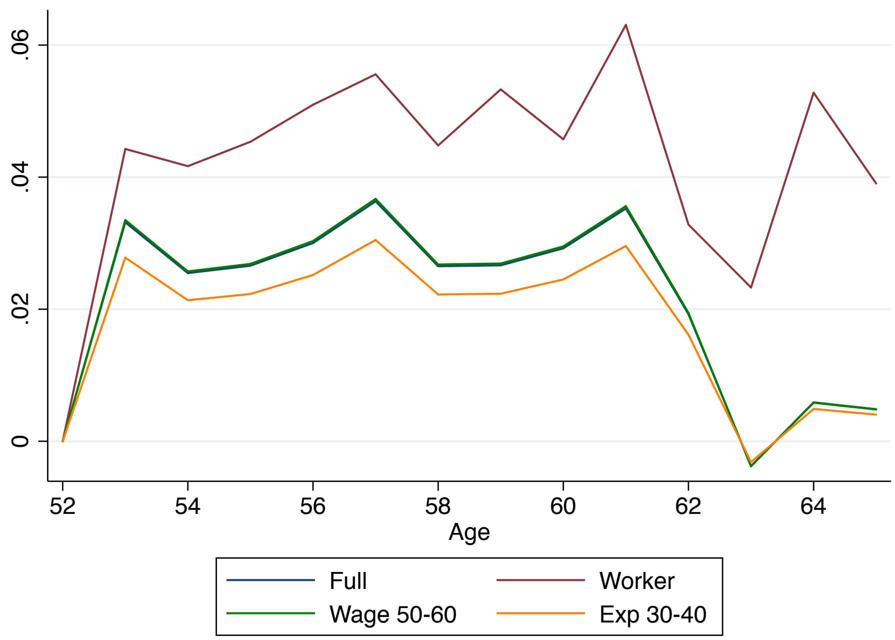
(a) $a = 52$ (b) $a = 53$

Note: Dashed lines represent the 95% confidence intervals for the Full sample.

Figure 3: Estimated $Var(\phi_{a,t})$ by Year

Note: Dashed lines represent the 95% confidence intervals for the Full sample.

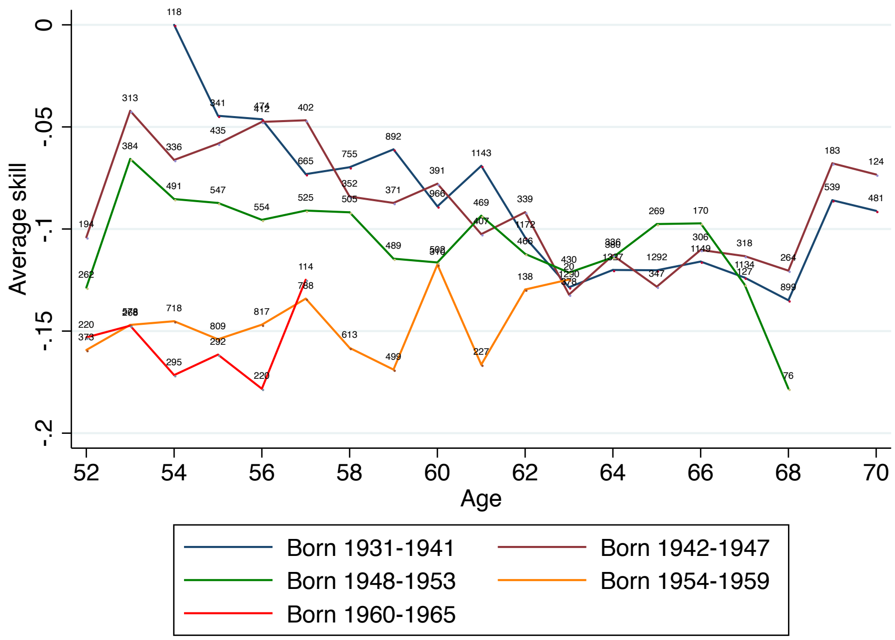
Figure 4: Estimated Return to Skill (λ_t) by Year



(a) Age effects

(b) Time Effects

Figure 5: $\widehat{\alpha}_{a,t}$ Regression Coefficients



Note: Numbers above data points reflect the number of observations.

