Multidimensional Skills and the Returns to Schooling: Evidence from an Interactive Fixed Effects Approach and a Linked Survey-Administrative Dataset*†

Mohitosh Kejriwal *Purdue University*

Xiaoxiao Li *Villanova University*

Evan Totty *U.S. Census Bureau*

November 13, 2019

Purdue University Economics Department Working Paper No. 1316

Abstract

This paper presents new evidence on the returns to schooling based on an interactive fixed effects framework that allows for multiple unobserved skills with associated prices that are potentially time-varying. Skills and prices are both allowed to be correlated with schooling. The modeling approach can also accommodate individual-level heterogeneity in the returns to schooling. The framework thus constitutes a substantive generalization of most existing approaches that assume ability is unidimensional and/or returns are homogeneous. Our empirical analysis employs a unique panel dataset on earnings and education 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). Our preferred specification yields a point estimate of the average marginal returns to schooling of about 2.8-4.4 percent relative to ordinary least squares and two stage least squares estimates which lie in the range 7.7-12.7 percent. A decomposition of the aggregate least squares bias shows that the omitted ability component is responsible for a larger fraction of the bias relative to the heterogeneity component. Finally, our heterogeneity analysis suggests larger returns for individuals born in more recent years, the presence of sheepskin effects, and considerable within-group heterogeneity.

*We thank the editor and anonymous referees for their thoughtful comments and efforts towards improving our manuscript. We are grateful to Timothy N. Bond, Chao Fu, Christopher Kilby, Junsoo Lee, Jie Ma, Peter Orazem, Miguel Sarzosa, Bryan Stuart, Justin Tobias, John Voorheis, and participants at a number of workshops and conferences for helpful comments. We are thankful for Xuewen Yu and Haiqing Zhao's exceptional research assistance. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the Census Bureau or other organizations. All results have been reviewed and approved by the Census Bureau disclosure review board: CBDRB-FY19-CED001-B0014, CBDRB-FY19-CED001-B0025.

†Contact information: Evan Totty (corresponding author), evan.scott.totty@census.gov; Mohitosh Kejriwal, mkejriwa@purdue.edu; Xiaoxiao Li, xiaoxiao.s.li@villanova.edu

1 Introduction

The human capital hypothesis (Becker, 1962) states that in a competitive market, higher education leads to higher human capital and therefore higher wages. This hypothesis has led to decades of empirical discussion on the average marginal return to education based primarily on the Mincer regression (Mincer, 1974). The debate has centered around the omitted ability bias, with the assumption being that ordinary least squares (OLS) estimates of the growth rate of earnings with schooling are likely to be overstated due to the positive association between earnings and ability as well as ability and schooling (Griliches, 1977). In an attempt to correct for the potential upward bias, a large body of empirical work has emerged over the past four decades that adopted various econometric strategies to account for the endogeneity of schooling which could potentially deliver a reliable estimate of the returns to schooling. Such strategies include the use of instrumental variables (IV) estimates (e.g., Angrist and Krueger, 1991), utilizing within family variation in schooling (e.g., Ashenfelter and Krueger, 1994), and the use of observable proxies for ability (e.g., Heckman, Stixrud, and Urzua, 2006). However, each strategy suffers from its own set of issues and collectively they produce conflicting and sometimes surprising results (Card, 2001; Heckman, Lochner, and Todd, 2006; Caplan, 2018). This has led to a call for new panel data approaches utilizing large administrative datasets (Heckman, Lochner, and Todd, 2006; Altonji, 2010).

This paper adopts an interactive fixed effects or common factor framework for estimating the returns to schooling that allows for multiple unobserved skills with associated prices that are potentially time-varying. The skills are represented by the factor loadings while their prices are represented by the common factors. Additive individual and/or time fixed effects are obtained as special cases of this framework. Skills and prices are both allowed to be correlated with schooling which addresses the endogeneity of the latter without resorting to external instruments or proxies for ability. The modeling approach can also accommodate individual-level heterogeneity in the returns to schooling. The framework thus constitutes a substantive generalization of most existing approaches that assume ability is unidimensional and/or returns are homogeneous. Moreover, it allows us to quantify two important sources of bias: one from ignoring the interactive fixed effects structure (the ability bias) and the other from ignoring potential parameter heterogeneity. Estimation is carried out using the methods developed by Bai (2009), Pesaran (2006), and Song (2013) that facilitate consistent estimation of the growth rate of earnings with schooling and enable statistical inference via asymptotically valid standard errors.

