The Efficacy of Ability Proxies for Estimating the Returns to Schooling: A Factor Model-Based Evaluation*

Mohitosh Kejriwal
Purdue University

Xiaoxiao Li
Villanova University

Evan Totty
U.S. Census Bureau

May 10, 2021

Abstract

This paper revisits the issue of estimating the returns to schooling within a framework that allows multiple unobserved skills with potentially time-varying prices where both skills and prices are possibly correlated with schooling. A common approach to addressing the problem of ability bias emanating from the unobservability of skills is to augment the earnings-schooling regression with various measures of cognitive and non-cognitive ability as proxies for these skills. While the adequacy of such measures has long been questioned, a rigorous assessment of their effectiveness in estimating the magnitude of ability bias has received relatively less attention. This paper undertakes a formal evaluation of the proxy approach by adopting a methodology that models the skills and their prices using an unobserved factor structure in which the factor loadings represent the latent skills and the common factors their associated prices. Our factor model approach allows consistent estimation of the returns to schooling without the need to rely on proxies thereby providing a flexible test bed for quantifying the contribution of the proxies to the aggregate least squares bias. The factor model estimators may also be viewed as implicitly estimating the measurement error inherent in the proxies. Our empirical results using panel data from the NLSY79 show that the estimated ability bias lies between 39-51% for our most general specification. A bias decomposition analysis indicates that the commonly used proxies can explain very little of the estimated bias. Direct tests for the viability of the proxies further confirm their inadequacy in capturing the underlying latent skills.

Keywords: returns to schooling, ability proxies, bias decomposition, factor model, interactive fixed effects, latent skills

JEL Classification: C33, C38, I26

---

*We are grateful to Timothy Bond, Kevin Lang, Debasri Mukherjee, Susan Pozo, and Miguel Sarzosa for useful comments. The opinions in this paper are those of the authors and may not represent the views of the U.S. Census Bureau or other organizations.
Contact information: Mohitosh Kejriwal (mkejriwa@purdue.edu), Xiaoxiao Li (xiaoxiao.s.li@villanova.edu), Evan Totty (evan.scott.totty@gmail.com).

1 Introduction

Estimating the returns to schooling has been an issue of long-standing interest in labor economics and the economics of education. Understanding the causal relationship between earnings and education is crucial for determining the effect of policy interventions. The traditional workhorse in empirical analyses of earnings determination has been the so-called Mincer regression (Mincer, 1974) that relates the log of earnings to years of schooling and a quadratic function of labor market experience. The challenge to obtaining an unbiased estimate of the returns to schooling arises from the fact that unobserved ability is correlated with both earnings and schooling so that the ordinary least squares (OLS) estimator is subject to an omitted variable bias. Under the assumption that the correlations are both positive, OLS leads to an upward biased estimate of the returns to schooling. This issue of ability bias has spawned the development of a vast literature that has employed a variety of econometric techniques to control for this bias thereby producing a wide range of estimates which are often in conflict with each other. The more commonly adopted techniques include instrumental variables, utilizing within family variation in schooling, and the use of observable proxies for ability [see Card (2001), Heckman et al. (2006) and Caplan (2018) for comprehensive reviews of this literature]. This paper focuses on the third approach, henceforth referred to as the ability proxy approach.

The ability proxy approach attempts to control for ability bias by augmenting the regression model with various measures of cognitive and non-cognitive abilities. Cognitive abilities are typically measured using IQ, GPA, scores on the Armed Forces Qualification Test (AFQT) or individual/principal components of the Armed Services Vocational Aptitude Battery (ASVAB) test. Non-cognitive abilities are typically measured using the Rotter Locus of Control Scale, the Rosenberg Self-Esteem Scale or their principal components. Heckman et al. (2018) provide a comparison of standard OLS estimates to estimates controlling for ability proxies using Bartlett cognitive and non-cognitive factors, and find that the latter are about 20-50% smaller, depending on the specification. Ashworth et al. (2020) report similar reductions in comparing the basic Mincer regressions to regressions that include ability proxies and actual experience using panel data. Caplan (2018) suggests informal bounds on the degree of ability bias based on a detailed survey of the literature that employs ability proxies: cognitive and non-cognitive ability bias ranges between 20-30% and 5-15% respectively, with overall ability bias between 25-45%.

The success of the ability proxy approach in delivering a reliable estimate of the returns to schooling depends critically on the extent to which these proxies represent the underlying true unobserved abilities. A major criticism of this approach is that the proxies, particularly those measuring non-cognitive ability or "soft skills" such as conscientiousness, conformity, self-esteem, etc., are far from perfect as they are likely affected by family background characteristics (Carneiro et al., 2003; Heckman et al., 2006). In fact, even cognitive measures such as AFQT scores are not without their critics; for instance, Das and Polachek (2019) argue that AFQT scores reflect analytic academic capabilities but do not get at facets of ability that matter with respect to real-world accomplishments. Griliches (1977) shows that mismeasured ability variables lead to biased estimates and proposes controlling for measurement error by instrumenting for schooling and test scores using family background variables. Such an approach, however, relies on the assumption being that the family variables affect earnings only via schooling or test scores. A further potential weakness of the proxy approach is that it is predicated on the assumption that education does not enhance ability, i.e., no reverse causation (Hansen et al., 2004). If this assumption fails, correcting for ability leads to misleadingly low estimates of the schooling effect and hence overstates the severity of ability bias.[1] This issue can be addressed by measuring ability and then estimating the

[1] See, e.g., Belzil and Hansen (2002), Hansen et al. (2004), Williams (2019).

effect of subsequent education on earnings.[2]

Despite the aforementioned limitations, a rigorous assessment of the effectiveness of the ability proxy approach in estimating the magnitude of ability bias has received little attention. Our paper attempts to fill this gap. We undertake a formal evaluation of this approach by adopting a methodology that models the skills and their prices using an unobserved factor structure in which the factor loadings represent the latent skills and the common factors their associated prices. Skills and prices are both allowed to be correlated with schooling. The factor model (FM, henceforth) approach, developed by Pesaran (2006) and Bai (2009), exploits the large panel nature of the dataset and the within individual variation in earnings and schooling to achieve identification and thus facilitates consistent estimation of the returns to schooling without the need to rely on test score proxies. The approach therefore provides a flexible test bed for quantifying the contribution of the proxies to the aggregate least squares bias. Bias decomposition methods are used to ascertain the proportion of bias that can be explained by each of the proxies or their principal components. Using the estimated factor loadings, we also conduct a battery of direct tests for the viability of the proxies. While these tests have typically been used to evaluate observable proxies for the common factors in macroeconomic/financial applications (Bai and Ng, 2006), we appear to be the first (to our knowledge) to apply them for the purpose of evaluating proxies for the factor loadings. Finally, we formally connect the FM and ability proxy approaches by showing that the FM estimators may be viewed as implicitly estimating the measurement error inherent in the proxies.

Our empirical results using panel data from the National Longitudinal Survey of Youth (NLSY79) show that the OLS point estimate of the returns to schooling is about 8% and lies within the range typically reported in the literature [see, e.g., Gunderson and Oreopolous

[2] Research on cognitive ability bias find little evidence favoring reverse causation when IQ or AFQT and related tests are used to measure cognitive ability. With regard to non-cognitive abilities, the evidence is mixed [see Caplan (2018) for further discussion and references].

(2020)]. The FM estimates based on our most general specification are discernibly smaller, between 3.9-4.9%, depending on the particular FM estimator employed so that the estimated OLS bias lies between 39-51%. A bias decomposition analysis indicates that, conditional on the factor structure, the commonly used ability proxies can explain very little of the estimated bias. Tests for directly evaluating the efficacy of the proxies further confirm the inadequacy of the proxies in capturing the underlying latent skills.

The FM approach to estimating the effect of schooling on earnings was recently employed by Westerlund and Petrova (2018) and Kejriwal et al. (2020) (KLT, henceforth). Westerlund and Petrova (2018) apply the FM approach to estimate the returns to schooling using data from the NLSY79 and find smaller returns than OLS but do not provide an evaluation of the proxy approach. KLT employ a unique panel dataset on education and earnings over the period 1978-2011 based on respondents from the Survey of Income and Program Participation (SIPP) linked with tax and benefit data from the Internal Revenue Service (IRS) and Social Security Administration (SSA). Their FM estimates of the returns to schooling are considerably smaller than the OLS estimates, a finding consistent with the results of the current paper.[3] Their analysis, however, does not consider the ability proxy approach given that the SIPP does not provide individual measures of ability. Accordingly, the objective of our paper is to compare and contrast the estimates obtained using the OLS, proxy and FM approaches with data from the NLSY79 and thereby directly assess the effectiveness of the proxies. Another important difference with the analysis in KLT is that they restrict their sample to individuals with positive earnings in each year (i.e., a balanced panel) while we allow for individuals with missing data in our sample. Consequently, the methods we apply are modifications of those employed in KLT to deal with unbalanced panels.

The rest of the paper is organized as follows. Section 2 presents the empirical framework

[3]Their preferred specification yields a point estimate of the average marginal returns to schooling of about 2.7% relative to OLS and 2SLS estimates which lie in the range 10.7-44.4%.

and describes the econometric methods employed. Section 3 details the data used and the associated summary statistics. Section 4 presents the empirical results. Section 5 concludes. Supplementary materials (not for publication) are included in Appendices A-C.

2 Empirical Framework

This section lays out the factor model framework and provides a description of the various econometric techniques that are used in the empirical analysis. Section 2.1 details the two FM estimators, namely, the interactive fixed effects and common correlated effects estimators. It also provides a formal connection between the ability proxy and FM approaches via a measurement error interpretation of the FM estimators. Section 2.2 presents a decomposition, originally developed by Gelbach (2016) in the context of the linear regression model, of the bias incurred by the OLS estimator into the components explained by the proxies and the factor structure in order to assess their relative contributions. Section 2.3 describes a set of procedures designed to test the validity of the ability proxies based on the methods developed by Bai and Ng (2006).

2.1 The Factor Model And Related Estimators

The unbalanced panel data factor model is specified as

$$y_{it} = c_i + s_{it}\beta_1 + s_{it}^2\beta_2 + e_{it}\gamma_1 + e_{it}^2\gamma_2 + s_{it}e_{it}\gamma_3 + p_i'\delta_t + v_{it} \tag{1}$$

$$v_{it} = \lambda_i' f_t + u_{it} \tag{2}$$

$$i = 1, ..., N; \quad t = t_i \in \mathcal{J}_i \equiv \{t_i(1), t_i(2), ..., t_i(T_i)\}$$

where y_{it} and s_{it} represent, respectively, the (log of) hourly wage and the years of schooling completed for person i at period t, p_i is a $(k \times 1)$ vector of ability proxies with potentially time-varying returns δ_t, and e_{it} is a measure of labor market experience.[4] The set $\mathcal{J}_i$ includes the time indices of the non-missing observations for person i. For each i, there are T_i observations available at times $\{t_i(1), t_i(2), ..., t_i(T_i)\}$ and T_i can be different across i. Let T denote the length of the complete time period, i.e., $T = \max_{i \leq N}\{t_i(T_i)\}$.

The error term v_{it} has an interactive fixed effects structure and is composed of a common component ($\lambda_i' f_t$) and an idiosyncratic component (u_{it}). Here λ_i represents a ($r \times 1$) vector of unmeasured skills (factor loadings), such as innate abilities, while f_t is a ($r \times 1$) vector of unobserved, possibly time-varying, prices (or common factors) of the unmeasured skills.[5] Both the loadings and factors are allowed to be correlated with schooling so that the same skills and prices that influence earnings may also influence schooling.[6] The number of common components r is assumed unknown. Note that while the returns to the skill components ($\lambda_i' f_t$) are identified, the skills and their prices are not separately identified.[7] The estimated factors and loadings thus only estimate a rotation of the underlying true parameters and so cannot be given a direct economic interpretation.[8]

---

[4] As in KLT, we use age and its square to measure actual experience. An alternative measure that uses potential experience defined as (age-schooling-6) is also common in this literature [e.g., Angrist and Newey (1991)]. This measure is not appropriate in our context given that individuals may simultaneously work and go to school in our sample [see also Ginther (2000) for further arguments favoring age over potential experience].

[5] While we refer to the factor loadings as skills/abilities, there are other time-invariant determinants with possibly time-varying prices, such as motivation and persistence, that can be captured by the factors loadings as well.

[6] More formally, the factors and factor loadings from equation (2) may also appear in a model for the independent variables. The two estimators we describe in sections 2.1.1 and 2.1.2 make somewhat different assumptions about the model for the independent variables. Bai (2009) allows the loadings (λ_i) alone, the factors (f_t) alone, or their interaction ($\lambda_i' f_t$) to enter the model for schooling. Pesaran (2006) allows the same factors to enter the equation for schooling with possibly different loadings from those that enter the equation for earnings.

[7] For an arbitrary ($r \times r$) invertible matrix A, we have $F\Lambda' = FAA^{-1}\Lambda' = F^*\Lambda^{*\prime}$, so that a model with common factors $F = (f_1, ..., f_T)'$ and loadings $\Lambda = (\lambda_1, ..., \lambda_N)'$ is observationally equivalent to a model with factors $F^* = (f_1^*, ..., f_T^*)'$ and loadings $\Lambda^* = (\lambda_1^*, ..., \lambda_N^*)'$ where $F^* = FA$ and $\Lambda^* = \Lambda A^{-1\prime}$.

[8] We are following Bai (2009) and Pesaran (2006), whose FM estimators we will describe below, in referring to skills/abilities as factor loadings and the common time shocks or prices as factors. Heckman et al. (2006) and others in the returns to schooling literature also estimate what they refer to as a factor model, although the methods are not the same. Importantly, they refer to the skills/abilities as the factors rather than the factor loadings. We provide more discussion on the two methods in Section 4.5.