Figure 6: Average Skill Profiles by Cohort

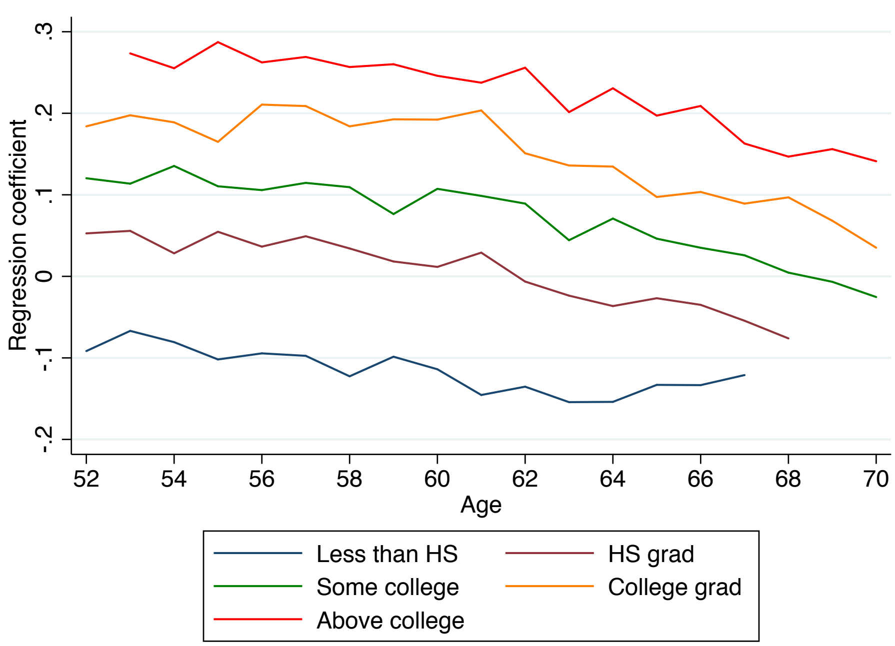


Figure 7: Estimated Average Skill Profiles by Education

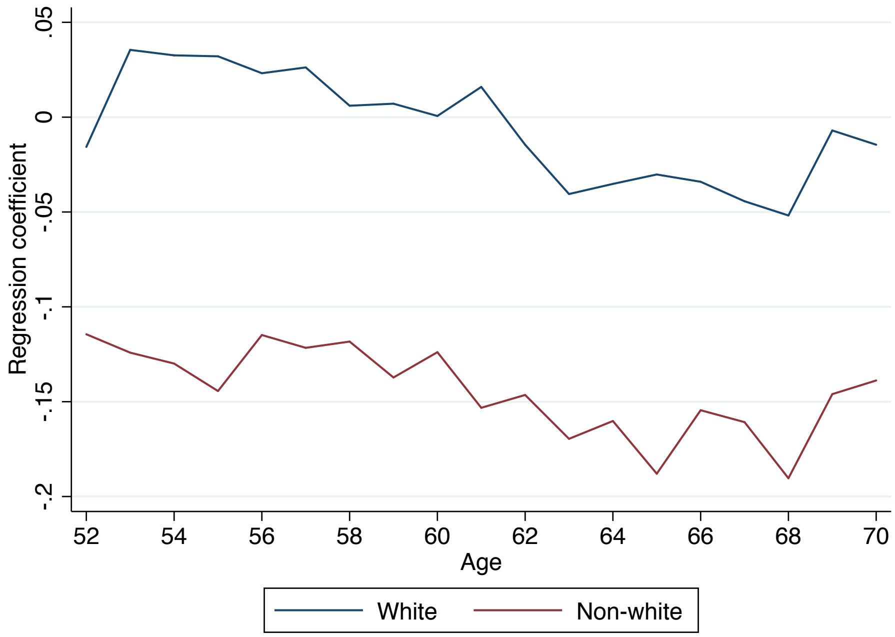
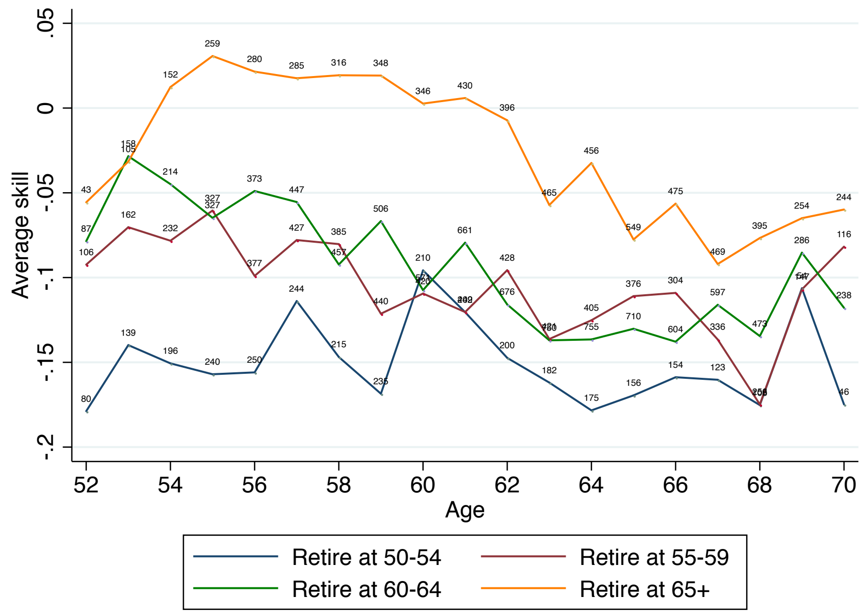
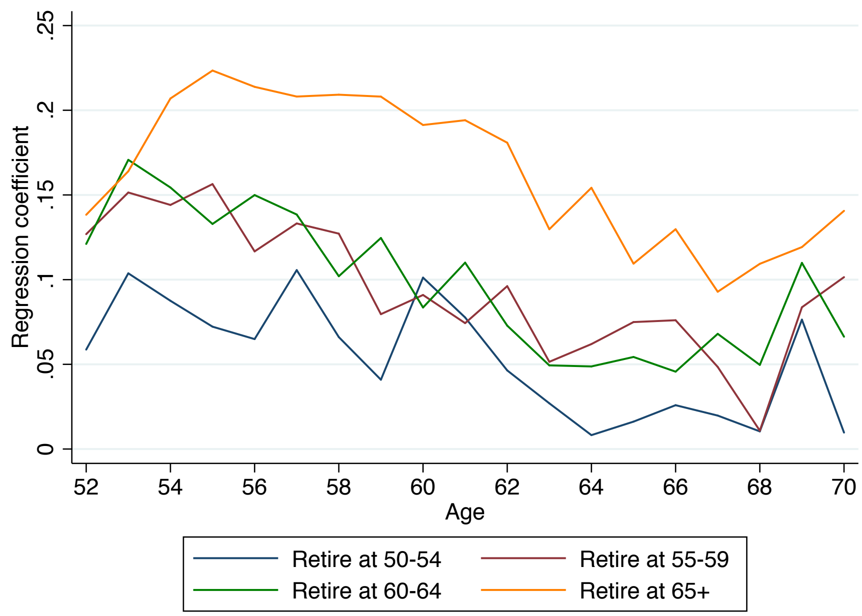