Using a common factor structure to model the earnings function is, however, not new. Hause (1980) employs an interactive effects framework (referring to it as "the fine structure of earnings") to decompose the covariance matrix of earnings time series into ability and on-the-job training components and evaluate the empirical significance of the latter. Heckman and Scheinkman (1987) employ a multifactor model for earnings in order to test the hypothesis of uniform pricing across sectors of the economy. More recently, Carneiro et al. (2003) use the common factor structure as a dimension reduction tool to model the dependence across unobservable ability components and estimate counterfactual distributions of outcomes while Heckman, Stixrud, and Urzua (2006) show that a low-dimensional vector of latent cognitive and non-cognitive skills modeled using a factor structure explains a variety of behavioral and labor market outcomes (see also Heckman et al., 2017). Westerlund and Petrova (2017) apply the interactive fixed effects framework to the returns to schooling and find smaller returns than OLS. However, their analysis was an empirical illustration of the performance of Pesaran's (2006) common correlated effects estimator under asymptotic collinearity, and leaves much room for work.[1] Our contribution differs from these studies in that we exploit the time series variation in schooling over the sample period as well as the high-dimensional nature of the panel dataset to simultaneously address the twin issues of heterogeneity in returns to schooling and the endogeneity of schooling thus enabling us to disentangle the biases associated with ignoring one or both of these features.

[1]This includes the use of a larger dataset, additional estimators (Bai, 2009; Song, 2013), a variety of specifications to account for heterogeneity and experience, relation of the results to both the IV and the ability proxy literature, and accommodation of individual-level heterogeneity in the returns to schooling, all of which we address in this paper.

Our empirical analysis employs a unique panel dataset on earnings and education 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). Combining nine SIPP survey panels and administrative earnings data provides a panel dataset that is of high quality, has a long time dimension, and includes a large number of individuals. Administrative data on earnings is advantageous relative to survey data due to rising measurement error and non-response in survey data (Abowd and Stinson, 2013; Meyer et al., 2015). This is particularly relevant for estimating the returns to schooling, given that the nature of earnings misreporting in survey data tends to vary with earnings and education levels (Chenevert et al., 2016; Cristia and Schwabish, 2009; Pedace and Bates, 2000). The linked dataset has a much larger time dimension and cross-section dimension than in the few existing panel studies on returns to schooling, which usually rely on the Panel Study of Income Dynamics (PSID) or the National Longitudinal Study of Youth (NLSY) (e.g., Angrist and Newey, 1991; Ashworth et al., 2017; Koop and Tobias, 2004; Westerlund and Petrova, 2017).[2]

Previewing our results, we first replicate the well established finding in the literature that the IV estimate of the growth rate of earnings due to schooling is larger than the corresponding OLS estimate, both using cross-section and panel data. The IV estimate is based on using the quarter of birth interacted with the year of birth as instruments following Angrist and Krueger (1991). Next, our interactive fixed effects estimates are found to be considerably smaller than the OLS estimates, regardless of whether a pooled or heterogenous model is estimated. Our preferred specification based on models with heterogeneous coefficients yields a point estimate of the average marginal returns to schooling of about 2.8-4.4 percent relative to OLS and two stage least squares (2SLS) estimates which lie in the range 7.7-12.7 percent. While both omitted ability and heterogeneity biases contribute to the overall OLS bias, a decomposition of the aggregate least

[2]Two other recent examples of panel analysis use administrative data from Norway and Sweden (Bhuller et al., 2017; Nybom, 2017).

squares bias shows that the omitted ability component is responsible for a larger fraction of the bias relative to the heterogeneity component. Overall, our results are more similar to the ability proxy literature, which finds smaller returns than OLS, than the IV literature, although we find even larger positive bias and smaller marginal returns. Lastly, we analyze both across-group and within-group heterogeneity in the returns to schooling. Although we find minimal evidence of heterogeneous returns across race, Hispanic status, or foreign born status, our results indicate that returns are larger for individuals born in more recent years. Our findings are also suggestive of "sheepskin effects" and considerable heterogeneity within demographic groups and education levels.

The rest of the paper is organized as follows. Section 2 discusses issues related to the existing econometric strategies in the literature. Section 3 introduces the interactive effects framework including a brief description of the associated estimation methods. Section 4 details the administrative data used to conduct the empirical analysis. Section 5 presents the estimated specifications and results. Section 6 concludes. Appendices provide detailed derivations and additional empirical results, including robustness to the inclusion of higher order terms for schooling and experience.

2 Issues in the Existing Literature

In order to motivate the approach taken in this paper, it is useful to first highlight the issues associated with the different econometric strategies that have been employed in the literature to correct for the omitted ability bias inherent in OLS estimates of the returns to schooling. These issues have turned out to be of considerable importance from an empirical standpoint and have contributed to a general lack of consensus about the appropriate methodology to adopt when estimating the returns to schooling. We first discuss the two main approaches that are based on utilizing cross-sectional data: the IV approach and the ability proxy approach. This is followed by an assessment of existing panel data studies including a discussion of the relative advantages of our approach which should further help delineate our contribution to the literature.