A set of person fixed effects c_i is included to control for time-invariant person characteristics such as gender, race, mother's and father's education, and the number of siblings. We also consider a variant of (1)-(2) that includes a set of demographics-by-year fixed effects. Such a specification allows us to investigate the extent to which the factor structure in (2) can be interpreted in terms of time varying returns to observable time-invariant characteristics (see Section 4).

The quadratic schooling specification has been routinely adopted in empirical studies; see, e.g., Lemieux (2006). The concavity of log earnings as a function of years of schooling arises in a simple human capital investment model in which individuals have different preferences (discount rates) but all face the same concave production function (the return to a year of schooling declines as years of schooling increase). Mincer (1996) shows that in a model where individuals have heterogenous preferences and earnings opportunities, average log earnings may either be a convex or a concave function of years of schooling. The importance of allowing for interaction between schooling and experience is underscored in Heckman et al. (2006) who show that log earnings-experience profiles are not parallel across schooling levels. The parameter of interest in our analysis is the marginal returns to schooling (MRTS) evaluated at the overall mean (across individuals and time) which, from (1), is given by

$$\begin{aligned} MRTS &= \beta_1 + 2\bar{s}\beta_2 + \bar{e}\gamma_3 \qquad (3) \\ \bar{s} &= N^{-1}\textstyle\sum_{i=1}^{N} T_i^{-1} \sum_{t \in \mathcal{J}_i} s_{it} \\ \bar{e} &= N^{-1}\textstyle\sum_{i=1}^{N} T_i^{-1} \sum_{t \in \mathcal{J}_i} e_{it} \end{aligned}$$

It is important to emphasize that unlike Heckman et al. (2006), our paper does not attempt to distinguish between the role of cognitive and non-cognitive skills in explaining the behavior of earnings. Rather, we are interested in estimating the average rate of growth of earnings with schooling given by (3) employing the interactive fixed effects structure as a device to control for the different components of ability that may affect earnings and are potentially correlated with schooling. Further, unlike KLT, we do not consider individual-level regressions that allow the estimation of individual-specific schooling coefficients since our analysis includes the ability proxies which are individual-specific with possibly time-varying returns.[9] It is also useful to note that the validity of the FM estimators discussed hereafter depends on the assumption of a common factor structure where each factor makes a nontrivial contribution to the variance of v_{it}.

2.1.1 The Interactive Fixed Effects (IFE) Estimator

The IFE approach to estimating the factor model (1)-(2) was originally proposed by Bai (2009) for the balanced panel framework. In this approach, the regression coefficients and the factor structure are jointly estimated using an iterative principal components algorithm. The extension to unbalanced panels was considered in Bai et al. (2015). They allow for various patterns of missing data including block missing where a group of individuals can join or drop out of the sample at a given time period, regular missing where the missing event occurs at the same time frequency for all the individuals, and random missing where some of the data are randomly missing without any obvious pattern. Bai et al. (2015) develop an estimation procedure based on adapting the EM algorithm. The algorithm consists of two loops where the inner loop carries out the EM, while the outer loop estimates the regression coefficients. In what follows, we first discuss the steps of their EM algorithm and then describe the iterations of the full algorithm.

Let $x_{it} = (s_{it}, s_{it}^2, e_{it}, e_{it}^2, s_{it}e_{it}, p_i'1(t=1), ..., p_i'1(t=T))'$, $\alpha = (\beta_1, \beta_2, \gamma_1, \gamma_2, \gamma_3, \delta_1, ..., \delta_T)'$.

[9] For instance, Altonji and Pierret (2001), Cunha et al. (2011), and Lin et al. (2018) find that returns to cognitive skills increase with age.

Then the model in (1)-(2) can be written as

$$y_{it} = c_i + x'_{it}\alpha + v_{it} = c_i + x'_{it}\alpha + \lambda'_i f_t + u_{it} \tag{4}$$

First assume that $\{c_i\}$ and α are known and define $y^{\alpha}_{it} = y_{it} - c_i - x'_{it}\alpha$. The E-step replaces the missing observations in y^{α}_{it} with an estimate of $\lambda'_i f_t$ to "complete" the dataset. The complete dataset then consists of observations $\{\breve{y}^{\alpha}_{it}\}$, where

$$\breve{y}^{\alpha}_{it} = \begin{cases} y_{it} - c_i - x'_{it}\alpha & \text{if not missing} \\ \hat{\lambda}'_i \hat{f}_t & \text{if missing} \end{cases} \tag{5}$$

given some estimator $\hat{\lambda}'_i \hat{f}_t$. The M-step entails conducting a standard principal components analysis (PCA) on the complete dataset to obtain updated estimates of $\hat{\lambda}'_i \hat{f}_t$. These estimates are in turn used to update $\breve{y}^{\alpha}_{it}$ in (5). The EM algorithm proceeds by iterating on the E and M steps until convergence.

The full algorithm can now be described by the following set of iterations:

1. Obtain an initial estimate of $\lambda'_i f_t$ using the one-step CCEP estimator described in Section 2.1.2.[10]

2. Estimate α by regressing $\widetilde{y}_{it} - \hat{\lambda}'_i \hat{f}_t$ on $\widetilde{x}_{it}$, where $\widetilde{y}_{it} = y_{it} - \bar{y}_i$, $\widetilde{x}_{it} = x_{it} - \bar{x}_i$ with $\bar{y}_i = T_i^{-1} \sum_{j=1}^{T_i} y_{i,t_i(j)}$ and $\bar{x}_i = T_i^{-1} \sum_{j=1}^{T_i} x_{i,t_i(j)}$. Denote this estimate $\hat{\alpha}$.

3. Apply the EM algorithm as described above to $\widetilde{y}_{it} - \widetilde{x}'_{it}\hat{\alpha}$ and update the estimate of $\lambda'_i f_t$.

[10] While it is not necessary to use the one-step CCE estimator to obtain the initial values, it is computationally convenient. An alternative method to generate initial values is to use a functional PCA algorithm as in Bai et al. (2015).

4. Iterate (2)-(3) until convergence.

To estimate the number of factors r, we implement the above algorithm for one to ten factors and choose as our estimate $\hat{r}$ the number that minimizes $\left|\widehat{MRTS}(s+1) - \widehat{MRTS}(s)\right|$ for $s = 0, ..., 9$, where $\widehat{MRTS}(s)$ is the estimate of $MRTS$ assuming s factors.[11] The confidence interval for the estimator is constructed using a wild cluster bootstrap procedure described in Appendix A.

To evaluate the finite sample performance of their estimator, Bai et al. (2015) conduct Monte Carlo simulations under block, regular and random missing patterns in the data generating process. They show that their proposed estimator performs well in terms of both bias and variance regardless of whether the common factors are smooth functions of time or stochastic and non-smooth.

2.1.2 The Common Correlated Effects (CCE) Estimator

An alternative approach to estimation proposed by Pesaran (2006) treats the factors as nuisance parameters rather than parameters of interest. In this approach, the common factors f_t are proxied for using the cross-sectional averages of the dependent and independent variables. The regression in (1) is augmented using these cross-sectional averages (with their coefficients allowed to be individual-specific) and estimated using OLS. This estimator, which does not require knowledge of the number of factors, is referred to as the common correlated effects pooled (CCEP) estimator. While Pesaran (2006)'s original proposal was developed in the balanced panel framework, it was subsequently extended to unbalanced panels by Zhou and Zhang (2016). To describe this estimator, let $z_{it} = (y_{it}, s_{it}, s_{it}^2, s_{it}e_{it})'$ and $\bar{z}_t = N_t^{-1}\sum_{i=1}^{N} z_{it}I_{it}$ with $N_t = \sum_{i=1}^{N} 1(t \in \mathcal{J}_i)$ and $I_{it} = 1(t \in \mathcal{J}_i)$. Further, let $y_{i,T_i} =$

[11] This approach to factor selection is adopted from Kim and Oka (2014). We consider this to be a reasonable approach in our context since, in contrast to statistical information criteria based on overall goodness of fit (Bai and Ng, 2002), it is tailored to the parameter of interest.

$(y_{it_i(1)}, ..., y_{it_i(T_i)})'$, $X_{i,T_i} = (x_{it_i(1)}, ..., x_{it_i(T_i)})'$, $\bar{Z}_{T_i} = (\bar{z}_1, ..., \bar{z}_{T_i})'$ and $\bar{H}_{T_i} = (1_{T_i}, \bar{Z}_{T_i})$ with 1_{T_i} a $(T_i \times 1)$ vector of ones. Then the unbalanced panel one-step CCEP estimator can be expressed as

$$\hat{\alpha}_{ccep} = \left(\sum_{i=1}^{N} X'_{i,T_i} M_{H_i} X_{i,T_i}\right)^{-1} \left(\sum_{i=1}^{N} X'_{i,T_i} M_{H_i} y_{i,T_i}\right)$$

where $M_{H_i} = I_{T_i} - \bar{H}_{T_i} \left(\bar{H}'_{T_i}\bar{H}_{T_i}\right)^{-1} \bar{H}'_{T_i}$ and I_{T_i} is the identity matrix of order T_i. Assuming that T is fixed and observations are randomly missing, Zhou and Zhang (2016) show that $\hat{\alpha}_{ccep}$ is $\sqrt{N}$-consistent and asymptotically normal and derive a formula for its asymptotic variance and the corresponding estimated standard error.

In addition to the one-step estimator, we also consider a version of the two-step estimator proposed by Pesaran (2006) in the balanced panel case that involves combining the CCE and principal component approaches. In the first step, we obtain the estimate $\hat{\alpha}_{ccep}$ and construct the complete dataset in a manner similar to (5):

$$\breve{y}_{it} = \begin{cases} \widetilde{y}_{it} - \widetilde{x}'_{it}\hat{\alpha}_{ccep} & \text{if not missing} \\ \widehat{\lambda'_i f_t} & \text{if missing} \end{cases}$$

where $\widehat{\lambda'_i f_t}$ is an estimate of $\lambda'_i f_t$ obtained from the one-step CCE estimation procedure using the cross-sectional averages $\bar{z}_t$. In the second step, the EM algorithm described in Section 2.1.1 is applied to $\breve{y}_{it}$ and obtain updated estimates $\{\hat{\lambda}_i, \hat{f}_t \ ; \ i = 1, ...N, \ t = 1, ..., T\}$. The number of factors is set to four in accordance with a rank condition that is required to validate the CCE approach (see discussion below). The estimates $\hat{f}_t$ are then used in the second step as regressors instead of the cross-sectional averages:

$$y_{it} = c_i + x'_{it}\alpha + \lambda'_i\hat{f}_t + \xi_{it} \tag{6}$$

Estimating (6) by least squares yields the two-step CCE estimator $\hat{\alpha}_{ccep\text{-}2}$. The confidence interval for this estimator is estimated using a wild bootstrap procedure in a manner analogous to that for the IFE estimator in Section 2.1.1 (See Appendix A).

Both the one-step and two-step CCE estimators can be sensitive to a particular rank condition which requires that the total number of factors does not exceed the total number of observed variables. While the original analysis of Pesaran (2006) suggested that the rank condition could be relaxed, recent work by Westerlund and Urbain (2013) reinstated its importance by showing that without this condition, the validity of the CCE approach hinges on the assumption that the factor loadings of the dependent variable and the regressors are uncorrelated. The relevance of this rank condition can explain potential discrepancies between the CCE and IFE estimators in practice given that the latter does not require this condition to be satisfied.[12]

2.1.3 A Measurement Error Interpretation of the Factor Model Estimators

We now provide a formal connection between the FM and ability proxy approaches. In particular, we show that the FM estimators can be viewed as correcting the measurement error in the proxies by implicitly estimating this error. To make this connection transparent, suppose that the true (population) data generating process (omitting the experience terms for simplicity) is given by,

$$y_{it} = c_i + s_{it}\beta_1 + s_{it}^2\beta_2 + \lambda_i' f_t + u_{it} \tag{7}$$

[12]The rank condition is potentially very relevant in our context, given that our empirical analysis includes the cross-sectional averages of only four variables: $y_{it}, s_{it}, s_{it}^2, s_{it}e_{it}$. Thus, the rank condition implies $r \leq 4$. Note that we only include cross-sectional averages of observables that involve earnings and schooling, given that the age controls (our measure of experience) are equivalent to the inclusion of a deterministic time trend (see Appendix B of KLT for details).

Suppose the $(k \times 1)$ vector of observed proxies is generated as

$$p_i = A'\lambda_i + \eta_i = p_i^* + \eta_i \tag{8}$$

where A is a $(r \times k)$ matrix of coefficients; p_i^* is the true (unobserved) value of the proxies, and η_i is a classical measurement error assumed to be uncorrelated with p_i^*. From (8),

$$p_i'A' = \lambda_i'AA' + \eta_i'A'$$

Assuming AA' is full rank [i.e., $k \geq r$],

$$\begin{aligned} p_i'A'(AA')^{-1} &= \lambda_i' + \eta_i'A'(AA')^{-1} \\ p_i'B &= \lambda_i' + \eta_i'B \end{aligned}$$

where $B = A'(AA')^{-1}$, which implies

$$\lambda_i' = (p_i' - \eta_i')B \tag{9}$$

Substitute (9) in (7),

$$\begin{aligned} y_{it} &= c_i + s_{it}\beta_1 + s_{it}^2\beta_2 + p_i'Bf_t + \{u_{it} - \eta_i'Bf_t\} \\ &= c_i + s_{it}\beta_1 + s_{it}^2\beta_2 + p_i'\breve{\delta}_t + \breve{u}_{it} \end{aligned} \tag{10}$$

where $\breve{\delta}_t = Bf_t$ and $\breve{u}_{it} = u_{it} - \eta_i'Bf_t$. Estimating (10) by least squares would then lead to biased estimates of the regression coefficients due to the correlation between p_i and η_i. In the FM approach (specifically, the IFE and 2 step CCE estimators), $(p_i' - \eta_i')B$ is effectively replaced by the factor loading estimate $\hat{\lambda}_i$, i.e., the FM method implicitly estimates the measurement error inherent in the proxies. Thus, the FM methods estimate the regression coefficients from

$$y_{it} = c_i + s_{it}\beta_1 + s_{it}^2\beta_2 + \hat{\lambda}_i'\hat{f}_t + \left(\lambda_i' f_t - \hat{\lambda}_i'\hat{f}_t\right) + u_{it}$$

where the sampling error involved in estimating the measurement error is reflected in the term $\left(\lambda_i' f_t - \hat{\lambda}_i'\hat{f}_t\right)$. As shown in Bai (2009), this error is negligible as the cross-sectional and time series dimensions become large, which guarantees consistent estimates for (β_1, β_2).