Figure 8: Estimated Average Skill Profiles by Race



Note: Numbers above data points reflect the number of observations.

(a) Average Skill



(b) Estimated Average Skill

Figure 9: Average Skill Profiles by Retirement Age

Appendix A Identification Details

In this appendix, we provide the regularity conditions for the identification result in Section 2.2. The result is based on Theorem 1 in Hu and Schennach (2008). For completeness of the arguments, we rewrite all regularity conditions using notation in the current setup. Suppose that we have two test measures, $J = 2$, which is the minimum requirement. The model at (a^*, t^*) can be rewritten as follows:

$$
\begin{aligned}
w_{i,a^*,t^*} &= \theta_{i,a^*,t^*} + \varepsilon_{i,a^*,t^*} \\
T_{i,1,a^*,t^*} &= \tau_1(\theta_{i,a^*,t^*}) + \eta_{i,1,a^*,t^*} \\
T_{i,2,a^*,t^*} &= G_2\left(\tau_2(\theta_{i,a^*,t^*}) + \eta_{i,2,a^*,t^*}\right).
\end{aligned}
$$

We collect the necessary regularity conditions below:

Assumption 1 *The observations $(w_{i,a^*,t^*}, T_{i,1,a^*,t^*}, T_{i,2,a^*,t^*})$ generated from the model above satisfy the following conditions:*

i. The joint density of $(\theta_{i,a^*,t^*}, w_{i,a^*,t^*}, T_{i,1,a^*,t^*}, T_{i,2,a^*,t^*})$ is bounded, and so are all their marginal and conditional densities. Furthermore, the joint density of $(\theta_{i,a^*,t^*}, w_{i,a^*,t^*}, T_{i,1,a^*,t^*})$ is continuous, and so are all their marginal and conditional densities.

ii. The random variables w_{i,a^*,t^*}, $T_{i,1,a^*,t^*}$, and $T_{i,2,a^*,t^*}$ are mutually independent conditional on θ_{i,a^*,t^*}.

iii. The conditional density functions $f_{w_{i,a^*,t^*}|T_{i,1,a^*,t^*}}(w|t)$ and $f_{\theta_{i,a^*,t^*}|w_{i,a^*,t^*}}(\theta|w)$ form a bounded complete family of distributions indexed by t and w, respectively.

iv. For all $\theta_1, \theta_2 \in \Theta$, the set $\{t_2 : f_{T_{i,2,a^*,t^*}|\theta_{i,a^*,t^*}}(t_2|\theta_1) \neq f_{T_{i,2,a^*,t^*}|\theta_{i,a^*,t^*}}(t_2|\theta_2)\}$ has positive probability whenever $\theta_1 \neq \theta_2$.

v. We normalize that $E[w_{i,a^*,t^*}|\theta_{i,a^*,t^*}] = \theta_{i,a^*,t^*}$ and that $E[\varepsilon_{i,a^*,t^*}|\theta_{i,a^*,t^*}] = E[\eta_{i,1,a^*,t^*}|\theta_{i,a^*,t^*}] = E[\eta_{i,2,a^*,t^*}|\theta_{i,a^*,t^*}] = 0$.

Condition (i) is a mild restriction on the distribution and allows $T_{i,2,a^*,t^*}$ to be discrete. Conditions (ii), (iv), and (v) are immediately satisfied from the model construction. For example, the strict monotonicity of $\tau_2(\cdot)$ implies condition (iv). The completeness assumption in condition (iii) is widely used in the nonparametric identification literature and is satisfied in many classes of distributions, e.g. the exponential family. See, Hu and Schennach (2008) for further discussions on the completeness assumption.