The IV approach is based on exploiting natural variation in the data caused by exogenous influences on the schooling decision. For instance, the seminal study of Angrist and Krueger (1991) uses an individual's quarter of birth (interacted with year of birth or state of birth in some specifications) as an instrument for schooling based on the observation that compulsory schooling laws tend to lead individuals born earlier in the year to have less schooling relative to those born later in the year. Surprisingly, however, the IV estimates were found to be consistently larger than the OLS estimates thereby presenting an empirical puzzle regarding the interpretation of the IV estimates (see Card, 2001, Table II, for a summary of this literature). One potential explanation for the larger IV estimates is in terms of the Local Average Treatment Effect (LATE) on a selected sample (Imbens and Angrist, 1994). That is, if the instrument has a larger impact on individuals with higher marginal returns to schooling, the IV procedure will tend to produce an overestimate of the average marginal returns to education.[3] Heckman, Lochner, and Todd (2006) and Heckman, Urzua, and Vytlacil (2006), however, point out that the LATE interpretation of the IV estimate assumes away heterogeneity in the response of schooling choices to instruments, via the monotonicity assumption. Card (2001) discusses other explanations for the puzzle including attenuation bias in the OLS estimates due to measurement error in schooling, short term credit constraints and specification search bias.[4,5] Carneiro and Heckman (2002) argue, using AFQT as a measure of ability, that the observed pattern of results can simply be a consequence of using poor or invalid instruments that are either only weakly correlated with schooling or correlated with ability. Heckman, Lochner, and Todd (2006) conclude in their survey of the literature that the IV approach is of limited use in uncovering a reliable estimate of the returns to schooling.[6]

[3]Note, the LATE issue arises when the marginal returns are heterogeneous, which could occur either when the relationship between earning and schooling is linear, but the coefficients are individual specific; or when the coefficients are homogeneous, but schooling enters the model nonlinearly.

[4]Card (2001) notes that measurement error in schooling cannot explain the observed difference in OLS and IV estimates while Carneiro and Heckman (2002) show that IV can exceed OLS even in the absence of credit constraints.

[5]Oreopoulos (2006) approximated the average treatment effect by looking at compulsory schooling policy change that affected a large group of people in U.K. and suggested that even when the sample is not subject to selection problems and credit constraints, the IV estimate is still larger than OLS and therefore the empirical puzzle remains.

[6]As one of the referees pointed out, criticism over the validity of the IV strategy depends on the choice of the

The ability proxy approach employs observable proxies for ability in order to mitigate the impact of the ability bias. Common proxies for cognitive ability include GPA, AFQT scores and other components in the ASVAB tests while those for non-cognitive ability include the Rotter Locus of Control Scale which measures the degree of control individuals feel they possess over their life and the Rosenberg Self-Esteem Scale which measures perceptions of self-worth (Heckman, Stixrud, and Urzua, 2006).[7] Heckman et al. (2017) 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 percent smaller, depending on the specification. Similar reductions are reported by Ashworth et al. (2017) in comparing the basic Mincer regressions to regressions that include ability proxies and actual experience using the NLSY panel data.[8] A major challenge facing this literature is that the ability proxies, particularly those measuring non-cognitive ability or "soft skills" such as conscientiousness, conformity, self-esteem, etc., are far from perfect resulting in biased estimates of the schooling effect (Heckman, Stixrud, and Urzua, 2006).[9] Our paper contributes to the literature by providing a rigorous framework that allows the data to speak regarding the importance of multi-dimensional abilities without relying on imperfect proxies. Our preferred specification based on models with heterogeneous coefficients suggests a reduction in the average marginal returns to schooling between 44-64 percent relative to OLS.

In contrast to the cross-section methods, the panel data approach identifies the effect of schooling based on time-series variation within individuals. Angrist and Newey (1991) and Koop and Tobias (2004) use panel data from the NLSY to estimate the returns to schooling (more precisely, the percentage growth rate of earnings due to schooling) although their modeling approaches are different. Both studies, however, assume that individual fixed effects can effectively capture

instrument so that making general statements about the IV estimates of the returns to schooling is difficult.

[7]While GPA is commonly used as a measure of cognitive ability, there is evidence indicating that GPA captures a mix of cognitive and non-cognitive skills (Humphries and Kosse, 2017).

[8]Based on reviewing the earlier evidence, Caplan (2018, Chapter 3) suggests that cognitive ability bias is between 20-30 percent while non-cognitive ability bias is between 5-15 percent. He interprets the ability bias in the literature as a lower bound on the true bias due to the imperfect measure of abilities, especially the non-cognitive abilities.

[9]Even cognitive ability measures, such as AFQT scores, are subject to criticism (Polachek et al., 2015).

the potential endogeneity of schooling. Angrist and Newey (1991) employ a standard panel data framework with homogeneous coefficients where unobserved heterogeneity is controlled for using individual and time fixed effects. They find that the fixed effects estimates are roughly twice as large as the OLS estimates which runs counterintuitive to the notion that ability bias tends to overstate the OLS returns and suggests that individual fixed effects are not sufficient to control for the potential upward bias. Koop and Tobias (2004) address the issue of cross-sectional heterogeneity in returns adopting a Bayesian framework to characterize the nature of such heterogeneity. Comparing results across a wide variety of specifications, they find strong evidence in favor of models that allow for heterogeneous slopes. Our modeling approach is considerably more general than those adopted in these studies in that we allow for multidimensional abilities with possibly time-varying prices as well as cross-sectional heterogeneity in the growth rate of earnings with schooling. In addition, our empirical analysis uses a linked survey-administrative dataset which offers important advantages over survey-based data.