Another important implication that follows from (10) is that the coefficients β_1 and β_2 can be consistently estimated from a specification that includes both proxies and the interactive fixed effects even if the proxies are subject to measurement error since the error will simply be controlled for by the factor structure. So this general specification allows us to directly quantify how useful the proxies are in terms of explaining the OLS bias or, equivalently, how large is the measurement error in each proxy (see section 2.2).

2.2 Bias Decomposition

In order to quantify the relative contribution of the proxies and the factor structure to the aggregate OLS bias, we apply a bias decomposition due to Gelbach (2016). This decomposition is useful from a practical standpoint since it does not rely on the sequence in which covariates are added to the model. Sequence sensitivity can lead to different bias estimates for different sequences when added covariates are intercorrelated.

To begin with, let $S_{it} = (s_{it}, s_{it}^2)'$, $\widetilde{S}_{it} = S_{it} - \bar{S}_i$, $\widetilde{S}_{i,T_i} = (\widetilde{S}_{it_i(1)}, ..., \widetilde{S}_{it_i(T_i)})'$. Then, consider the following model (again, omitting experience terms to simplify the exposition) estimated by least squares:

$$y_{it} = c_i + s_{it}\beta_1 + s_{it}^2\beta_2 + v_{it} = c_i + S_{it}'\beta + v_{it}, \quad \beta = (\beta_1, \beta_2)' \tag{11}$$

Denote the OLS estimate from (11) by $\hat{\beta}_{OLS}$. Next, estimate the following model using one of the FM methods discussed above:

$$y_{it} = c_i + S_{it}'\beta + p_i'\delta_t + \lambda_i' f_t + u_{it}$$

Denote the FM estimate of β by $\hat{\beta}_{FM}$, the estimates of λ_i and f_t by $\hat{\lambda}_i$ and $\hat{f}_t$, respectively, and the estimate of δ_t by $\hat{\delta}_t$. Then, following Gelbach (2016), we can write

$$\begin{aligned}
\hat{\beta}_{OLS} &= \left(\sum_{i=1}^{N}\sum_{t\in\mathcal{J}_i}\widetilde{S}_{it}\widetilde{S}_{it}'\right)^{-1}\sum_{i=1}^{N}\sum_{t\in\mathcal{J}_i}\widetilde{S}_{it}\widetilde{y}_{it} = \left(\sum_{i=1}^{N}\sum_{t\in\mathcal{J}_i}\widetilde{S}_{it}\widetilde{S}_{it}'\right)^{-1}\sum_{i=1}^{N}\sum_{t\in\mathcal{J}_i}\widetilde{S}_{it}\left(\widetilde{S}_{it}'\hat{\beta}_{FM} + p_i'\hat{\delta}_t + \hat{\lambda}_i'\hat{f}_t\right) \\
&= \hat{\beta}_{FM} + \left(\sum_{i=1}^{N}\sum_{t\in\mathcal{J}_i}\widetilde{S}_{it}\widetilde{S}_{it}'\right)^{-1}\sum_{i=1}^{N}\sum_{t\in\mathcal{J}_i}\widetilde{S}_{it}p_i'\hat{\delta}_t + \left(\sum_{i=1}^{N}\sum_{t\in\mathcal{J}_i}\widetilde{S}_{it}\widetilde{S}_{it}'\right)^{-1}\sum_{i=1}^{N}\sum_{t\in\mathcal{J}_i}\widetilde{S}_{it}\hat{\lambda}_i'\hat{f}_t \quad (12)
\end{aligned}$$

Denoting the second and third term in (12) by C_1 and C_2, respectively, the estimated aggregate OLS bias can be expressed as

$$C = \hat{\beta}_{OLS} - \hat{\beta}_{FM} = C_1 + C_2 \qquad (13)$$

In (13), C_1 is the component of the bias explained by the proxies and C_2 is the component explained by the factor structure. Denoting $p_i = (p_{i1}, ..., p_{ik})'$, $\hat{\delta}_t = (\hat{\delta}_{t1}, ..., \hat{\delta}_{tk})'$, we can further decompose C_1 as

$$C_1 = \left(\sum_{i=1}^{N}\sum_{t\in\mathcal{J}_i}\widetilde{S}_{it}\widetilde{S}_{it}'\right)^{-1}\sum_{i=1}^{N}\sum_{t\in\mathcal{J}_i}\widetilde{S}_{it}\left(\sum_{j=1}^{k}p_{ij}\hat{\delta}_{tj}\right) = \sum_{j=1}^{k}C_{1j}$$

where

$$C_{1j} = \left(\sum_{i=1}^{N}\sum_{t\in\mathcal{J}_i}\widetilde{S}_{it}\widetilde{S}_{it}'\right)^{-1}\sum_{i=1}^{N}\sum_{t\in\mathcal{J}_i}\widetilde{S}_{it}p_{ij}\hat{\delta}_{tj}$$

is the contribution of proxy j to the aggregate bias C. Hence, we can ascertain not only the proportions of the estimated OLS bias that can be explained by the ability proxies and the factor structure, but also the proportion that can be explained by each of the proxies.

2.3 Validation Tests for Proxies

The final component of our empirical analysis entails conducting a battery of tests designed to directly evaluate the effectiveness of the proxies at measuring the true unobserved skills. In particular, we test if each individual proxy can be represented by a linear combination of the factor loadings (skills) as well as if the proxies as a set are adequate for the skills. Bai and Ng (2006) develop procedures to test if an observable macro variable is a linear combination of the common factors and if a set of macro variables can represent the factors. Here we modify their approach and apply it to the loadings instead of the factors. These tests are justified by the fact that the space spanned by the common factors/factor loadings can be consistently estimated when the sample size is large in both the cross-section and time series dimensions [see Bai and Ng (2006)].

We first discuss the individual proxy tests. Denoting the $(k \times 1)$ vector of proxies as $p_i = (p_{i1}, ..., p_{ik})'$, the goal is to test the null hypothesis,

$$H_0\colon p_{ij} = \varphi_j' \lambda_i \quad \forall i = 1, ..., N \tag{14}$$

for any given j. The construction of the test statistics involves the following steps:

1. Estimate the following model using either the IFE or the two-step CCE approach[13]:

$$y_{it} = c_i + s_{it}\beta_1 + s_{it}^2\beta_2 + e_{it}\gamma_1 + e_{it}^2\gamma_2 + s_{it}e_{it}\gamma_3 + \lambda_i' f_t + u_{it}$$

[13] The proxy validation tests are not constructed for the one step CCE approach since this method does not provide an estimate of the factor loadings, given that the factor structure is controlled for using cross-sectional averages of the observables.

2. Save the estimates of the loadings $\{\hat{\lambda}_i\}_{i=1}^{N}$.

3. Estimate the regression

$$p_{ij} = \varphi_j' \hat{\lambda}_i + \varepsilon_i \quad \text{for each } j \in \{1, ..., k\}$$

and compute the t-test as $t_i(j) = \frac{\hat{p}_{ij} - p_{ij}}{\sqrt{\widehat{var}(\hat{p}_{ij})}}$, where $\hat{p}_{ij} = \hat{\varphi}_j' \hat{\lambda}_i$ is the fitted value and $\widehat{var}(\hat{p}_{ij}) = \hat{\varphi}_j' \widehat{var}(\hat{\lambda}_i) \hat{\varphi}_j$ with $\widehat{var}(\hat{\lambda}_i)$ computed using a wild bootstrap procedure.

Once $t_i(j)$ has been obtained, various test statistics can be constructed. Let Φ_ν be the ν percentage point of the standard normal distribution. Then, define $\mathcal{A}(j) = N^{-1} \sum_{i=1}^{N} 1(|t_i(j)| > \Phi_\nu)$. This statistic is the frequency that $t_i(j)$ exceeds the ν percent critical value and allows p_{ij} to deviate from $\hat{p}_{ij}$ for a prespecified proportion of cross-sectional observations as specified by ν. Thus, if $\nu = .025$, $\mathcal{A}(j)$ should be close to .05.

The second statistic we compute is the coefficient of determination from a regression of p_{ij} onto $\hat{\lambda}_i$ for a given j. This statistic is computed as

$$R^2(j) = \frac{N^{-1} \sum_{i=1}^{N} (\hat{p}_{ij} - \bar{\hat{p}}_j)^2}{N^{-1} \sum_{i=1}^{N} (p_{ij} - \bar{p}_j)^2}$$

with $\bar{\hat{p}}_j = N^{-1} \sum_{i=1}^{N} \hat{p}_{ij}$ and $\bar{p}_j = N^{-1} \sum_{i=1}^{N} p_{ij}$. This statistic should be unity if the proxy is an exact linear combination of the factor loadings, and zero if the proxy is irrelevant. It thus provides a direct measure of the effectiveness of the proxy in representing the true unobserved skills. A $(1 - 2\nu)$ percent confidence interval for $R^2(j)$ can be obtained as

$$(R^{2-}(j), R^{2+}(j)) = \left(R^2(j) - 2\Phi_\nu \frac{R(j)[1 - R^2(j)]}{\sqrt{N}}, \; R^2(j) + 2\Phi_\nu \frac{R(j)[1 - R^2(j)]}{\sqrt{N}} \right)$$

The final set of tests is designed to evaluate the adequacy of the proxies as a set by examining the canonical correlations between p_i and $\hat{\lambda}_i$. To this end, let $S_{\lambda\lambda}$ and S_{pp} be the sample variance-covariance matrices of $\hat{\lambda}_i$ and p_i, respectively, and $S_{\lambda p}(S_{p\lambda})$ be the $r\times k$ $(k\times r)$ sample cross-covariance matrix between p_i and $\hat{\lambda}_i$. The squared canonical correlations, denoted, $\hat{\rho}^2_m,\ m = 1, ..., \min(k, r)$, are the largest eigenvalues of the $(r \times r)$ matrix $S_{\lambda\lambda}S_{\lambda p}S^{-1}_{pp}S_{p\lambda}$. If each proxy is an exact linear combination of the loadings, the non-zero population correlations should all be unity. A $(1-2\nu)$ percent confidence interval for the population canonical correlations $\rho^2_m,\ m = 1, ..., \min(k, r)$, can be obtained as

$$(\hat{\rho}^{2-}_m, \hat{\rho}^{2+}_m) = \left(\hat{\rho}^2_m - 2\Phi_\nu\frac{\hat{\rho}_m[1-\hat{\rho}^2_m]}{\sqrt{N}},\ \hat{\rho}^2_m + 2\Phi_\nu\frac{\hat{\rho}_m[1-\hat{\rho}^2_m]}{\sqrt{N}}\right)$$

3 Data

We use the National Longitudinal Survey of Youth (NLSY79) for our empirical analysis. The information collected by the NLSY79 includes schooling choices, labor market outcomes, cognitive and non-cognitive test scores, family background and individual characteristics, for a nationally representative sample of 12,686 young men and women who were 14-22 years old in 1979. These individuals were interviewed yearly through 1994 and every other year thereafter. Importantly, the cognitive and non-cognitive tests were administered between 1979 and 1980. Therefore, we construct a panel dataset with hourly earnings, years of schooling, and other covariates from 1981 to 2016 to mitigate the concern about reverse causality from schooling to test scores.

To evaluate the efficacy of cognitive ability proxies, we follow the literature and exploit the Armed Services Vocational Aptitude Battery (ASVAB) which was administered in 1980 to NLSY79 respondents and is composed of ten intelligence tests (see Heckman and Vytlacil, 2001, Table 2, for a detailed description of each subarea). A widely used measure of cognitive ability, the Armed Forces Qualification Test (AFQT), is a composite score constructed from four ASVAB subtests: Arithmetic Reasoning, Mathematics Knowledge, Paragraph Comprehension, and Word Knowledge (e.g. Neal and Johnson 1996; Altonji et al. 2012; Deming 2017; Das and Polachek 2019; Williams 2019; Ashworth et al. 2020). In this paper, we follow Koop and Tobias (2004) and make use of all ten ASVAB subtests, which include the four in the AFQT, General Science, Coding Speed, Numeric Operations, as well as three sections commonly referred as the technical composites: Auto and Shop Information, Electronics Information, and Mechanical Comprehension to measure mechanical ability (Prada and Urzúa, 2017). A principal component analysis (PCA) of all ten standardized test scores reveals that the first principal component explains 64% of the variance using individuals in our final sample, which we denote as the cognitive factor in our PCA approach, a commonly employed method in the literature (Cawley et al., 1997; Heckman and Vytlacil, 2001; Koop and Tobias, 2004; Heckman et al., 2006). In addition, we also directly use the standardized test scores of all ten ASVAB subtests as a robustness check, which we refer to as the non-PCA approach.[14]

As for the non-cognitive ability proxies, we follow Heckman et al. (2006) and utilize two attitudinal scales available in the NLSY79, the Rotter Locus of Control Scale and the Rosenberg Self-Esteem Scale. The Rotter Locus of Control Scale, designed to measure the extent to which individuals believe they have internal control (e.g. self-motivation) over their lives as opposed to the extent that external forces (e.g. fate, luck) control their life outcomes (Rotter, 1966), was collected as part of the initial interviews in 1979. The Rotter scale is re-scored in the internal direction such that a high score indicates beliefs in internal control. The Rosenberg Self-Esteem Scale, which describes a degree of approval or disapproval toward oneself (Rosenberg, 1965), was administered during the 1980 interviews. Higher scores indicate higher self-esteem. Following Heckman et al. (2006), we extract the first three principal components as the non-cognitive factors in our PCA approach using

[14] An alternative approach that uses only AFQT and its square as the proxy variables yielded very similar results to the two approaches shown in the paper. These results are available upon request.