Under Assumption 1, Theorem 1 in Hu and Schennach (2008) holds, and we can identify the joint and conditional densities $f_{T_{2,a^*,t^*},\theta_{a^*,t^*}}(\cdot,\cdot)$, $f_{w_{a^*,t^*}|\theta_{a^*,t^*}}(\cdot|\cdot)$, and $f_{T_{1,a^*,t^*},|\theta_{a^*,t^*}}(\cdot|\cdot)$ from Equation (6) therein. The measurement function $\tau_1(\cdot)$ is the conditional mean function of T_{1,a^*,t^*} given θ_{a^*,t^*} and can be identified by

$$\tau_1(\theta) = E[T_{1,a^*,t^*}|\theta] = \int t_1 f_{T_{1,a^*,t^*}|\theta_{a^*,t^*}}(t_1|\theta)dt_1.$$

The marginal density $f_{\theta_{a^*,t^*}}(\cdot)$ is identified by integrating the joint density:

$$f_{\theta_{a^*,t^*}}(\theta) = \int f_{T_{2,a^*,t^*},\theta_{a^*,t^*}}(t_2,\theta)dt_2.$$

Finally, the marginal densities $f_{\varepsilon_{a^*,t^*}}(\cdot)$ and $f_{\eta_{1,a^*,t^*}}(\cdot)$ are identified by the standard deconvolution method:

$$\begin{aligned}
\varphi_{\varepsilon_{a^*,t^*}}(t) &= \varphi_{w_{a^*,t^*}}(t)/\varphi_{\theta_{a^*,t^*}}(t) \\
\varphi_{\eta_{1,a^*,t^*}}(t) &= \varphi_{T_{1,a^*,t^*}}(t)/\varphi_{\tau_1(\theta_{a^*,t^*})}(t),
\end{aligned}$$

where $\varphi_x(\cdot)$ denotes the characteristic function of x.

Appendix B Step 1 Estimation

Derivation of Equation (4):

$$
\begin{aligned}
& f_{w_{a^*,t^*},T_{1,a^*,t^*},\ldots,T_{J,a^*,t^*}}(w,T_1,\ldots,T_J) \\
&= \int_\Theta f_{w_{a^*,t^*},T_{1,a^*,t^*},\ldots,T_{J,a^*,t^*},\theta_{a^*,t^*}}(w,T_1,\ldots,T_J,\theta)d\theta \\
&= \int_\Theta f_{w_{a^*,t^*}|T_{1,a^*,t^*},\ldots,T_{J,a^*,t^*},\theta_{a^*,t^*}}(w|T_1,\ldots,T_J,\theta) \\
&\quad \times f_{T_{1,a^*,t^*}|T_{2,a^*,t^*},\ldots,T_{J,a^*,t^*},\theta_{a^*,t^*}}(T_1|T_2,\ldots,T_J,\theta)\times\cdots\times f_{T_{J,a^*,t^*}|\theta_{a^*,t^*}}(T_J|\theta)f_{\theta_{a^*,t^*}}(\theta)d\theta \\
&= \int_\Theta f_{w_{a^*,t^*}|\theta_{a^*,t^*}}(w|\theta)f_{T_{1,a^*,t^*}|\theta_{a^*,t^*}}(T_1|\theta)\times\cdots\times f_{T_{J,a^*,t^*}|\theta_{a^*,t^*}}(T_J|\theta)f_{\theta_{a^*,t^*}}(\theta)d\theta \\
&= \int_\Theta f_{\varepsilon_{a^*,t^*}|\theta_{a^*,t^*}}(\varepsilon|\theta)f_{\eta_{1,a^*,t^*}|\theta_{a^*,t^*}}(\eta_1|\theta)\times\cdots\times f_{\eta_{J,a^*,t^*}|\theta_{a^*,t^*}}(\eta_J|\theta)f_{\theta_{a^*,t^*}}(\theta)d\theta \\
&= \int_\Theta f_{\varepsilon_{a^*,t^*}}(w-\theta;\boldsymbol{\beta}_{\varepsilon_{a^*,t^*}}) \\
&\quad \times f_{\eta_{1,a^*,t^*}}(T_1-\tau_1(\theta;\boldsymbol{\beta}_{\tau_1});\boldsymbol{\beta}_{\eta_{1,a^*,t^*}})\times\cdots\times f_{\eta_{J,a^*,t^*}}(T_J-\tau_J(\theta;\boldsymbol{\beta}_{\tau_J});\boldsymbol{\beta}_{\eta_{J,a^*,t^*}})f_{\theta_{a^*,t^*}}(\theta;\boldsymbol{\beta}_{\theta_{a^*,t^*}})d\theta.
\end{aligned}
$$