A potential drawback of the panel data approach is that it requires a sample of individuals with continuous earnings while increasing schooling. This may include, for example, traditional students who also work while obtaining a bachelors degree or individuals who return to school later in life, whether to finish an uncompleted degree or for additional degrees. This sample could be different from the traditional idea of a student who completes degrees consecutively and does not work while in school. Setting aside sample selection effects, there could also be issues comparing time-series earnings before, during, and after schooling, since earnings before or during schooling could be part-time or seasonal work and not truly reflect an individual's earning potential (Card, 1995; Lazear, 1977). That said, we believe these concerns are mitigated somewhat by the facts that: (1) we do replicate well-established results in terms of the absolute and relative magnitude of OLS and IV estimates from the cross-section literature; (2) we set annual minimum earnings restrictions equal to the federal minimum wage multiplied by 800 hours, following the criterion adopted in Koop and Tobias (2004); (3) we find similar sample statistics and cross-section estimates if we instead use a sample that does not require continuous earnings while in school; and (4) 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. Furthermore, unlike the cross-section approach, the use of panel data allows us to formally test for heterogeneity in the returns to schooling as well as explore its nature across and within subgroups.

3 Empirical Framework

This section presents the interactive fixed effects framework that forms the basis of our empirical analysis aimed at estimating the growth rate of earnings with years of schooling. Conditional on the common factor structure embedded in the framework that represents multiple skills with time varying prices, one can further derive not only the aggregate OLS and IV biases but also provide a decomposition of the biases in terms of their omitted ability and heterogeneity components. Section 3.1 lays out the modeling framework including a description of the alternative estimation approaches. Section 3.2 discusses the intuition underlying the omitted ability and heterogeneity biases while Appendix A outlines the derivations and details regarding the computation of the two sources of bias. A potential explanation for the pattern of results obtained from the empirics can be given based on these derivations.

3.1 The Interactive Fixed Effects Model

The general interactive fixed effects model with heterogeneous coefficients is specified as

$$y_{it} = c_i + s_{it}\beta_i + w_{it}'\gamma_i + v_{it} \tag{1}$$

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

where y_{it} and s_{it} represent, respectively, the (log of) annual earnings and the years of schooling completed for person $i = 1,...,N$ at period $t = 1,...,T$, and w_{it} is a vector of observable characteristics that influence wages and are potentially correlated with education (e.g., experience). We include a set of person fixed effects c_i to control for time-invariant person characteristics such as gender and race. The parameter β_i measures the percentage change in annual earnings for person i due to an additional year of schooling. This parameter does not necessarily represent an internal rate of return to schooling unless the only costs of schooling are earnings foregone, and markets are perfect (Heckman, Lochner, and Todd, 2006). The error term v_{it} 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.[10] Both loadings and the factors are potentially correlated with the observables (s_{it}, w_{it}). The number of common components r is assumed unknown. The object of interest is the average marginal return $[E(\beta_i)]$ in the population. Note that while the returns to each of the skill components $(\lambda_i' f_t)$ are identified, the skills and their prices are not separately identified.[11] That is, the estimated factors and their loadings only estimate a rotation of the underlying true parameters and so cannot be given a direct economic interpretation. Unlike Heckman, Stixrud, and Urzua (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 rate of growth of earnings with schooling 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.

Various panel data specifications used in the literature can be obtained as special cases of (1) and (2). The standard panel data model with person and time fixed effects considered by

[10]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.

[11]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 $\Lambda^* = (\lambda_1^*, ..., \lambda_N^*)'$ where $F^* = FA$ and $\Lambda^* = \Lambda A^{-1\prime}$.

Angrist and Newey (1991) is obtained by setting $\beta_i = \beta,\ \gamma_i = \gamma,\ \lambda_i = \lambda$. Koop and Tobias (2004) consider a restricted version of (1) and (2) that allows heterogeneity in returns to schooling but assumes that the endogeneity of schooling (i.e., the ability bias) is fully accounted for by the individual fixed effects c_i. Thus, their model does not allow for multiple skill components with time varying prices. We consider estimating model (1) and (2) using two alternative econometric procedures: the principal components approach (Bai, 2009; Song, 2013) and the common correlated effects approach (Pesaran, 2006). We now briefly describe each of these methods.

3.1.1 The Principal Components Approach

Bai (2009) advocates an iterative principal components approach that treats the common factors and their loadings as parameters which are jointly estimated with the regression coefficients assuming cross-sectional homogeneity of the latter. Under both large N and large T, the estimator is shown to be $\sqrt{NT}$-consistent and asymptotically normal under mild conditions on the idiosyncratic components that allow for (weak) correlation and heteroskedasticity in both dimensions. To ensure that the asymptotic distribution is centered around zero, a bias corrected estimator is proposed. Our empirical analysis employs the bias corrected estimator which we refer to as the interactive fixed effects (IFE) estimator.