4 items from the Rotter Locus of Control Scale and 10 items from the Rosenberg Self-Esteem Scale, after carefully adjusting the direction of individual items to ensure consistency before applying standardization.[15] Our analysis reveals that while the three components cumulatively explain about 49% of the variance using individuals in our final sample, the first principal component only accounts for 31% of the variance. This is consistent with the evidence that there seems to be no single "g factor" among the non-cognitive items (Heckman et al., 2006). For the non-PCA approach, we include the standardized scores on the Rotter and Rosenberg composite scales as measures of non-cognitive abilities.

To remain consistent with the literature, we restrict the sample of analysis to white males who are at least 16 years old in a given year to analyze a population that historically is strongly attached to the labor market, and least likely to experience wage discrimination (Ginther, 2000).[16] Following Koop and Tobias (2004), we further limit the sample to individuals who reported working at least 30 weeks and 800 hours in the previous year to ensure the hourly wage variable reasonably captures individuals' full earning potential. After carefully removing observations with missing covariates, outlier observations in the top and bottom 1% of the wage distribution, and individuals with negative or abnormally large (years of schooling changes by more than the number of years between interviews) schooling changes, our final sample contains $2,420$ individuals for a total of $33,248$ person-year observations. In this unbalanced panel dataset, the majority of individuals have more than ten years of observations (Figure 1).[17]

[15] Specifically, the 1st-3rd items in the Rotter Locus of Control Scale are re-scored in the internal direction such that the higher the score, the more internal the individual. Similarly, the 1st, 2nd, 4th, 6th, and 7th items in the Rosenberg Self-Esteem Scale are recalculated such that higher score indicates higher self-esteem. See Heckman et al., 2006, Web Appendix, Table S27 and S28, for detailed descriptions of individual items in the Rotter and Rosenberg scales.

[16] Individuals in our sample are in their prime working age, ranging from 16-59 years old.

[17] Although there are many papers that use panel data and within individual variation to study returns to schooling (Angrist and Newey, 1991; Ashworth et al., 2020; Kejriwal et al., 2020; Koop and Tobias, 2004; Westerlund and Petrova, 2018), a potential concern is that it requires individuals who work before or while they finish schooling. This raises potential concerns related to sample selection and to econometric issues of comparing earnings before, during, and after schooling, since earnings before or during schooling could

Table 1 presents descriptive statistics for not only the main sample of analysis, but also how the summary statistics change as we sequentially add the selection criteria to obtain our final sample. Column (1) shows a baseline sample of 9,224 individuals who are at least 16 years old in a given year with no missing covariates. Focusing our attention on white males decreases the sample by more than two thirds to 2,954 individuals in column (2). It is worth noting that, hourly wage earning is 24% higher on average for white males compared to the baseline sample. The AFQT score is more than a third of a standard deviation higher for white males which is also reflected in the higher test scores of the ASVAB individual components. As for non-cognitive abilities, the white male group appears to have a relatively higher tendency to believe in internal control (higher Rotter locus of control score), and higher self-esteem (higher Rosenberg self-esteem score) on average. Other notable changes include higher mother's and father's education levels (+0.82 and +1.06 years, respectively), and lower number of siblings (-0.58). Column (3) imposes wages and hours worked restrictions, but only shrinks the sample by 111 individuals, presumably due to the aforementioned reason that white males above age 16 are strongly attached to the labor market. After trimming the wage outliers in Column (4), we lose another person, and the number of observations decreases to 37,452. The average wage converted to 1999 dollars drops to about 16 dollars an hour, and the standard deviation is greatly reduced to a reasonable level. Finally, column (5) displays summary statistics of the sample of analysis

be part-time or seasonal work and not truly reflect an individual's earning potential (Card, 1995; Lazear, 1977). Moreover, within-individual fixed effects estimators are sensitive to measurement error, which raises concerns over whether much of our schooling variation is measurement error. We believe these concerns are mitigated by the facts that: (1) we set minimum annual weeks and hours worked restrictions following Koop and Tobias (2004), in order to reduce the likelihood that observed wages do not reflect an individual's true earning potential; (2) we carefully remove individuals with abnormal schooling changes, thus removing obvious measurement error in schooling; (3) we extend the estimators from Bai (2009) and Pesaran (2006) to unbalanced panels, thus allowing us to relax the requirement in Kejriwal et al. (2020) that individuals have no gaps in earnings; (4) our OLS and ability proxy results are very similar to the literature that does not use panel data; and (5) other research has shown that the student population that works during school is large (Bacolod and Hotz, 2006; Bound et al., 2012; Carnevale et al., 2015; Hotz et al., 2002), and is thus an important population itself.

after we carefully exclude individuals with clearly misreported schooling changes. Most covariates, especially years of schooling and age, remain fairly stable across columns (2)-(5) and are not sensitive to additional restrictions imposed. Since ability proxies are the focus of this paper, we compute one cognitive and three non-cognitive factors for individuals in the final sample as shown in column (5). Appendix B shows more information on the variation in years of schooling for the final sample.

4 Empirical Results

The empirical results of our analysis are organized into five subsections. Section 4.1 presents the set of estimated specifications that differ according to whether ability proxies, demographics and/or interactive fixed effects are included. Section 4.2 reports the parameter estimates from the different specifications that facilitate the estimation of the aggregate OLS bias as well as underscore the difference between the use of ability proxies vis-a-vis interactive fixed effects to control for the underlying unobserved skills. Section 4.3 presents the bias decomposition results that allow us to quantify the contribution of the proxies to the OLS bias. Section 4.4 contains the results from proxy validation tests. Finally, Section 4.5 discusses the findings of our analysis in relation to the existing literature in order to provide further perspective on our results.

4.1 Estimated Specifications

We estimate a total of eight specifications which are summarized in Table 2. Depending on whether demographics and/or interactive fixed effects are included, we have the following four groups of specifications:

- *Group 1* [Specifications 1-2]: These specifications do not include any demographic controls or interactive fixed effects. Specification 1 includes only the basic age controls while Specification 2 appends Specification 1 with the ability proxies.

- *Group 2* [Specifications 3-4]: These specifications are the counterparts of those in Group 1 that include the demographic controls.

- *Group 3* [Specifications 5-6]: These specifications are the counterparts of those in Group 1 that include interactive fixed effects.

- *Group 4* [Specifications 7-8]: These specifications include both the demographic controls and interactive fixed effects.

4.2 Parameter Estimates

Table 3 reports the estimates of the MRTS (in percentage form) for the specifications presented in Table 2. We consider two alternative methods to account for proxies, namely, the PCA approach and the non-PCA (individual proxies) approach (see Section 3). Panels A and B present the results without and with the demographic controls, respectively. The demographic controls employed are mother's and father's level of education as well as the number of siblings.

Consider first the results without the demographic controls (Panel A). Several features of the findings are noteworthy. First, the OLS estimate from the base specification that only includes the age controls is about 8.3% which lies in the range typically reported in the literature.[18] Second, a comparison of the OLS estimates for the specifications in Group 1 [columns (1)-(3)] reveals that, regardless of the particular approach used to control for the proxies, their inclusion only marginally reduces the MRTS estimate. Specifically, the PCA and non-PCA approaches entail reductions of only 4% and 19% in the OLS estimate, respectively. Thus, conditional on the age controls, schooling is only marginally correlated

[18]For instance, Gunderson and Oreopolous (2020) report a range of 6-10% for OLS estimates based on their survey of the returns to schooling literature.

with the proxies. Third, a comparison of the estimates for the specifications in Group 3 [columns (4)-(12)] with the OLS estimate in column (1) shows that the inclusion of interactive effects make the MRTS estimate discernibly smaller, with the corresponding estimate of the OLS bias ranging between 36-43% of the OLS estimate. Fourth, conditional on the interactive fixed effects, the proxies have rather limited explanatory power as evidenced by the very similar estimates in columns (4)-(6) for the IFE approach and columns (7)-(12) for the CCE approach. On the other hand, conditional on the proxies, the interactive effects retain considerable explanatory power as indicated by the comparison of estimates in columns (2)-(3) with their counterparts that allow for interactive effects [columns (5)-(6) for the IFE approach and (8)-(9)/(11)-(12) for the CCE approach]. Fifth, the IFE and CCE approaches to account for interactive effects yield similar estimates ranging between 4.7-5.3%. Hence, the particular method used to control for the factor structure matters much less compared to using ability proxies. Finally, model fit as measured by the adjusted R^2 improves considerably with the inclusion of the factor structure while the inclusion of the proxies only has a marginal impact.

When demographic controls are included (Panel B), the OLS estimate in the base specification is slightly smaller, about 8%, though still consistent with the estimates from the literature. As in Panel A, including the proxies only engineers a marginal reduction in the estimate of the schooling effect. The FM estimates are noticeably smaller, ranging between 3.9-4.9% for the most general specification that also includes the proxies. The corresponding estimate of the ability bias ranges between 39-51% of the OLS estimate. Importantly, the inclusion of interactive fixed effects in a specification with demographics entails a much larger decline in the MRTS estimate relative to the inclusion of demographics in a specification with interactive fixed effects. For instance, the IFE estimate with demographic controls but without proxies [column (4)] is about 3 percentage points lower than the corresponding OLS estimate in column (1) while the difference in the IFE estimate with and without the demographic controls is only about 0.3 points [difference in the column (4) estimate between Panels A and B]. This finding suggests that the factor structure cannot be interpreted in terms of time-varying returns to time-invariant individual characteristics. Appendix C contains two robustness checks on the results in Table 3. First, Table C.1 shows the sensitivity of the MRTS estimate based on the IFE approach to the number of factors.[19] The results indicate that the estimate is fairly stable once five or more factors are included in the model. Second, Table C.2 reports the results for specifications that include a quartic age term.[20] The results are qualitatively very similar to those presented in Table 3.[21]

Overall, the pattern of estimates with and without the demographic controls are qualitatively similar and show that the factor structure contains variation independent of proxies and/or demographics that is correlated with schooling. Moreover, when employing the factor model framework as a test bed for evaluation, the ability proxies are of little use in controlling for ability bias.

4.3 Bias Decomposition

Tables 4 and 5 presents the results from the bias decomposition analysis described in Section 2.2 that serves to provide a robust mechanism for isolating the OLS bias components attributable to the proxies vis-à-vis the factor structure. We compute the aggregate OLS bias, the contribution of the factor loadings and proxies, as well as the cognitive/non-cognitive principal components of the proxies in the PCA approach and the individual proxies in the non-PCA approach.

[19] We also computed the number of factors using the information criteria in Bai and Ng (2002) but those turned out to offer little guidance. In particular, two of the three criteria (IC_{p1}, IC_{p2}) always select the maximum number of factors (ten) while IC_{p3} always selects three factors.

[20] See Murphy and Welch (1990) and Cho and Phillips (2018) for empirical evidence favoring a quartic specification for experience.

[21] The results from the bias decomposition analysis as well as the proxy tests for the model with quartic age have the same qualitative pattern as those reported in Tables 4-7. These results are available upon request.

Table 4 reports the findings based on specifications where the interactive fixed effects are not included so that ability bias is controlled for only using the proxies. Columns (1)-(2) and (3)-(4) present the results excluding and including the demographic controls, respectively. In addition to point estimates, we report 95% confidence intervals constructed using the approach proposed in Gelbach (2016). For the PCA approach, the cognitive factor emerges as the chief contributor to the aggregate bias while the contribution of each of the non-cognitive factors is comparatively rather limited. For the individual proxies, the cognitive measures turn out to be relatively much more important overall than their non-cognitive counterparts, consistent with the PCA results. Among the cognitive measures, Mathematical Knowledge appears as the leading contender for explaining the OLS bias. When demographics are included, the aggregate contribution of the proxies diminishes slightly although the qualitative pattern of the results remains similar. The fact that cognitive measures appear to be more important than non-cognitive measures in explaining ability bias is consistent with prior work using the ability proxy approach (Caplan, 2018).

Table 5 reports the results when the interactive fixed effects are included in addition to the proxies. A comparison of these results with those in Table 4 serves to clearly delineate the role of proxies in controlling for the OLS bias. Panels A and B report the results excluding and including the demographic controls, respectively. Here the 95% confidence intervals are computed using a wild cluster bootstrap procedure (Appendix A). Regardless of whether the demographic controls are included and the particular proxy approach adopted, the main takeaway from these results is that the contribution of the proxies to the aggregate OLS bias is quantitatively quite small relative to the contribution of the factor structure. Thus, conditional on the interactive fixed effects, the proxies are only marginally correlated with earnings. Finally, it is reassuring that the overall pattern of results is remarkably similar across the three FM approaches.

In stark contrast to the results in Table 4 where the cognitive factor accounted for the bulk of the bias, this factor is now almost as uninformative as its non-cognitive counterparts once demographics are controlled for. Among the individual proxies, the non-cognitive measures again appear to contribute much less to the bias relative to the cognitive measures. Interestingly, the relative importance of the individual proxies depends crucially on whether demographics are included. For instance, while Auto and Shop Information appears as the dominant bias component when demographics are excluded, its contribution is considerably diminished by the inclusion of demographics with Mathematical Knowledge now emerging as the leading contributor. The rather minor contribution of the proxies to the OLS bias relative to the factor structure reflects the presence of substantial measurement error in the proxies that is captured by the factor structure (see equation 10). On the other hand, while the proxy approach can be potentially used to study which types of abilities contribute to the overall bias, the skills estimated by the FM approach cannot be assigned an economic interpretation given that the approach only identifies the space of the latent skills (see, e.g., Bai and Ng, 2002).