Consider mixtures of normal distributions for the distributions of ε_{a^*,t^*}, η_{j,a^*,t^*} of continuous measure, and θ_{a^*,t^*}:

$$
f_{\varepsilon_{a^*,t^*}}(w-\theta;\boldsymbol{\beta}_{\varepsilon_{a^*,t^*}}) = \sum_{n_\varepsilon} p_{\varepsilon_{a^*,t^*},n_\varepsilon}\frac{1}{\sqrt{2\pi\sigma^2_{\varepsilon_{a^*,t^*},n_\varepsilon}}}\exp\left(-\frac{(w-\theta-\mu_{\varepsilon_{a^*,t^*},n_\varepsilon})^2}{2\sigma^2_{\varepsilon_{a^*,t^*},n_\varepsilon}}\right),
$$

$$
\begin{aligned}
& f_{\eta_{j,a^*,t^*}}(T_j-\tau_j(\theta;\boldsymbol{\beta}_{\tau_j});\boldsymbol{\beta}_{\eta_{j,a^*,t^*}}) \\
&= \sum_{n_{\eta_j}} p_{\eta_{j,a^*,t^*},n_{\eta_j}}\frac{1}{\sqrt{2\pi\sigma^2_{\eta_{j,a^*,t^*},n_{\eta_j}}}}\exp\left(-\frac{(T_j-\tau^j(\theta;\boldsymbol{\beta}_{\tau_j})-\mu_{\eta_{j,a^*,t^*},n_{\eta_j}})^2}{2\sigma^2_{\eta_{j,a^*,t^*},n_{\eta_j}}}\right),
\end{aligned}
$$

$$
f_{\theta_{a^*,t^*}}(\theta;\boldsymbol{\beta}_{\theta_{a^*,t^*}}) = \sum_{n_\theta} p_{\theta_{a^*,t^*},n_\theta}\frac{1}{\sqrt{2\pi\sigma^2_{\theta_{a^*,t^*},n_\theta}}}\exp\left(-\frac{(\theta-\mu_{\theta_{a^*,t^*},n_\theta})^2}{2\sigma^2_{\theta_{a^*,t^*},n_\theta}}\right),
$$

where $\boldsymbol{\beta}_x := (p_{x,n_x},\mu_{x,n_x},\sigma_{x,n_x})$ and $\sum_{n_x} p_{x,n_x} = 1$ for $x = \varepsilon_{a^*,t^*},\eta_{j,a^*,t^*}$, and θ_{a^*,t^*}.[28] Note that the location restriction (i.e. $E[x]=0$) implies that $\sum_{n_x} p_{x,n_x}\mu_{x,n_x} = 0$.

[28]For expositional purposes, we assume that the number of distributions for each random variable mixture (i.e. n_ε, n_{η_j}, and n_θ) do not vary with age and time.

If the test measure T_j is discrete, we assume that it is generated from an ordered probit model. Suppose it has K_j discrete values: $T_j \in \{1, ..., K_j\}$. We need to estimate $K_j - 1$ cutoff values for the ordered probit, i.e., $\chi_j := (\chi_{j,1}, \chi_{j,2}, ..., \chi_{j,K_j-1})$. The density function is:

$$f_{\eta_{j,a^*,t^*}}(T_j - \tau_j(\theta; \boldsymbol{\beta}_{\tau_j}); \chi_j) = \sum_{k=1}^{K_j} \mathbb{1}(T_j = k) \left[\Phi\left(\chi_{j,k} - \tau_j(\theta; \boldsymbol{\beta}_{\tau_j})\right) - \Phi\left(\chi_{j,k-1} - \tau_j(\theta; \boldsymbol{\beta}_{\tau_j})\right)\right],$$

with $\chi_{j,0} = -\infty$ and $\chi_{j,K_j} = \infty$. $\Phi(\cdot)$ is the cdf of the standard normal distribution.

Since we have 4 measures: two are continuous (T_1 and T_2) and two are discrete (T_3 and T_4), the log-likelihood for individual i at age a^* in year t^* is:

$$\begin{aligned}
\ell_{i,a^*,t^*} = \log \int_{-\infty}^{\infty} & \left[\sum_{n_\varepsilon} p_{\varepsilon_{a^*,t^*},n_\varepsilon} \frac{1}{\sqrt{2\pi\sigma^2_{\varepsilon_{a^*,t^*},n_\varepsilon}}} \exp\left(-\frac{(w_{i,a^*,t^*} - \theta - \mu_{\varepsilon_{a^*,t^*},n_\varepsilon})^2}{2\sigma^2_{\varepsilon_{a^*,t^*},n_\varepsilon}}\right)\right] \\
\times & \left[\sum_{n_{\eta_1}} p_{\eta_{1,a^*,t^*},n_{\eta_1}} \frac{1}{\sqrt{2\pi\sigma^2_{\eta_{1,a^*,t^*},n_{\eta_1}}}} \exp\left(-\frac{(T_{i,1,a^*,t^*} - \tau_1(\theta; \boldsymbol{\beta}_{\tau_1}) - \mu_{\eta_{1,a^*,t^*},n_{\eta_1}})^2}{2\sigma^2_{\eta_{1,a^*,t^*},n_{\eta_1}}}\right)\right] \\
\times & \left[\sum_{n_{\eta_2}} p_{\eta_{2,a^*,t^*},n_{\eta_2}} \frac{1}{\sqrt{2\pi\sigma^2_{\eta_{2,a^*,t^*},n_{\eta_2}}}} \exp\left(-\frac{(T_{i,2,a^*,t^*} - \tau_2(\theta; \boldsymbol{\beta}_{\tau_2}) - \mu_{\eta_{2,a^*,t^*},n_{\eta_2}})^2}{2\sigma^2_{\eta_{2,a^*,t^*},n_{\eta_2}}}\right)\right] \\
\times & \left[\sum_{k_3=1}^{K_3} \mathbb{1}(T_{i,3,a^*,t^*} = k_3) \left[\Phi\left(\chi_{3,k_3} - \tau_3(\theta; \boldsymbol{\beta}_{\tau_3})\right) - \Phi\left(\chi_{3,k_3-1} - \tau_3(\theta; \boldsymbol{\beta}_{\tau_3})\right)\right]\right] \\
\times & \left[\sum_{k_4=1}^{K_4} \mathbb{1}(T_{i,4,a^*,t^*} = k_4) \left[\Phi\left(\chi_{4,k_4} - \tau_4(\theta; \boldsymbol{\beta}_{\tau_4})\right) - \Phi\left(\chi_{4,k_4-1} - \tau_4(\theta; \boldsymbol{\beta}_{\tau_4})\right)\right]\right] \\
\times & \left[\sum_{n_\theta} p_{\theta_{a^*,t^*},n_\theta} \frac{1}{\sqrt{2\pi\sigma^2_{\theta_{a^*,t^*},n_\theta}}} \exp\left(-\frac{(\theta - \mu_{\theta_{a^*,t^*},n_\theta})^2}{2\sigma^2_{\theta_{a^*,t^*},n_\theta}}\right)\right] d\theta \\
= \log & \sum_{n_\varepsilon, n_{\eta_1}, n_{\eta_2}, k_3, k_4, n_\theta} p_{\varepsilon_{a^*,t^*},n_\varepsilon} p_{\eta_{1,a^*,t^*},n_{\eta_1}} p_{\eta_{2,a^*,t^*},n_{\eta_2}} p_{\theta_{a^*,t^*},n_\theta} \\
\times & \frac{1}{\sqrt{2^4\pi^4\left(\sigma_{\varepsilon_{a^*,t^*},n_\varepsilon}\sigma_{\eta_{1,a^*,t^*},n_{\eta_1}}\sigma_{\eta_{2,a^*,t^*},n_{\eta_2}}\sigma_{\theta_{a^*,t^*},n_\theta}\right)^2}} \\
\times & \mathbb{1}(T_{i,3,a^*,t^*} = k_3) \times \mathbb{1}(T_{i,4,a^*,t^*} = k_4) \\
\times & \int_{-\infty}^{\infty} \exp\Bigg(-\frac{(w_{i,a^*,t^*} - \theta - \mu_{\varepsilon_{a^*,t^*},n_\varepsilon})^2}{2\sigma^2_{\varepsilon_{a^*,t^*},n_\varepsilon}} - \frac{(T_{i,1,a^*,t^*} - \tau_1(\theta; \boldsymbol{\beta}_{\tau_1}) - \mu_{\eta_{1,a^*,t^*},n_{\eta_1}})^2}{2\sigma^2_{\eta_{1,a^*,t^*},n_{\eta_1}}}
\end{aligned}$$

$$
- \frac{(T_{i,2,a^*,t^*} - \tau_2(\theta;\boldsymbol{\beta}_{\tau_2}) - \mu_{\eta_{2,a^*,t^*,n_{\eta_2}}})^2}{2\sigma^2_{\eta_{2,a^*,t^*,n_{\eta_2}}}} - \frac{(\theta - \mu_{\theta_{a^*,t^*,n_\theta}})^2}{2\sigma^2_{\theta_{a^*,t^*,n_\theta}}}\Bigg)
$$

$$
\times \left[\Phi\left(\chi_{3,k_3} - \tau_3(\theta;\boldsymbol{\beta}_{\tau_3})\right) - \Phi\left(\chi_{3,k_3-1} - \tau^3(\theta;\boldsymbol{\beta}_{\tau_3})\right)\right]
$$

$$
\times \left[\Phi\left(\chi_{4,k_4} - \tau_4(\theta;\boldsymbol{\beta}_{\tau_4})\right) - \Phi\left(\chi_{4,k_4-1} - \tau^4(\theta;\boldsymbol{\beta}_{\tau_4})\right)\right] d\theta.
$$

$\boldsymbol{\beta}_{a^*,t^*}$ is estimated by maximizing the log-likelihood function:

$$
\widehat{\boldsymbol{\beta}}_{a^*,t^*} = \arg\max_{\beta_{a^*,t^*} \in \mathcal{B}} \frac{1}{|\mathcal{I}_{c^*}|} \sum_{i \in \mathcal{I}_{c^*}} \ell_{i,a^*,t^*}.
$$