Let $x_{it} = (s_{it}{}', w_{it}{}')'$ and $\widetilde{x}_{it} = x_{it} - T^{-1}\sum_{t=1}^{T} x_{it}$ with $\widetilde{y}_{it}$ defined analogously. Letting $\phi = (\beta', \gamma')'$, the IFE estimator is obtained by iteratively solving the following pair of equations:

$$\hat{\phi}\left[\{f_t\}_{t=1}^{T}, \{\lambda_i\}_{i=1}^{N}\right] = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{x}_{it}'\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}(\widetilde{y}_{it} - \lambda_i' f_t)\right) \quad (3)$$

$$\widetilde{y}_{it} - \widetilde{x}_{it}'\phi = \lambda_i' f_t + \widetilde{u}_{it} \quad (4)$$

In particular, given the factors and loadings, we compute $\hat{\phi}$ from (3) and given ϕ, we compute the factors and loadings from (4) using principal components. Two choices of initial values are employed to start the iteration and the one that leads to the lower sum of squared residuals upon convergence is chosen as the final set of estimates. The first choice sets $f_t = 0$ for all t in (3) while the second sets $\phi = 0$ in (4). The tolerance level for the convergence of the sum of squared residuals was set at 10^{-10} and convergence was achieved within 500 iterations across all estimated specifications.

Song (2013) develops a heterogeneous version of the IFE estimator that allows the regression coefficients to be unit-specific. The estimator is obtained by taking the cross-sectional average of the individual specific IFE estimates and is shown to be $\sqrt{N}$-consistent for the average return in the population. We refer to this estimator as the IFEMG (MG denoting mean group) estimator.

Both the IFE and IFEMG estimators require a choice on the number of common factors. Bai (2009) proposes estimating the number of factors employing the information criterion procedure of Bai and Ng (2002). Specifically, the number of factors is obtained by minimizing the criterion

$$IC(k) = \ln\left[(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^{2}(k)\right] + k\left(\frac{N+T}{NT}\right)g(N,T)$$

over $k \in [0, k_{\max}]$, where $k_{\max}$ is a prespecified upper bound. The residuals $\{\hat{u}_{it}(k)\}$ are obtained from principal components estimation assuming k factors and $g(N,T)$ is a penalty function. When estimating a pooled model as in Bai (2009), the IFE estimate is used to construct the residual series while estimating a heterogeneous version as in Song (2013) entails the use of the individual level IFE estimate. We set $k_{\max} = 10$ and use $g(N,T) = \ln\left(\frac{NT}{N+T}\right)$ which corresponds to the "IC_{p1}" criterion in Bai and Ng (2002).

3.1.2 The Common Correlated Effects (CCE) Approach

Pesaran (2006) proposes to proxy for the unobserved common factors f_t using cross-sectional averages of the dependent and independent variables, i.e., unlike the principal components approach, the factors are treated as nuisance parameters rather than parameters of interest. Estimation is based on augmenting the regression (1) with the cross-sectional averages and does not require knowledge of the number of factors. Two estimators are suggested: (1) the common correlated effects mean group (CCEMG) estimator which allows for heterogeneous coefficients and is obtained by estimating person-specific time series regressions using OLS and taking the average of the person-specific estimates; (2) the common correlated effects pooled (CCEP) estimator which pools the observations over the cross-section units and achieves efficiency gains when the slope parameters are the same across units.[12]

Based on a random coefficients formulation for the regression coefficients as well as the factor loadings, both estimators are shown to be $\sqrt{N}$-consistent and asymptotically normal as the cross-section dimension (N) and the time series dimension (T) jointly diverge to infinity. The finite sample performance of both estimators can be sensitive to a particular rank condition which requires that the number of factors does not exceed the total number of observed variables (see the Monte Carlo evidence in section 7 of Pesaran, 2006).

Pesaran (2006, p.1000) also suggests a two-step approach to estimation that involves combining the CCE and principal components approaches. For the model specified in (1) and (2), the first step entails obtaining the residuals

$$\hat{v}_{it} = y_{it} - \hat{c}_i - s_{it}\hat{\beta}_i - w_{it}'\hat{\gamma}_i$$

where $(\hat{c}_i, \hat{\beta}_i, \hat{\gamma}_i)'$ denote the individual level CCE estimates. The factors are then estimated by principal components treating the residuals as observed data where the number of factors is again selected based on the information criterion discussed in section 3.1.1. In the second step, the factor estimates (say $\{\hat{f}_t\}_{t=1}^T$) are then directly used as regressors in the regression equation

$$y_{it} = c_i + s_{it}\beta_i + w_{it}'\gamma_i + \lambda_i'\hat{f}_t + \xi_{it} \tag{5}$$

[12]The coefficients of the cross-sectional averages are, however, allowed to be individual-specific in both the pooled and heterogeneous specifications.