4.4 Proxy Validation Tests

As a complement to the preceding analysis, we conduct the proxy validation tests discussed in Section 2.3 that are designed to directly assess the suitability of the proxies in representing the latent skills. Table 6 reports the results for the $\mathcal{A}(j)$ and $R^2(j)$ statistics. The $\mathcal{A}(j)$ statistics are computed with $\nu = .025$ so that if proxy j was indeed a linear combination of the factor loadings as specified by the null hypothesis, $\mathcal{A}(j)$ should be close to .05. However, the results make it evident that this is not the case for any of the proxies or their principal components. When demographic controls are excluded (Panel A), the $\mathcal{A}(j)$ values all exceed 89% and 93% for the IFE and CCE-2 approaches, respectively, thereby providing strong evidence against the null hypothesis. When these controls are included (Panel B), the values are comparatively a bit lower, though still in excess of 75% for the IFE approach and 81% for the CCE-2 approach.

Turning to the $R^2(j)$ statistics, the relevance of the proxies can be gauged by how close these statistics are to unity.[22] We present both point estimates and 95% confidence intervals. These results paint a very similar picture as the $\mathcal{A}(j)$ statistics and offer little support to the proxies: the point and interval estimates are all below .06 regardless of the estimation method as well as whether demographic controls are included.

Finally, Table 7 presents the results from evaluating the adequacy of the proxies as a set using the canonical correlations test. Both point estimates and 95% confidence intervals are presented. For the IFE approach, the largest nine (eight) squared sample canonical correlations are reported for the case without (with) demographic controls, given the estimated number of factors in these cases (when proxies are not included). For the CCE-2 approach, the largest four correlations are reported given that the number of factors is set at four. The results are remarkably consistent between the two estimation approaches. The point estimate of ρ_1^2 is about 0.11 for both approaches when demographic controls are excluded and between 0.38-0.42 when such controls are included. The point and interval estimates for ρ_m^2, $m > 1$, are all less than 10% regardless of the estimation approach and the inclusion/exclusion of demographic controls. Overall, these estimates fall well below the threshold of 0.95 suggested by Bai and Ng (2006) to validate the relevance of the proxies. In summary, the results from all three proxy tests serve to further strengthen the evidence against the efficacy of the ability proxies in estimating the returns to schooling.

4.5 Discussion

We now provide a discussion of the results presented in the preceding subsections with particular emphasis on their relationship to the existing literature. Caplan (2018, Chapter 3)

[22] Bai and Ng (2006) suggest a value below 0.95 for $R^2(j)$ and $\hat{\rho}_m^2$ as indicative of non-negligible measurement errors.

surveys a large number of studies that use ability proxies to control for ability bias. He finds that while controlling for cognitive ability and/or non-cognitive ability generally reduces the education premium, there is considerable variation across studies on the magnitude of the reduction depending on the specific proxies employed. Recognizing the imperfect nature of the proxies (especially for soft skills), he suggests that ability bias ranges between 25-45% (with 20-30% from cognitive ability bias and 5-15% from non-cognitive ability bias) with the lower limit corresponding to a "cautious" estimate and the upper limit to a "reasonable" estimate. Our analysis provides a formal econometric framework for estimating the magnitude of this bias by relying on the interactive fixed effects structure and suggests a range between 39-51% for the most general specification that includes the demographic controls.

Heckman et al. (2006) use data from the NLSY79 to demonstrate the importance of both cognitive and non-cognitive skills in explaining schooling decisions and wages among a diverse array of outcomes and dimensions of behavior. They are more concerned with the role of these latent cognitive and non-cognitive skills in explaining a wide variety of outcomes and in carefully accounting for endogeneity and simultaneity among those outcomes than they are with estimating the average marginal returns to schooling. Yet, in a similar vein, our FM estimates in Section 4.2 and the bias decomposition results in Section 4.3 show that the OLS estimates are subject to a substantial upward bias thereby suggesting a rather important role for the underlying unobserved skills that are captured by the factor structure. An important difference between our analysis and that in Heckman et al. (2006) is that their identification strategy is based on assuming the existence of two sets of measurements (each with at least two elements), with one set measuring cognitive skills and the other set measuring non-cognitive skills. Further, latent cognitive ability is only allowed to affect scores on cognitive measures while latent non-cognitive ability is only allowed to affect scores on non-cognitive measures. Their approach has the advantage that it leads to a direct interpretation of their skills as being cognitive and non-cognitive because of the proxies used in the models, which allows them to study the impact of different types of skills on a wide variety of outcomes. In contrast, our analysis does not rely on the measured proxies to identify the unobserved factor structure but instead exploits the large panel nature of the dataset and the within individual variation in earnings and schooling to achieve identification. Our approach has the advantage of not imposing restrictions on the relationship between outcomes and different abilities like cognitive and non-cognitive skills. It also has the advantage of not relying on ability proxies from the NLSY79 to identify skills, which is important not only because the proxies might be measured with substantial error, as we show, but also because we are not limited to the type and number of abilities for which proxies exist.

More recently, Polachek et al. (2015) and Das and Polachek (2019) derive a nonlinear earnings function based on a structural human capital life cycle model and obtain five ability measures. Of these, three correspond directly to one's ability to create human capital, one to skill depreciation, and one to a person's time discount rate. They argue that these measures are jointly the most important determinants of individual earnings variation. In particular, Das and Polachek (2019) find that including these five measures in a regression of log earnings on schooling and a quadratic in experience increases the adjusted R^2 by 19% and reduces the schooling coefficient from 13.4% to 6.5%. Moreover, conditional on these ability measures, including AFQT has little effect - the adjusted R^2 only increases by 0.001 while the schooling coefficient drops from 6.5% to 6.1%. One explanation for this finding is that while standardized tests capture analytic academic capabilities, they do not necessarily indicate a proficiency that translates into real-world accomplishments. The results of our reduced form analysis can be viewed as complementary to these two studies. While their ability measures are derived from a structural model, we control for the underlying skills using the interactive fixed effects structure. Consistent with their results, we find that the inclusion of interactive fixed effects leads to a substantive reduction in the estimate of the schooling effect and that conditional on this structure, the ability proxies only have a marginal impact on returns to schooling and adjusted R^2.

Finally, KLT also use the FM estimators from Bai (2009) and Pesaran (2006). KLT estimate the average marginal return to schooling using FM estimators and a new linked survey-administrative dataset thereby simultaneously addressing the twin issues of heterogeneity in returns to schooling and the endogeneity of schooling. The linked SIPP-administrative data does not contain ability proxies and thus cannot speak to the efficacy of ability proxies for removing ability bias. However, it is informative to compare OLS and FM estimates between the results in KLT and this paper. The preferred results in KLT with heterogeneity in returns to schooling (Table 5) show an average marginal return of 2.8-4.4% relative to an OLS return of 7.8% for a reduction of 44-64%. Our most general specification shows an average marginal return of 3.9-4.9% relative to an OLS return of 8% for a reduction of 39-51%. The similarity in both average marginal returns to schooling and ability bias despite different data sources, differences in sample construction (e.g., balanced versus unbalanced panel), and different specifications (e.g., heterogenous versus pooled models) strengthens the credibility of our results.

5 Conclusion

A vast body of empirical work has emerged over the past few decades that adopted a variety of econometric approaches to address the issue of ability bias when estimating the returns to schooling. The ability proxy approach entails the inclusion of various measures of cognitive and non-cognitive ability in the earnings-schooling regression as proxies for the unobserved skills. Central to the viability of such an approach is whether these proxies capture the underlying skills that matter for earnings. While it is well known that mismeasured proxies render the resulting estimates biased, attempts at quantifying the magnitude of this bias appear to be scarce. Proposed solutions in this context are sensitive to the violation of exclusion restrictions that rule out direct effects of family background variables on earnings [e.g., Griliches (1977)]. This paper conducts a formal assessment of this approach by adopting an interactive fixed effects methodology in which unobserved ability is allowed to be multidimensional where each component is characterized by its own contribution to earnings with skill prices that can vary over time. The endogeneity of schooling is accounted for through estimation of or proxying for the skill prices without the need to rely on proxies for ability which facilitates the decomposition of the aggregate least squares bias into components that can and cannot be explained by the proxies.

The empirical results show that while the OLS estimates lie within the range typically reported in the literature, our preferred factor model estimators are discernibly smaller, suggesting the presence of substantial ability bias. Bias decomposition methods find that the performance of ability proxies in controlling for this bias is rather disappointing - the commonly used proxies can explain very little of the estimated ability bias. We corroborate our findings with a set of proxy validation tests that directly evaluate the relevance of the proxies in capturing the underlying individual abilities.

While the quadratic specification for schooling adopted in our empirical analysis is common in the literature, an interesting extension is to allow for the possibility that educational credentials (degree completion) have a direct effect on earnings, i.e., the presence of *sheepskin effects* [e.g., Hungerford and Solon (1987)]. In particular, if the value of additional years of school is partly related to the value of degree attainment rather than knowledge obtained in each year, then returns may be larger for individuals who complete bachelor's and graduate degrees than for individuals who drop out of college. An extended specification that controls for these degree effects would be useful in decomposing the returns to schooling into its human capital and sheepskin components, thereby allowing us to at least partly speak to the relative importance of the human capital and signaling theories within the interactive fixed effects framework (Caplan, 2018). Further, the role of ability bias could be explored by comparing the contribution of the sheepskin component from the least squares and interactive fixed effects specifications, given that the importance of sheepskin effects may be overstated due to the presence of ability bias (Lange and Topel, 2006).

References

Altonji, J. G., P. Bharadwaj, and F. Lange (2012). Changes in the characteristics of american youth: Implications for adult outcomes. *Journal of Labor Economics 30*(4), 783–828.

Altonji, J. G. and C. R. Pierret (2001). Employer learning and statistical discrimination. *The quarterly journal of economics 116*(1), 313–350.

Angrist, J. D. and W. K. Newey (1991). Over-identification tests in earnings functions with fixed effects. *Journal of Business & Economic Statistics 9*(3), 317–323.

Ashworth, J., V. J. Hotz, A. Maurel, and T. Ransom (2020). Changes across cohorts in wage returns to schooling and early work experiences. *Journal of Labor Economic*, forthcoming.

Bacolod, M. and V. J. Hotz (2006). Cohort changes in the transition from school to work: Evidence from three nls surveys. *Economics of Education Review 25*(4), 351–373.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77*(4), 1229–1279.

Bai, J., Y. Liao, and J. Yang (2015). Unbalanced panel data models with interactive effects. In *The Oxford Handbook of Panel Data*. Oxford University Press.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bai, J. and S. Ng (2006). Evaluating latent and observed factors in macroeconomics and finance. *Journal of Econometrics 131*(1-2), 507–537.

Belzil, C. and J. Hansen (2002). Unobserved ability and the return to schooling. *Econometrica 70*(5), 2075–2091.

Bound, J., M. Lovenheim, and S. Turner (2012). Increasing time to baccalaureate degree in the united states. *Education Finance and Policy 7*(4), 375—-424.

Caplan, B. (2018). *The case against education: Why the education system is a waste of time and money*. Princeton University Press.

Card, D. (1995). Earnings, schooling, and ability revisited. *Research in Labor Economics 14*, 23–48.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica 69*(5), 1127–1160.

Carneiro, P., K. T. Hansen, and J. J. Heckman (2003). 2001 lawrence r. klein lecture estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review 44*(2), 361–422.

Carnevale, A. P., N. Smith, M. Melton, and E. Price (2015). Learning while earning: The new normal. *Georgetown University Center on Education and the Workforce*.

Cawley, J., K. Conneely, J. Heckman, and E. Vytlacil (1997). Cognitive ability, wages, and meritocracy. In *Intelligence, genes, and success*, pp. 179–192. Springer.

Cho, J. S. and P. C. Phillips (2018). Sequentially testing polynomial model hypotheses using power transforms of regressors. *Journal of applied econometrics 33*(1), 141–159.

Cunha, F., F. Karahan, and I. Soares (2011). Returns to skills and the college premium. *Journal of Money, Credit and Banking 43*, 39–86.

Das, T. and S. W. Polachek (2019). Microfoundations of earnings differences. In *The Palgrave Handbook of Economic Performance Analysis*, pp. 9–76. Springer.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics 132*(4), 1593–1640.

Gelbach, J. B. (2016). When do covariates matter? and which ones, and how much? *Journal of Labor Economics 34*(2), 509–543.

Ginther, D. K. (2000). Alternative estimates of the effect of schooling on earnings. *Review of Economics and Statistics 82*(1), 103–116.

Griliches, Z. (1977). Estimating the returns to schooling: Some econometric problems. *Econometrica: Journal of the Econometric Society 45*(1), 1–22.

Gunderson, M. and P. Oreopolous (2020). Returns to education in developed countries. In *The Economics of Education*, pp. 39–51. Elsevier.

Hansen, K. T., J. J. Heckman, and K. J. Mullen (2004). The effect of schooling and ability on achievement test scores. *Journal of Econometrics 121*(1-2), 39–98.

Heckman, J. and E. 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 83*(1), 1–12.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2018). Returns to education: The causal effects of education on earnings, health, and smoking. *Journal of Political Economy 126*(S1), S197–S246.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006). Earnings functions, rates of return and treatment effects: The mincer equation and beyond. *Handbook of the Economics of Education 1*, 307–458.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics 24*(3), 411–482.

Hotz, V. J., L. C. Xu, M. Tienda, and A. Ahituv (2002). Are there returns to the wages of young men from working while in school? *The Review of Economics and Statistics 84*(2), 221–236.