Standard Errors

Define the score of the log-likelihood for observation i as follows:

$$
\widehat{\boldsymbol{S}}_{i,a^*,t^*} = \boldsymbol{S}_{i,a^*,t^*}(\widehat{\boldsymbol{\beta}}_{a^*,t^*}) = \frac{\partial \ell_{i,a^*,t^*}(\widehat{\boldsymbol{\beta}}_{a^*,t^*})}{\partial \boldsymbol{\beta}^\top_{a^*,t^*}},
$$

and the Hessian:

$$
\widehat{\boldsymbol{H}}_{i,a^*,t^*} = \boldsymbol{H}_{i,a^*,t^*}(\widehat{\boldsymbol{\beta}}_{a^*,t^*}) = \frac{\partial^2 \ell_{i,a^*,t^*}(\widehat{\boldsymbol{\beta}}_{a^*,t^*})}{\partial \boldsymbol{\beta}^\top_{a^*,t^*} \partial \boldsymbol{\beta}_{a^*,t^*}}.
$$

The asymptotic variance matrix is:

$$
\widehat{\boldsymbol{V}}_{ML,a^*,t^*} = \left(\sum_{i \in \mathcal{I}_{c^*}} \widehat{\boldsymbol{H}}_{i,a^*,t^*}\right)^{-1} \left(\sum_{i \in \mathcal{I}_{c^*}} \widehat{\boldsymbol{S}}_{i,a^*,t^*} \widehat{\boldsymbol{S}}^\top_{i,a^*,t^*}\right) \left(\sum_{i \in \mathcal{I}_{c^*}} \widehat{\boldsymbol{H}}_{i,a^*,t^*}\right)^{-1}.
$$

Appendix C GMM Estimation with a Linear Measure

Let $\boldsymbol{\Lambda}$ be a vector of parameters to be estimated in the second stage. We use the generalized methods of moments (GMM) to estimate $\boldsymbol{\Lambda}$. Suppose the total number of covariances is M and let $m = 1, ..., M$ be the index of the covariances. Define the theoretical covariance vector as $\boldsymbol{h}(\boldsymbol{\Lambda}) = (h_1(\boldsymbol{\Lambda}), ..., h_M(\boldsymbol{\Lambda}))^\top$. Let $d_{i,m}$ be the indicator of whether individual i contributes to the m^{th} covariance. Then we can write individual i's contribution to the m^{th} moment as $g_m(\boldsymbol{z_i}, \boldsymbol{\Lambda})$ where $\boldsymbol{z_i}$ includes $d_{i,m}$, individual i's log wage residuals, and cognitive measures. $g_m(\boldsymbol{z_i}, \boldsymbol{\Lambda})$ is equal to $d_{i,m}$ times the difference between the product of corresponding de-meaned variables and the theoretical covariance. For example, individual i's contribution to the moment involving covariance $Cov(w_{a,t}, T_{j,a+k,t+k})$ is $g_m(\boldsymbol{z_i}, \boldsymbol{\Lambda}) = d_{i,m}\big[(w_{i,a,t} - \bar{w}_{a,t})(T_{i,j,a+k,t+k} - \bar{T}_{j,a+k,t+k}) - h_m(\boldsymbol{\Lambda})\big]$.

Let $\boldsymbol{g}(\boldsymbol{z}, \boldsymbol{\Lambda}) = (g_1(\boldsymbol{z}, \boldsymbol{\Lambda}), ..., g_M(\boldsymbol{z}, \boldsymbol{\Lambda}))^\top$. Then the following moment condition holds at the true parameter $\boldsymbol{\Lambda}_0$:

$$\mathbb{E}\big[\boldsymbol{g}(\boldsymbol{z}, \boldsymbol{\Lambda}_0)\big] = 0.$$

The GMM estimator $\widehat{\boldsymbol{\Lambda}}$ solves

$$\min_{\boldsymbol{\Lambda}} \left[\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{g}(\boldsymbol{z_i}, \boldsymbol{\Lambda})\right]^\top \boldsymbol{W} \left[\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{g}(\boldsymbol{z_i}, \boldsymbol{\Lambda})\right],$$

where $\boldsymbol{W}$ is the weighting matrix.