Given that the consistency of $\hat{\beta}_i$ hinges on the validity of the aforementioned rank condition, we replace $\hat{\beta}_i$ with the CCEMG estimate when computing the first step residuals. The estimate of β_i obtained from OLS estimation of (5) will be referred to as the "two-step CCE" estimate and the corresponding mean group version as CCEMG-2. For the pooled analog of (1), the first step residuals are obtained using the CCEP estimate and the resulting estimate is referred to as CCEP-2. Our empirical analysis reports both the one and two-step CCE estimates. A potential advantage of the two-step approach is that the second-step estimate is based on factors estimated by principal components instead of observable proxies and is therefore possibly less sensitive to the fulfillment of the rank condition.[13]

3.2 Omitted Ability and Heterogeneity Biases

In the interactive effects environment, there are at least two potential sources of bias that can arise in panel OLS/IV estimation of the returns to schooling. The first is the omitted ability bias that emanates from ignoring the common factor structure (2). While OLS estimation treats the ability components as part of the error term leading to endogeneity of the schooling variable, the IV estimator can be subject to bias if the instruments are inappropriate in that they are correlated with the factor structure. The second source of bias arises from estimating a pooled specification when the true regression coefficients are heterogeneous. In practice, the two biases may reinforce or offset each other depending on their signs. The interactive effects framework allows us to separately estimate the bias associated with each of the two sources.

Appendix A provides analytical expressions for the aforementioned sources of bias including conditions under which one would expect a given pattern in the relative magnitude of the regression parameter estimates. These expressions can then be employed to estimate the biases using the

[13]The rank condition is potentially very relevant in this application, given that our empirical analysis based on panel data includes a small number of observed variables (2-3, depending on the specification). Note that we only include cross-sectional averages of the earning and schooling variables in the CCE approach, given that age controls are equivalent to the inclusion of a deterministic time trend (see Appendix B for details).

interactive fixed effects estimates of the factor structure. Comparison of the component-specific OLS and IV biases allows us to isolate components that are responsible for exacerbating the IV bias relative to OLS from those where the instruments are effective at mitigating the bias. For instance, the instruments may reduce the bias associated with an ability component that is negatively correlated with schooling (e.g., high school skills) while worsening the bias associated with a component positively correlated with schooling (e.g., college skills).[14]

4 Data

4.1 Linked Survey-Administrative Data

Linked survey-administrative data come from the U.S. Census Bureau Gold Standard File (GSF) which links respondents from the SIPP with tax and benefit data from the IRS and SSA.[15] The linked dataset includes respondents' survey information from the SIPP for the years they were in the survey and annual tax and benefit information that ranges from 1978-2011 for some variables and 1951-2011 for others.[16] This allows us to construct a panel dataset with annual earnings, annual years of schooling, and other covariates. Annual earnings comes from the SSA's Detailed Earnings Record (DER), which is based on W-2 records for employed workers and Schedule C records for self-employed workers, including deferred earnings, and is available from 1978-2011. We construct a longitudinal years of schooling variable using the educational history information in the SIPP, which includes not just the highest level of schooling completed, but also the year each level was completed. More details on the construction of this variable can be found in Appendix C.

[14]We borrow the language of "high school skills" versus "college skills" from Heckman, Lochner, and Todd (2006, p. 390). One can also think of it as "mechanical skills" versus "cognitive/non-cognitive skills" (Prada and Urzua, 2017).

[15]We use version 6.0 of the GSF. Outside researchers can access a synthetic version of the GSF, known as SIPP Synthetic Beta. Researchers can then have their results validated on non-synthetic data. More information is available in Benedetto et al. (2018).

[16]Nine SIPP panels are linked: 1984, 1990, 1991, 1992, 1993, 1996, 2001, 2004, and 2008.

Administrative data on earnings is advantageous to survey data due to rising measurement error and non-response in survey data (Abowd and Stinson, 2013; Meyer et al., 2015). Previous work has shown that earnings data from surveys appears to be overstated at the bottom of the earnings distribution and understated at the top (Chenevert et al., 2016; Cristia and Schwabish, 2009; Pedace and Bates, 2000). Chenevert et al. (2016) also found that survey earnings data is overstated for lower education levels and understated for higher education levels. These findings have potential implications about the reliability of survey data for estimating the returns to schooling. Several of the aforementioned studies on the merits of administrative versus survey data analyze the same SIPP and SSA DER data that we use (Abowd and Stinson, 2013; Chenevert et al., 2016; Meyer et al., 2015).

Linked SIPP-administrative data therefore provides a unique panel dataset of education and earnings that is of high quality, has a long time dimension, and includes a large number of individuals. Most studies have relied on cross-section analysis (e.g., Angrist and Krueger, 1991; Card, 1995) or short panels (e.g., Carneiro et al., 2003; Carneiro and Heckman, 2002; Carneiro et al., 2011; Cunha et al., 2005). Studies that use longer panel data typically use either the PSID (e.g., Westerlund and Petrova, 2017) or the NLSY (e.g., Angrist and Newey, 1991; Ashworth et al., 2017; Koop and Tobias, 2004). The linked SIPP-administrative data has several advantages over the PSID and NLSY, including larger sample sizes, due to the combination of many SIPP panels;[17] more accurate earnings data, due to the removal of survey mis-reporting, non-response, and top-coding; less attrition, because longitudinal earnings data come from administrative records rather than repeated survey responses; and a longer time dimension for earnings, due to administrative earnings records that cover many years.

Heckman, Lochner, and Todd (2006) conclude in their survey of the literature that the solution to improving the estimation of returns to schooling lies in rich panel data and new econo-

[17]Most panel studies in the literature analyze approximately 1,000-2,000 individuals, with the extremes being 888 in Westerlund and Petrova (2017) and 3,695 in Cunha et al. (2005).

metric approaches. The use of linked survey-administrative data addresses the former of those recommendations. It also addresses the latter recommendation: the use of rich panel data allows for an interactive fixed effects framework which cannot be applied to cross-section or short panel data (see Altonji, 2010 for further discussion of these points).