Hungerford, T. and G. Solon (1987). Sheepskin effects in the returns to education. *The review of economics and statistics 69*(1), 175–177.

Kejriwal, M., X. Li, and E. Totty (2020). Multidimensional skills and the returns to schooling: Evidence from an interactive fixed-effects approach and a linked survey-administrative data set. *Journal of Applied Econometrics 35*(5), 548–566.

Kim, D. and T. Oka (2014). Divorce law reforms and divorce rates in the usa: An interactive fixed-effects approach. *Journal of Applied Econometrics 29*(2), 231–245.

Koop, G. and J. L. Tobias (2004). Learning about heterogeneity in returns to schooling. *Journal of Applied Econometrics 19*(7), 827–849.

Lange, F. and R. Topel (2006). The social value of education and human capital. *Handbook of the Economics of Education 1*, 459–509.

Lazear, E. P. (1977). Schooling as a wage depressant. *Journal of Human Resources 12*(2), 164–176.

Lemieux, T. (2006). The "mincer equation" thirty years after schooling, experience, and earnings. In *Jacob Mincer a pioneer of modern labor economics*, pp. 127–145. Springer.

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

Mincer, J. (1974). *Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2.* ERIC.

Mincer, J. (1996). Changes in wage inequality, 1970-1990. Technical report, National Bureau of Economic Research.

Murphy, K. M. and F. Welch (1990). Empirical age-earnings profiles. *Journal of Labor Economics 8*(2), 202–229.

Neal, D. A. and W. R. Johnson (1996). The role of premarket factors in black-white wage differences. *Journal of political Economy 104*(5), 869–895.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica 74*(4), 967–1012.

Polachek, S. W., T. Das, and R. Thamma-Apiroam (2015). Micro-and macroeconomic implications of heterogeneity in the production of human capital. *Journal of Political Economy 123*(6), 1410–1455.

Prada, M. F. and S. Urzúa (2017). One size does not fit all: Multiple dimensions of ability, college attendance, and earnings. *Journal of Labor Economics 35*(4), 953–991.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton university press.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and applied 80*(1), 1.

Westerlund, J. and Y. Petrova (2018). Asymptotic collinearity in cce estimation of interactive effects models. *Economic Modelling 70*, 331–337.

Westerlund, J. and J.-P. Urbain (2013). On the estimation and inference in factor-augmented panel regressions with correlated loadings. *Economics Letters 119*(3), 247–250.

Williams, B. (2019). Controlling for ability using test scores. *Journal of Applied Econometrics 34*(4), 547–565.

Zhou, Q. and Y. Zhang (2016). Common correlated effects estimation of unbalanced panel data models with cross-sectional dependence. *Journal of Economic Theory and Econometrics 27*(4), 25–45.

Figures and Tables

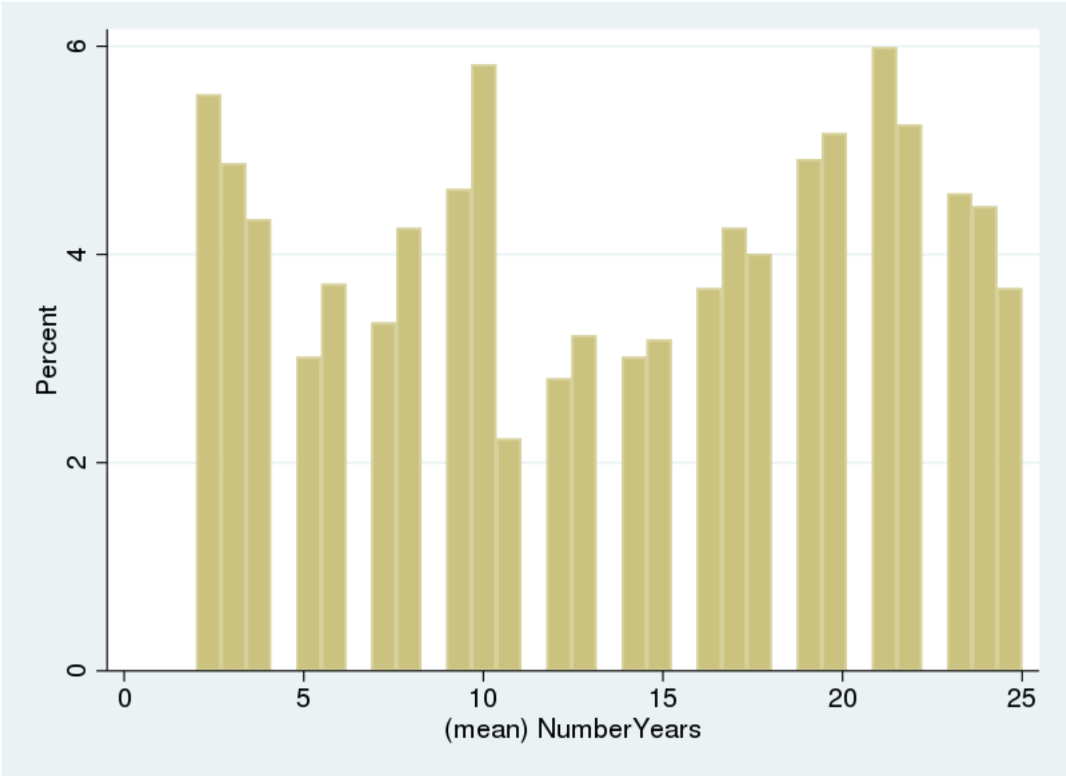

Figure 1: Distribution of Number of Years in the Final Sample

Table 1: Summary Statistics

	(1) No missing variables and age 16+	(2) White males	(3) Weeks and hours worked restrictions	(4) No outlier wages	(5) No abnormal schooling changes
Hourly wage inflation-adjusted	18.80	23.28	24.00	15.94	15.82
	(349.95)	(363.95)	(349.57)	(10.18)	(9.99)
Log hourly wage inflation-adjusted	2.37	2.56	2.64	2.60	2.60
	(0.74)	(0.76)	(0.73)	(0.58)	(0.57)
Years of school	13.19	13.29	13.39	13.34	13.21
	(2.36)	(2.48)	(2.47)	(2.44)	(2.28)
Age	32.66	32.09	33.05	32.78	32.82
	(10.20)	(10.11)	(9.92)	(9.79)	(9.78)
Mother's years of school	11.13	11.95	11.96	11.93	11.90
	(3.11)	(2.43)	(2.40)	(2.38)	(2.33)
Father's years of school	11.11	12.17	12.18	12.14	12.07
	(3.90)	(3.41)	(3.36)	(3.34)	(3.28)
Number of siblings	3.60	3.02	2.99	3.00	3.01
	(2.50)	(2.00)	(1.98)	(1.98)	(1.99)
AFQT standardized	0.13	0.50	0.52	0.51	0.50
	(1.00)	(0.99)	(0.97)	(0.97)	(0.96)
ASVAB standardized - General Science	0.12	0.61	0.63	0.62	0.61
	(0.97)	(0.88)	(0.87)	(0.87)	(0.86)
ASVAB standardized - Arithmetic Reasoning	0.12	0.57	0.60	0.59	0.58
	(1.00)	(0.99)	(0.97)	(0.97)	(0.97)
ASVAB standardized - Mathematical Knowledge	0.15	0.44	0.46	0.45	0.43
	(1.01)	(1.04)	(1.03)	(1.03)	(1.01)
ASVAB standardized - Paragraph Comprehension	0.13	0.31	0.33	0.33	0.32
	(0.95)	(0.87)	(0.85)	(0.85)	(0.85)
ASVAB standardized - Word Knowledge	0.12	0.41	0.43	0.43	0.43
	(0.93)	(0.81)	(0.79)	(0.79)	(0.79)
ASVAB standardized - Coding Speed	0.11	0.11	0.14	0.13	0.13
	(0.95)	(0.88)	(0.87)	(0.87)	(0.87)
ASVAB standardized - Numeric Operations	0.14	0.25	0.27	0.26	0.27
	(0.94)	(0.9)	(0.88)	(0.88)	(0.88)
ASVAB standardized - Auto and Shop Information	0.07	0.92	0.95	0.95	0.97
	(0.98)	(0.86)	(0.85)	(0.85)	(0.85)
ASVAB standardized - Electronic Information	0.08	0.75	0.78	0.78	0.79
	(0.98)	(0.89)	(0.88)	(0.88)	(0.87)
ASVAB standardized - Mechanical Comprehension	0.10	0.79	0.82	0.82	0.82
	(0.98)	(0.93)	(0.91)	(0.91)	(0.91)
Rotter locus of control standardized	0.03	0.14	0.15	0.15	0.14
	(0.98)	(0.98)	(0.98)	(0.98)	(0.97)
Rosenberg self-esteem standardized	0.04	0.10	0.11	0.11	0.12
	(0.98)	(0.96)	(0.96)	(0.96)	(0.96)
Cognitive factor					0.15
					(2.43)
Noncognitive factor 1					0.01
					(2.09)
Noncognitive factor 2					0.00
					(1.14)
Noncognitive factor 3					-0.02
					(1.08)
Number of Individuals (N)	9,224	2,954	2,843	2,842	2,420
Number of Observations ($N * T$)	145,473	45,291	38,457	37,452	33,248

Source: National Longitudinal Survey of Youth (NLSY79), 1981-2016.

Note: Each column reports averages and standard deviations (in parentheses) for the sample specified in the column. Column (1) starts with a baseline sample. Columns (2)-(5) sequentially add additional sample criteria until the final sample is shown in column (5). Hourly earnings are adjusted for inflation to 1999 dollars.

Table 2: Summary of Estimated Specifications

	Specification	Controls (beyond the base)	Estimator
1.	$y_{it} = c_i + s_{it}\beta_1 + s_{it}^2\beta_2 + e_{it}\gamma_1 + e_{it}^2\gamma_2 + s_{it}e_{it}\gamma_3 + v_{it}$	-	OLS
2.	$y_{it} = c_i + s_{it}\beta_1 + s_{it}^2\beta_2 + e_{it}\gamma_1 + e_{it}^2\gamma_2 + s_{it}e_{it}\gamma_3 + p_i'\delta_t + v_{it}$	proxies	OLS
3.	$y_{it} = c_i + s_{it}\beta_1 + s_{it}^2\beta_2 + e_{it}\gamma_1 + e_{it}^2\gamma_2 + s_{it}e_{it}\gamma_3 + w_i'\phi_t + v_{it}$	demographics	OLS
4.	$y_{it} = c_i + s_{it}\beta_1 + s_{it}^2\beta_2 + e_{it}\gamma_1 + e_{it}^2\gamma_2 + s_{it}e_{it}\gamma_3 + w_i'\phi_t + p_i'\delta_t + v_{it}$	proxies, demographics	OLS
5.	$y_{it} = c_i + s_{it}\beta_1 + s_{it}^2\beta_2 + e_{it}\gamma_1 + e_{it}^2\gamma_2 + s_{it}e_{it}\gamma_3 + \lambda_i' f_t + u_{it}$	interactive fixed effects	IFE, CCEP, CCEP-2
6.	$y_{it} = c_i + s_{it}\beta_1 + s_{it}^2\beta_2 + e_{it}\gamma_1 + e_{it}^2\gamma_2 + s_{it}e_{it}\gamma_3 + p_i'\delta_t + \lambda_i' f_t + u_{it}$	proxies, interactive fixed effects	IFE, CCEP, CCEP-2
7.	$y_{it} = c_i + s_{it}\beta_1 + s_{it}^2\beta_2 + e_{it}\gamma_1 + e_{it}^2\gamma_2 + s_{it}e_{it}\gamma_3 + w_i'\phi_t + \lambda_i' f_t + u_{it}$	demographics, interactive fixed effects	IFE, CCEP, CCEP-2
8.	$y_{it} = c_i + s_{it}\beta_1 + s_{it}^2\beta_2 + e_{it}\gamma_1 + e_{it}^2\gamma_2 + s_{it}e_{it}\gamma_3 + w_i'\phi_t + p_i'\delta_t + \lambda_i' f_t + u_{it}$	proxies, demographics, interactive fixed effects	IFE, CCEP, CCEP-2

Table 3: Estimates of the Return to Schooling (in %) for White Males

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	OLS	OLS	OLS	IFE	IFE	IFE	CCE	CCE	CCE	CCE-2	CCE-2	CCE-2
A. Without demographic controls												
Marginal return	8.31***	7.94***	6.73***	5.32***	5.28***	4.95***	5.20***	5.17***	4.84***	5.12***	5.10***	4.74***
	(6.48,10.13)	(6.12,9.75)	(4.90,8.56)	(1.02,7.74)	(1.03,7.36)	(1.40,6.99)	(2.28,8.12)	(2.25,8.09)	(1.74,7.94)	(3.77,8.81)	(3.82,8.86)	(3.85,9.24)
Demos-by-year FE	No	No	No	No	No	No	No	No	No	No	No	No
PCA Proxies-by-year FE	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No
Non-PCA Proxies-by-year FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Adjusted R-squared	0.3327	0.336	0.3471	0.8529	0.8335	0.6546	0.66	0.6418	0.4702	0.6703	0.652	0.4768
B. With demographic controls												
Marginal return	8.04***	7.77***	6.57***	5.05***	4.88***	4.13***	4.89***	4.70***	4.05**	4.76***	4.59***	3.94***
	(6.18,9.91)	(5.92,9.62)	(4.72,8.43)	(1.77,7.60)	(1.87,7.45)	(1.76,6.71)	(1.73,8.05)	(1.56,7.84)	(0.87,7.23)	(4.04,9.09)	(3.91,8.87)	(3.43,8.32)
Demos-by-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
PCA Proxies-by-year FE	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No
Non-PCA Proxies-by-year FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Adjusted R-squared	0.338	0.341	0.3503	0.5278	0.5253	0.4084	0.3703	0.369	0.3583	0.3749	0.3735	0.3613

Source: National Longitudinal Survey of Youth (NLSY79), 1981-2016.