Standard Errors

The GMM estimator $\widehat{\boldsymbol{\Lambda}}$ is asymptotically normal with a variance matrix

$$\boldsymbol{V}_{GMM} = (\boldsymbol{G}^\top \boldsymbol{W} \boldsymbol{G})^{-1}(\boldsymbol{G}^\top \boldsymbol{W} \boldsymbol{\Omega} \boldsymbol{W} \boldsymbol{G})(\boldsymbol{G}^\top \boldsymbol{W} \boldsymbol{G})^{-1}/N,$$

where $\boldsymbol{G}$ is the Jacobian of the vector of moments, $\mathbb{E}[\partial \boldsymbol{g}(\boldsymbol{z}, \boldsymbol{\Lambda}_0)/\partial \boldsymbol{\Lambda}_0^\top]$, and $\boldsymbol{\Omega} = \mathbb{E}[\boldsymbol{g}(\boldsymbol{z}, \boldsymbol{\Lambda}_0)\boldsymbol{g}(\boldsymbol{z}, \boldsymbol{\Lambda}_0)^\top]$. To calculate the asymptotic variance matrix, both expectations are replace by sample averages and evaluated at the estimated parameters:

$$\widehat{\boldsymbol{G}} = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial \boldsymbol{g}(\boldsymbol{z}_i, \widehat{\boldsymbol{\Lambda}})}{\partial \boldsymbol{\Lambda}^\top} = -\boldsymbol{W}^{-\frac{1}{2}} \frac{\partial \boldsymbol{h}(\widehat{\boldsymbol{\Lambda}})}{\partial \boldsymbol{\Lambda}^\top},$$

$$\widehat{\boldsymbol{\Omega}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{g}(\boldsymbol{z}_i, \widehat{\boldsymbol{\Lambda}})\boldsymbol{g}(\boldsymbol{z}_i, \widehat{\boldsymbol{\Lambda}})^\top.$$

We can test r linear parameter restrictions $H_0 : \boldsymbol{R\Lambda} = 0$ using Wald test statistic:

$$(\boldsymbol{R}\widehat{\boldsymbol{\Lambda}})^\top(\boldsymbol{R}\widehat{\boldsymbol{V}}_{\boldsymbol{GMM}}\boldsymbol{R})^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\Lambda}}) \xrightarrow{d} \chi^2_r.$$

Appendix D Data

D.1 Cognitive Measures

Details on the construction of the four cognitive measures are as follows:

- Word recall. In the data, there are two separate tasks to assess respondent's memory: one is immediate word recall and the other is delayed word recall. During the interview, the interviewer read a list of 10 nouns to the respondent and asked the respondent to recall as many words as possible from the list in any order. After approximately 5 minutes of answering other survey questions, the respondent was asked to recall the nouns previously presented. We construct a single measure which is the sum of the number of nouns that the respondent recalled in the two tasks. This measure ranges from 0 to 20.

- Serial 7's. This test asks the respondent to subtract 7 from the prior number, beginning with 100 for five trials. Correct subtractions are based on the prior number given, so that even if one subtraction is incorrect subsequent trials are evaluated on the given (perhaps wrong) answer. This test score ranges from 0 to 5.

- Quantitative reasoning. In HRS 2002, three questions were added to the core survey to assess respondents' numerical ability:

  1. "Next I would like to ask you some questions which assess how people use numbers in everyday life. If the chance of getting a disease is 10 percent, how many people out of 1,000 would be expected to get the disease?"
  2. "If 5 people all have the winning numbers in the lottery and the prize is two million dollars, how much will each of them get?"
  3. "Let's say you have $200 in a savings account. The account earns ten percent interest per year. How much would you have in the account at the end of two years?"

  We construct a single measure called quantitative reasoning using the answers from these three questions. For the first two questions, the respondent gets 1 if the answer is correct and 0 otherwise. For the last question, the respondent gets 2 if the answer is correct. If the respondent used 10% as a simple interest rate rather than a compound interest rate, i.e., answered 240 instead of 242, he gets 1. The quantitative reasoning measure is the sum of scores of all three questions and it ranges from 0 to 4.

- Retrieval fluency. This task was first incorporated in the HRS in the 2010 wave. During this task, respondents were asked to name as many animals as they could withing a 60-second time limit. The retrieval fluency measure is constructed as the number of total animal answered minus the number of incorrect names. The value of this measure ranges from 0 to 90.

D.2 Age

The age variable we use is the age at the end of the interview. According to the HRS, when there are different beginning and ending interview dates, most of the interview is usually conducted on the ending date. So it is recommended to use age at the end of interview date for respondent age at each interview.

The interval between interviews is usually 2 years. But about 5-10% of the sample was interviewed a year later than the wave year. For example, the normal case would be someone at age 52 interviewed in 1998 and age 54 interviewed in 2000. But it could be the case that he was interviewed in 2001 for the second interview at age 55. Another case could be he was aged 53 when interviewed in 1999 and aged 54 when interviewed in 2000. In these cases, we assume that age at the first interview is the age at that wave year and the subsequent interviews are two years apart. So for the first case, the wages we observe are $w_{a=52,t=1998}$ and $w_{a=55,t=2001}$ and we assume they are $w_{a=52,t=1998}$ and $w_{a=54,t=2000}$. For the second case, we observe $w_{a=53,t=1999}$ and $w_{a=54,t=2000}$ and we assume they are $w_{a=53,t=1998}$ and $w_{a=55,t=2000}$.

Another approach is to use the birth year to calculate age at each wave year. Then we would assume that we observe $w_{a=52,t=1998}$ and $w_{a=54,t=2000}$ for both the first and second cases. The results are quite similar and do not drive any particular patterns using this alternative approach.