4.2 Sample Selection and Summary Statistics

The main sample of analysis was selected based on eight selection criteria: (1) males; (2) age 16-65 during the entirety of 1978-2011; (3) at least 27 years of age at the time of their SIPP survey; (4) not currently enrolled in school at the time of their SIPP survey; (5) no missing information for variables included in the analysis; (6) positive earnings each year from 1978-2011; (7) meets annual minimum earnings thresholds; (8) at least one change in years of schooling during 1978-2011. The first five criteria are typical in the returns to schooling literature. The sample is restricted to males to analyze a population that historically is consistently and strongly attached to the labor market. The age range is limited to 16-65 to focus on prime working years. Criteria (3)-(5) select individuals who are most likely to have complete and accurate educational histories.

Criteria (6)-(8) are required due to the panel data approach. Positive earnings in each year is required to have a balanced panel sample. However, work while enrolled in school may often be part-time work. Thus, earnings may be artificially low during school, which could lead to biased estimates of the return to schooling from panel datasets. To ensure that we are analyzing earnings that reasonably represent individuals' full earning potential, we further limit the sample to individuals who meet minimum earnings thresholds during each and every year. We set the annual minimum earnings equal to the federal minimum wage multiplied by 800 hours, following the criterion adopted in Koop and Tobias (2004).[18] Finally, we further limit the sample to individuals

[18] Koop and Tobias (2004) have annual data on wages, hours, and weeks. Their sample restrictions to identify individuals whose earnings represent full earnings potential include limiting to individuals who work at least 30 weeks in the year, work at least 800 hours in the year, and have a wage of at least $1. We are forced to restrict based on annual earnings amounts due to the lack of annual weeks and hours worked information.

with at least one change in years of schooling so that we can estimate person fixed effects models.[19] Each of these criteria is typical in the panel data literature on estimating the returns to schooling: Angrist and Newey (1991), Koop and Tobias (2004), and Westerlund and Petrova (2017) analyze individuals who have positive earnings and increase their schooling at the same time, while Koop and Tobias (2004) also use minimum earnings and work thresholds.

Nonetheless, it is reasonable to be concerned that the final panel sample may not be representative of a typical individual's work and school experience. To investigate this, Table 1 shows summary statistics for not only the main sample of analysis, but also shows how the summary statistics change as we sequentially add the selection criteria to obtain our final sample. Panel A shows the balanced panel sample. Panel B shows a cross-section sample which is a subset of the panel sample for year 1990. We include both because, while our main results are based on a panel sample, we also provide some cross-section results below. Column (1) shows a baseline sample that applies criteria (1)-(5) listed above.[20] This sample has 47,500 individuals. Before applying the remaining selection criteria, we first show a comparative sample in column (2). This column restricts the baseline sample to individuals who have positive earnings every year after finishing their schooling. Whereas our restriction of positive earnings each year from 1978-2011 is required for the panel analysis, this restriction is more consistent with the cross-section literature.[21] This

[19]We checked the sensitivity of the results to an additional restriction that removed individuals with an earnings observation in the top one percent of the sample. We chose not to include this restriction in the main sample because (1) we do not have a precedent in the returns to schooling literature for what top percent to remove; (2) one advantage of using the administrative data rather than survey earnings data is the lack of top-coding; and (3) the results are similar. These results are available upon request.

[20]We show age in years in this table for demonstration, but for the analysis below we followed Angrist and Krueger (1991) and constructed age-in-quarters as the individual's age-in-quarters at the time of their SIPP survey. That is, the within-birthyear-birthquarter variation due to the differences in which quarter individuals were born and which quarter they were interviewed allows cross-section IV specifications that include birth year fixed effects, age controls, and the quarter-based IVs. When we moved to the panel setting, we calculated the age-in-quarters variable for their non-survey years by subtracting/adding four for each additional year away from the survey year for consistency and for the sake of estimating similar panel 2SLS specifications.

[21]The cross-section literature often relies on "potential experience" as a proxy for work experience when estimating returns to schooling, which explicitly builds in an assumption of positive earnings each year after finishing schooling: the proxy assumes that individuals are always either in school or working, but never both, and measures experience as $age - years\ of\ schooling - 6$. The cross-section literature typically also uses a similar sample selection criteria by analyzing individuals with positive earnings and not currently enrolled in school.

restriction decreases the sample by more than half, from 47,500 individuals to 22,000 individuals. Based on Panel A, the sample has higher average earnings (+$13,800), more average years of schooling (+0.73 years), lower average age (-0.63 years), and has higher means for the fraction White, non-Hispanic, native born, and married.[22]