Note: The dependent variable is the log of hourly wage adjusted to 1999 dollars. Our final sample contains 2,420 individuals for a total of 33,248 person-year observations. 95% confidence intervals are shown in parentheses. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*. Person fixed-effects and age controls are controlled for in all specifications (see Table 2 for details). Panel B presents estimates with demographic controls including mother's and father's education levels as well as the number of siblings. Columns (4)-(6) are based on Interactive Fixed Effects (IFE) (Bai, 2009) with the estimated number of factors to be 9 in Panel A, 8 in Panel B Columns (4)-(5), and 5 in Panel B Columns (6). Columns (7)-(9) are based on Common Correlated Effects (CCE) (Pesaran, 2006). Columns (10)-(12) are based on the two-step CCE procedure with the number of factors set to be 4 in both panels.

Table 4: Bias Associated with OLS Estimates, Due to Proxies Only

	(1)	(2)	(3)	(4)
	OLS - Without Demographics		OLS - With Demographics	
	PCA Proxy	Non-PCA Proxy	PCA Proxy	Non-PCA Proxy
Total Bias from Proxies	0.37***	1.58***	0.27***	1.47***
	(0.25,0.49)	(1.28,1.87)	(0.16,0.39)	(1.20,1.75)
Bias from Cognitive Factor	0.39***		0.32***	
	(0.27,0.50)		(0.22,0.41)	
Bias from Non-Cognitive Factor 1	-0.01		-0.02	
	(-0.05,0.03)		(-0.06,0.02)	
Bias from Non-Cognitive Factor 2	0.01		0.00	
	(-0.03,0.04)		(-0.04,0.04)	
Bias from Non-Cognitive Factor 3	-0.02		-0.03	
	(-0.06,0.02)		(-0.07,0.01)	
Bias from ASVAB_General Science		0.02		0.09
		(-0.30,0.33)		(-0.05,0.23)
Bias from ASVAB_Arithmetic Reasoning		-0.04		-0.03
		(-0.36,0.27)		(-0.18,0.13)
Bias from ASVAB_Mathematical Knowledge		1.12***		0.97***
		(0.75,1.49)		(0.70,1.25)
Bias from ASVAB_Paragraph Comprehension		0.15		0.09
		(-0.08,0.39)		(-0.06,0.25)
Bias from ASVAB_Word Knowledge		-0.22		0.11
		(-0.52,0.07)		(-0.02,0.25)
Bias from ASVAB_Coding Speed		-0.06		-0.02
		(-0.16,0.04)		(-0.08,0.04)
Bias from ASVAB_Numeric Operations		0.14		0.01
		(-0.03,0.03)		(-0.11,0.13)
Bias from ASVAB_Auto and Shop Information		-0.06		0.35***
		(-0.32,0.19)		(0.21,0.49)
Bias from ASVAB_Electronic Information		0.50***		-0.12**
		(0.21,0.80)		(-0.21,-0.02)
Bias from ASVAB_Mechanical Comprehension		0.02		0.05
		(-0.25,0.30)		(-0.01,0.10)
Bias from Rotter Scale		-0.01		-0.04*
		(-0.05,0.03)		(-0.08,-0.01)
Bias from Rosenberg Scale		0.01		-0.01
		(-0.05,0.07)		(-0.05,0.03)

Table 5: Bias Associated with OLS Estimates, Due to Common Factor Structure and Proxies

	(1)	(2)	(3)	(4)	(5)	(6)
	IFE		CCE		CCE-2	
	PCA Proxy	Non-PCA Proxy	PCA Proxy	Non-PCA Proxy	PCA Proxy	Non-PCA Proxy
A. Without demographic controls						
Total Bias	3.02*	3.36**	3.14***	3.47***	3.21***	3.56***
	(-0.20,6.22)	(0.46,6.26)	(3.00,3.28)	(3.31,3.63)	(0.81,5.60)	(1.26,5.86)
Bias from Factor Loadings	3.76**	3.23**	3.85***	3.36***	3.93***	3.43***
	(0.56,6.96)	(0.53,5.93)	(3.71,3.99)	(3.22,3.50)	(1.53,6.33)	(1.03,5.83)
Bias from Proxies	-0.73***	0.13	-0.71***	0.11***	-0.71***	0.14
	(-0.81,-0.65)	(-0.48,0.74)	(-0.69,-0.73)	(0.07,0.15)	(-0.93,-0.49)	(-0.60,0.88)
Bias from Cognitive Factor	-0.71***		-0.70***		-0.70***	
	(-0.79,-0.63)		(-0.68,-0.72)		(-0.92,-0.48)	
Bias from Non-Cognitive Factor 1	-0.01		0.00		0.00	
	(-0.03,0.01)		(-0.01,0.02)		(-0.01,0.02)	
Bias from Non-Cognitive Factor 2	-0.01		0.00		0.00	
	(-0.03,0.01)		(-0.01,0.02)		(-0.01,0.02)	
Bias from Non-Cognitive Factor 3	0.00		0.00		0.00	
	(-0.01,0.02)		(-0.01,0.02)		(-0.01,0.02)	
Bias from ASVAB_General Science		-0.15***		-0.19***		-0.16***
		(-0.23,-0.07)		(-0.23,-0.15)		(-0.22,-0.10)
Bias from ASVAB_Arithmetic Reasoning		-0.11***		-0.17***		-0.14***
		(-0.15,-0.07)		(-0.23,-0.11)		(-0.20,-0.08)
Bias from ASVAB_Mathematical Knowledge		-0.60***		-0.51***		-0.54***
		(-0.76,-0.44)		(-0.55,-0.47)		(-0.86,-0.22)
Bias from ASVAB_Paragraph Comprehension		0.33***		0.36***		0.34***
		(0.29,0.37)		(0.34,0.38)		(0.28,0.40)
Bias from ASVAB_Word Knowledge		0.30***		0.25***		0.27***
		(0.22,0.38)		(0.21,0.29)		(0.11,0.43)
Bias from ASVAB_Coding Speed		0.07***		0.06***		0.05***
		(0.05,0.09)		(0.04,0.08)		(0.03,0.07)
Bias from ASVAB_Numeric Operations		-0.19***		-0.19***		-0.19***
		(-0.23,-0.15)		(-0.21,-0.17)		(-0.25,-0.13)
Bias from ASVAB_Auto and Shop Information		0.62***		0.60***		0.62*
		(0.20,1.04)		(0.56,0.64)		(-0.06,1.30)
Bias from ASVAB_Electronic Information		-0.17***		-0.14***		-0.18***
		(-0.21,-0.13)		(-0.22,-0.06)		(-0.26,-0.10)
Bias from ASVAB_Mechanical Comprehension		0.03		0.03***		0.04
		(-0.09,0.15)		(0.01,0.05)		(-0.14,0.22)
Bias from Rotter Scale		-0.01*		-0.02***		-0.02**
		(-0.03,0.01)		(-0.04,-0.01)		(-0.04,-0.01)
Bias from Rosenberg Scale		0.02***		0.03***		0.03***
		(0.01,0.04)		(0.01,0.05)		(0.01,0.05)

Note: Bias estimates are based on the common factor estimates in Table 3 Panel A.

Table 5 continued: Bias Associated with OLS Estimates, Due to Common Factor Structure and Proxies

	(1)	(2)	(3)	(4)	(5)	(6)
	IFE		CCE		CCE-2	
	PCA Proxy	Non-PCA Proxy	PCA Proxy	Non-PCA Proxy	PCA Proxy	Non-PCA Proxy
B. With demographic controls						
Total Bias	3.16***	3.92***	3.35***	4.00***	3.46***	4.11***
	(0.42,4.10)	(1.59,6.25)	(3.17,3.53)	(3.82,4.18)	(1.17,5.75)	(1.93,6.29)
Bias from Factor Loadings	3.11***	4.43***	3.29***	4.44***	3.41***	4.59***
	(0.41,5.81)	(2.10,6.76)	(3.11,3.40)	(4.28,4.60)	(1.12,5.70)	(2.56,6.54)
Bias from Proxies	0.05	-0.51***	0.05***	-0.44***	0.05	-0.48*
	(-0.05,0.15)	(-0.78,-0.24)	(0.03,0.07)	(-0.42,-0.46)	(-0.09,0.19)	(-0.02,0.98)
Bias from Cognitive Factor	0.01		0.01		0.01	
	(-0.02,0.09)		(-0.01,0.03)		(-0.13,0.15)	
Bias from Non-Cognitive Factor 1	0.04***		0.03***		0.04***	
	(0.02,0.06)		(0.01,0.05)		(0.02,0.06)	
Bias from Non-Cognitive Factor 2	0.00		0.00		0.00	
	(-0.01,0.02)		(-0.01,0.02)		(-0.01,0.02)	
Bias from Non-Cognitive Factor 3	0.00		0.00		0.00	
	(-0.01,0.02)		(-0.01,0.02)		(-0.01,0.02)	
Bias from ASVAB_General Science		-0.24***		-0.20***		-0.23***
		(-0.28,-0.20)		(-0.22,-0.18)		(-0.33,-0.13)
Bias from ASVAB_Arithmetic Reasoning		0.10***		0.10**		0.09*
		(0.06,0.14)		(-0.12,-0.08)		(-0.01,0.19)
Bias from ASVAB_Mathematical Knowledge		-0.55***		-0.52**		-0.53**
		(-0.73,-0.37)		(-0.54,-0.50)		(-0.94,-0.12)
Bias from ASVAB_Paragraph Comprehension		-0.11***		-0.09***		-0.10***
		(-0.07,-0.15)		(-0.11,-0.07)		(-0.14,-0.06)
Bias from ASVAB_Word Knowledge		0.19***		0.19***		0.18***
		(0.15,0.23)		(0.17,0.21)		(0.10,0.26)
Bias from ASVAB_Coding Speed		0.01		0.01		0.01
		(-0.01,0.03)		(-0.01,0.03)		(-0.01,0.03)
Bias from ASVAB_Numeric Operations		-0.01		-0.01		-0.01
		(-0.03,0.01)		(-0.03,0.01)		(-0.15,0.13)
Bias from ASVAB_Auto and Shop Information		0.01		0.04***		0.03
		(-0.09,0.11)		(0.02,0.06)		(-0.07,0.13)
Bias from ASVAB_Electronic Information		0.08**		0.04***		0.06
		(0.02,0.14)		(0.02,0.06)		(-0.02,0.14)
Bias from ASVAB_Mechanical Comprehension		-0.02*		-0.02*		-0.02*
		(-0.04,0.01)		(-0.04,0.01)		(-0.04,0.01)
Bias from Rotter Scale		0.01		0.00		0.00
		(-0.01,0.03)		(-0.01,0.03)		(-0.01,0.03)
Bias from Rosenberg Scale		0.03***		0.03***		0.03
		(0.01,0.05)		(0.01,0.05)		(-0.03,0.09)

Note: Bias estimates are based on the common factor estimates in Table 3 Panel B.

Table 6: Validation Tests for Proxies: $A(j)$ and $R^2(j)$ Statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IFE				CCE-2			
	$A(j)$	$R^2(j)$	$R^{2-}(j)$	$R^{2+}(j)$	$A(j)$	$R^2(j)$	$R^{2-}(j)$	$R^{2+}(j)$
A. Without demographic controls								
Cognitive Factor	0.8983	0.0045	0.0045	0.0046	0.9339	0.0024	0.0023	0.0024
Non-Cognitive Factor 1	0.9393	0.0027	0.0027	0.0027	0.9533	0.0014	0.0014	0.0014
Non-Cognitive Factor 2	0.8938	0.0061	0.0061	0.0062	0.9525	0.0018	0.0018	0.0019
Non-Cognitive Factor 3	0.9145	0.0038	0.0038	0.0039	0.9372	0.0022	0.0022	0.0022
ASVAB_General Science	0.9417	0.0029	0.0029	0.0030	0.9595	0.0013	0.0013	0.0013
ASVAB_Arithmetic Reasoning	0.9360	0.0048	0.0047	0.0048	0.9388	0.0036	0.0036	0.0036
ASVAB_Mathematical Knowledge	0.8983	0.0085	0.0085	0.0086	0.9417	0.0037	0.0037	0.0037
ASVAB_Paragraph Comprehension	0.9161	0.0039	0.0039	0.0039	0.9459	0.0021	0.0021	0.0021
ASVAB_Word Knowledge	0.9417	0.0042	0.0042	0.0042	0.9550	0.0028	0.0028	0.0028
ASVAB_Coding Speed	0.8926	0.0059	0.0059	0.0060	0.9310	0.0032	0.0031	0.0032
ASVAB_Numeric Operations	0.9066	0.0052	0.0052	0.0053	0.9388	0.0030	0.0030	0.0030
ASVAB_Auto and Shop Information	0.9401	0.0030	0.0029	0.0030	0.9521	0.0021	0.0021	0.0021
ASVAB_Electronic Information	0.9467	0.0026	0.0026	0.0026	0.9533	0.0016	0.0016	0.0016
ASVAB_Mechanical Comprehension	0.9364	0.0035	0.0035	0.0035	0.9446	0.0025	0.0025	0.0025
Rotter Scale	0.9351	0.0050	0.0049	0.0050	0.9471	0.0026	0.0026	0.0026
Rosenberg Scale	0.9545	0.0018	0.0018	0.0018	0.9669	0.0008	0.0008	0.0008
B. With demographic controls								
Cognitive Factor	0.8884	0.0062	0.0062	0.0063	0.9198	0.0022	0.0022	0.0022
Non-Cognitive Factor 1	0.9318	0.0028	0.0028	0.0028	0.9488	0.0007	0.0007	0.0007
Non-Cognitive Factor 2	0.9165	0.0040	0.0040	0.0040	0.9417	0.0017	0.0017	0.0017
Non-Cognitive Factor 3	0.9140	0.0048	0.0048	0.0048	0.9236	0.0021	0.0021	0.0021
ASVAB_General Science	0.7905	0.0444	0.0436	0.0451	0.8554	0.0352	0.0346	0.0357
ASVAB_Arithmetic Reasoning	0.8037	0.0497	0.0488	0.0506	0.8388	0.0454	0.0446	0.0461
ASVAB_Mathematical Knowledge	0.8008	0.0411	0.0405	0.0418	0.8500	0.0342	0.0337	0.0347
ASVAB_Paragraph Comprehension	0.8362	0.0210	0.0208	0.0212	0.8727	0.0150	0.0149	0.0152
ASVAB_Word Knowledge	0.8223	0.0401	0.0394	0.0407	0.8690	0.0311	0.0307	0.0316
ASVAB_Coding Speed	0.8707	0.0093	0.0092	0.0093	0.9145	0.0049	0.0048	0.0049
ASVAB_Numeric Operations	0.8442	0.0189	0.0187	0.0192	0.8694	0.0157	0.0155	0.0158
ASVAB_Auto and Shop Information	0.7864	0.0577	0.0566	0.0588	0.8326	0.0503	0.0494	0.0512
ASVAB_Electronic Information	0.7946	0.0510	0.0501	0.0519	0.8397	0.0459	0.0452	0.0467
ASVAB_Mechanical Comprehension	0.7591	0.0572	0.0562	0.0583	0.8190	0.0517	0.0508	0.0526
Rotter Scale	0.9211	0.0077	0.0076	0.0077	0.9355	0.0049	0.0049	0.0049
Rosenberg Scale	0.9281	0.0050	0.0049	0.0050	0.9632	0.0012	0.0012	0.0012

Table 7: Validation Tests for Proxies: Canonical Correlations

	(1)	(2)	(3)	(4)	(5)	(6)
		IFE			CCE-2	
m	$\hat{\rho}_m^2$	$\hat{\rho}_m^{2-}$	$\hat{\rho}_m^{2+}$	$\hat{\rho}_m^2$	$\hat{\rho}_m^{2-}$	$\hat{\rho}_m^{2+}$
A. Without demographic controls						
1	0.1141	0.0903	0.1380	0.1060	0.0828	0.1292
2	0.0469	0.0304	0.0633	0.0186	0.0079	0.0292
3	0.0159	0.0060	0.0258	0.0143	0.0049	0.0237
4	0.0117	0.0032	0.0202	0.0082	0.0010	0.0153
5	0.0098	0.0020	0.0176			
6	0.0045	0.0000	0.0098			
7	0.0042	0.0000	0.0093			
8	0.0032	0.0000	0.0077			
9	0.0007	0.0000	0.0029			
B. With demographic controls						
1	0.3844	0.3540	0.4148	0.4196	0.3897	0.4496
2	0.0577	0.0396	0.0757	0.0254	0.0130	0.0378
3	0.0161	0.0061	0.0260	0.0109	0.0027	0.0191
4	0.0122	0.0035	0.0208	0.0070	0.0004	0.0136
5	0.0095	0.0018	0.0171			
6	0.0069	0.0003	0.0135			
7	0.0044	0.0000	0.0097			
8	0.0027	0.0000	0.0069			

Note: $m = 1, 2, ..., min[k; r]$ where k is the number of ability proxies, and r is the number of common components. $\hat{\rho}_m^2$ is the canonical correlation of order m between the estimated factor loadings and proxies, and $\hat{\rho}_m^{2-}$ and $\hat{\rho}_m^{2+}$ define the 95% confidence interval.

Appendix A: Wild Cluster Bootstrap Confidence Intervals

The confidence intervals for the IFE and CCEP-2 estimator described in Sections 2.1.1 and 2.1.2, respectively, are estimated using a wild cluster bootstrap procedure. An attractive feature of this procedure is that it is not only robust to heteroskedasticity but also preserves the time series dependence of the errors within clusters. Here we provide the details involved in implementing this procedure for the IFE estimator with analogous steps applicable to the CCEP-2 estimator.

Consider the model

$$\begin{aligned} y_{it} &= c_i + s_{it}\beta_1 + s_{it}^2\beta_2 + e_{it}\gamma_1 + e_{it}^2\gamma_2 + s_{it}e_{it}\gamma_3 + p_i'\delta_t + \lambda_i' f_t + u_{it} \qquad \text{(A.1)} \\ i &= 1, ..., N; \quad t = t_i \in \mathcal{J}_i \equiv \{t_i(1), t_i(2), ..., t_i(T_i)\} \end{aligned}$$

The steps for computing the bootstrap standard errors are enumerated as follows:

1. Obtain the IFE estimator from estimating model (A.1) using the procedure described in Section 2.1.1 and save the residuals $\{\hat{u}_{it};\ i = 1, ..., N,\ t \in \mathcal{J}_i\}$.

2. Generate bootstrap errors as

$$u_i^* = w_i \hat{u}_i, \quad w_i \overset{iid}{\sim} N(0,1)$$

   where $\hat{u}_i = (\hat{u}_{i,t_i(1)}, ..., \hat{u}_{i,t_i(T_i)})'$.

3. Generate bootstrap data from

$$y_{it}^* = \hat{c}_i + s_{it}\hat{\beta}_1 + s_{it}^2\hat{\beta}_2 + e_{it}\hat{\gamma}_1 + e_{it}^2\hat{\gamma}_2 + s_{it}e_{it}\hat{\gamma}_3 + p_i'\hat{\delta}_t + \hat{\lambda}_i'\hat{f}_t + u_{it}^*$$

   where "^" denotes the IFE estimate and u_{it}^* are the bootstrap errors obtained from step 2.

4. Estimate (A.1) using y_{it}^* instead of y_{it} and obtain the bootstrap estimates $\left(\hat{\beta}_{1,b}, \hat{\beta}_{2,b}, \hat{\gamma}_{3,b}\right)$. Use these estimates to compute the bootstrap estimate of the marginal return to schooling as

$$\widehat{MRTS}_b = \hat{\beta}_{1,b} + 2\bar{s}\hat{\beta}_{2,b} + \bar{e}\hat{\gamma}_{3,b}$$

5. Repeat steps (2)-(4) B times to obtain the set $\{\widehat{MRTS}_b\}_{b=1}^{B}$.

6. A $100(1-\alpha)\%$ equal-tailed confidence interval for the average marginal returns to schooling is constructed as

$$\left[\widehat{MRTS}(\alpha/2),\ \widehat{MRTS}(1-\alpha/2)\right]$$

where $\widehat{MRTS}(\alpha/2)$ and $\widehat{MRTS}(1-\alpha/2)$ are the $\alpha/2$ and $(1-\alpha/2)$ quantiles of the ordered collection $\{\widehat{MRTS}_b\}_{b=1}^{B}$, respectively.

Appendix B: Schooling Variable

Table B.1: Variation in Years of Schooling for the Final Sample

A. Changes Per Person

Total Number of Changes	Number of Individuals	Percent (in %)
0	1591	65.74
1	467	19.30
2	206	8.51
3	87	3.60
4	44	1.82
5	17	0.70
6	3	0.12
7	3	0.12
8	2	0.08

B. Changes Per Education Level

Education Level after Schooling Change	Number of Instances	Percent (in %)
9	3	0.21
10	8	0.55
11	80	5.49
12	249	17.10
13	211	14.49
14	227	15.59
15	172	11.81
16	242	16.62
17	125	8.59
18	80	5.49
19	38	2.61
20	21	1.44

Source: National Longitudinal Survey of Youth (NLSY79), 1981-2016.
Note: Panel A presents the number of schooling changes per person across our final sample. Panel B tabulates the schooling changes by education level after the change. For example, in the first row, we observed 3 instances of an individual's years of schooling changing to 9th grade, and 8 instances changing to 10th grade in the second row, and so on. Thus, Panel B indicates that most of the variation in schooling in our final sample comes from individuals completing grades years 12-17 (i.e., completing high schooling through 5 years of college).

Appendix C: Robustness Checks

Table C.1: IFE Estimates in % (Robustness) of the Return to Schooling for White Males

Number of Factors	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A. Without demographic controls, No proxy										
Marginal return	9.58***	4.54***	4.10***	5.02***	5.28***	5.24***	5.39***	5.29***	**5.32*** **	5.31***
	(6.69,10.39)	(2.31,9.45)	(2.50,8.55)	(2.68,8.20)	(2.23,8.29)	(1.68,7.96)	(1.39,8.12)	(1.30,8.00)	(1.02,7.74)	(1.36,7.58)
B. Without demographic controls, PCA proxy										
Marginal return	8.77***	4.61***	3.98***	5.00***	5.24***	5.19***	5.35***	5.25***	**5.28*** **	5.27***
	(5.90,9.76)	(2.43,8.98)	(2.32,7.95)	(2.48,8.19)	(2.10,7.84)	(1.67,7.82)	(1.35,7.76)	(1.21,7.66)	(1.03,7.36)	(1.17,7.02)
C. Without demographic controls, Non-PCA proxy										
Marginal return	6.77***	4.10***	2.52**	4.62***	4.95***	4.88***	4.99***	4.98***	**4.95*** **	4.96***
	(3.97,8.77)	(2.82,8.75)	(0.96,7.08)	(2.41,7.66)	(1.90,7.35)	(1.77,7.20)	(1.49,7.20)	(1.22,7.32)	(1.40,6.99)	(1.47,6.67)
D. With demographic controls, No proxy										
Marginal return	6.76***	5.09***	4.75***	3.00***	4.60***	4.92***	4.97***	**5.05*** **	5.06***	5.09***
	(5.01,10.04)	(3.26,9.87)	(1.09,7.75)	(2.67,7.93)	(2.12,7.32)	(2.16,7.47)	(1.96,7.77)	(1.77,7.60)	(1.98,7.53)	(2.02,7.00)
E. With demographic controls, PCA proxy										
Marginal return	6.88***	4.51***	2.90***	4.45***	4.79***	4.80***	4.88***	**4.88*** **	4.88***	4.94***
	(5.11,10.02)	(3.56,9.21)	(1.07,7.23)	(2.52,7.69)	(2.17,7.22)	(1.98,7.10)	(1.91,7.64)	(1.87,7.45)	(1.98,6.74)	(1.81,6.83)
F. With demographic controls, Non-PCA proxy										
Marginal return	5.24***	3.75***	2.39**	3.80***	**4.13*** **	4.14***	4.23***	4.24***	4.23***	4.28***
	(3.96,8.67)	(2.19,8.55)	(0.75,6.64)	(2.15,6.92)	(1.76,6.71)	(1.78,6.82)	(1.52,6.69)	(1.52,6.69)	(1.71,6.69)	(1.64,6.26)

Source: National Longitudinal Survey of Youth (NLSY79), 1981-2016.
Note: The dependent variable is the log of hourly wage adjusted to 1999 dollars. Our final sample contains 2,420 individuals for a total of 33,248 person-year observations. 95% confidence intervals are shown in parentheses. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*. Person fixed-effects and age controls are controlled for in all specifications (see Table 2 for details). Panel D, E, and F present estimates with demographic controls including mother's and father's education levels, and the number of siblings. The number of factors are chosen based on the criterion described in section 2.1.1, and the selected estimates are highlighted in bold.

Table C.2: Estimates of the Return to Schooling (in %) for White Males (Robustness with Age Quartic)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	OLS	OLS	OLS	IFE	IFE	IFE	CCE	CCE	CCE	CCE-2	CCE-2	CCE-2
A. Without demographic controls												
Marginal return	7.33***	6.89***	5.54***	5.11***	5.00***	4.76***	4.97***	4.92***	4.62***	4.91***	4.85***	4.54***
	(5.52,9.14)	(5.08,8.70)	(3.72,7.36)	(0.89,6.52)	(1.34,7.12)	(0.99,6.58)	(2.07,7.87)	(2.02,7.82)	(1.56,7.68)	(3.23,8.36)	(3.38,8.40)	(3.48,8.80)
Demos-by-year FE	No	No	No	No	No	No	No	No	No	No	No	No
PCA Proxies-by-year FE	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No
Non-PCA Proxies-by-year FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Adjusted R-squared	0.3346	0.338	0.3497	0.8529	0.7022	0.6289	0.6604	0.6423	0.4706	0.6705	0.6519	0.4769
B. With demographic controls												
Marginal return	7.06***	6.46***	5.52***	5.09***	4.87***	4.23***	4.92***	4.71***	4.04**	4.82***	4.63***	3.96***
	(5.21,8.90)	(4.63,8.30)	(3.67,7.36)	(1.93,7.21)	(1.80,6.83)	(1.50,6.37)	(1.76,8.08)	(1.57,7.85)	(0.88,7.20)	(4.15,9.16)	(3.97,8.90)	(3.34,8.19)
Demos-by-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
PCA Proxies-by-year FE	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No
Non-PCA Proxies-by-year FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Adjusted R-squared	0.3406	0.3444	0.3527	0.5279	0.5254	0.5114	0.3703	0.369	0.3583	0.3749	0.3736	0.3614

Source: National Longitudinal Survey of Youth (NLSY79), 1981-2016.

Note: The dependent variable is the log of hourly wage adjusted to 1999 dollars. Our final sample contains 2,420 individuals for a total of 33,248 person-year observations. 95% confidence intervals are shown in parentheses. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*. Person fixed-effects and (quartic) age controls are controlled for in all specifications. Panel B presents estimates with demographic controls including mother's and father's education levels as well as the number of siblings. Columns (4)-(6) are based on Interactive Fixed Effects (IFE) (Bai, 2009) with the estimated number of factors to be 9 in Panel A, 8 in Panel B Columns (4)-(5), and 5 in Panel B Columns (6). Columns (7)-(9) are based on Common Correlated Effects (CCE) (Pesaran, 2006). Columns (10)-(12) are based on the two-step CCE procedure with the number of factors set to be 4 in both panels.