Column (3) imposes the criterion that individuals must have positive earnings each year from 1978-2011. This only further decreases the sample from 22,000 individuals to 18,500. The average earnings (+$320) and average years of schooling (-0.09 years) do not change much compared the effect of imposing positive earnings after finishing schooling. The average age increases (+0.68 years) and the fraction White, non-Hispanic, native born, and married continues to increase. Column (4) imposes the annual minimum earnings criterion. This decreases the sample from 18,500 individuals to 12,000. The only particularly notable changes arising from this restriction are higher annual earnings (+$4,290), as expected, and increased mean age (+0.94 years). Finally, column (5) imposes the criterion of at least one change in schooling during 1978-2011. This decreases the sample from 12,000 individuals to 3,600. The notable changes in sample means arising from this restriction are mean years of school (+0.62 years) and mean age (-1.83 years).[23] In summary, comparing the balanced panel and minimum earnings restrictions in columns (3) and (4) to the comparative sample in column (2), these criteria do not further decrease the sample size dramatically compared to the decrease from the baseline sample when imposing the common assumption in the literature of positive earnings after finishing school. Comparing the final sample in column (5) to the comparative sample, the only differences in the summary statistics are relatively small changes in average annual earnings due to the minimum earnings threshold (+$4,580), years of school (+0.53 years), age (-0.21 years), and small increases in the fraction White, non-Hispanic, native born, and married.

[22]The large standard deviation for annual earnings is due to the fact that the earnings data is very skewed to the right.

[23]The relatively large changes in sample size and mean age are due to losing individuals who never complete high school and also older individuals who already completed their schooling by 1978, both of whom would have no observed changes in years of schooling.

To further alleviate concerns that our main sample may not be representative, we begin our analysis below by estimating cross-section specifications based on both our main sample in column (5) of Panel B and the comparative sample in column (2) in order to replicate the well-known pattern of OLS/2SLS results from the literature. Furthermore, the main panel sample is used to generate OLS and 2SLS estimates, in addition to estimates from specifications based on an interactive fixed effects structure. Thus, to the extent that estimates of the return to schooling are larger or smaller based on panel analysis, all of the estimates should be affected by this, such that comparing estimates from OLS/2SLS with estimates from interactive fixed effects specifications still illustrates the effect of allowing for multiple unobserved skills whose prices can vary over time.

5 Empirical Results

The empirical results are organized into five subsections. Section 5.1 presents the set of specifications estimated that differ according to whether cross-section or panel data are employed, whether the regression parameters are allowed to be heterogeneous, and whether interactive fixed effects are incorporated. Section 5.2 reports the cross-section estimates which replicate the robust empirical finding in the literature that the IV estimate of the returns to schooling exceeds the OLS estimate. The former is based on using the quarter of birth interacted with year of birth as instruments following the seminal paper by Angrist and Krueger (1991). Our choice of IV is driven by the fact that it is one of the most widely used in the literature and the only available one in our dataset, although we acknowledge its limitations with respect to both relevance and exogeneity (Bound et al., 1995; Buckles and Hungerman, 2013; Barua and Lang, 2016). Section 5.3 presents the panel OLS, 2SLS, and interactive fixed effects estimates obtained by pooling the data across cross-section units assuming homogeneous parameters. Section 5.4 contains results for models that allow heterogeneity in the returns to schooling. Section 5.5 details the bias estimates from both pooled and heterogeneous models. Finally, Section 5.6 conducts a more in-depth analysis of the nature and degree of heterogeneity by examining the distribution of returns for various subgroups of the population.

5.1 Estimated Specifications

We estimate a total of seven specifications that are summarized in Table 2. We group the specifications as follows:

- **Group 1 [Specification 1]:** Cross-section OLS and 2SLS regressions of log annual earnings on schooling to verify the "IV $>$ OLS" result commonly found in empirical studies. We estimate the specification

  $$y_i = c + s_i\beta + a_i\rho_1 + a_i^2\rho_2 + u_i$$

  where a_i denotes the age of individual i. The age variables are included to account for the actual experience (we discuss this issue further below).

- **Group 2 [Specifications 2-3]:** Standard panel data specifications that include person or time fixed effects to control for unobserved heterogeneity. When person fixed effects are included, it takes the form

  $$y_{it} = c_i + s_{it}\beta + a_{it}\rho_1 + a_{it}^2\rho_2 + u_{it} \tag{6}$$

  where a_{it} denotes the age of individual i at period t. We also estimate specifications with time instead of person fixed effects. Angrist and Newey (1991) consider a specification of the form (ignoring demographic controls)

  $$y_{it} = c_i + \delta_t + s_{it}\beta + pe_{it}\rho_1 + pe_{it}^2\rho_2 + u_{it}$$

  where pe_{it} denotes potential experience and is computed as $pe_{it} = a_{it} - s_{it} - 6$, where they define s_{it} as the highest grade completed. They estimate a reduced form schooling effect (expressed as a function of s_{it} and a_{it}) based on the observation that the effect of schooling conditional on potential experience is not identified. We present a derivation in the Appendix B which shows that the effect of actual experience can be accounted for by including age and its square as controls as in (6).

- **Group 3 [Specifications 4-5]:** This group contains specifications that include interactive fixed effects while assuming that the regression coefficients are homogeneous. The nesting model takes the form

$$y_{it} = c_i + s_{it}\beta + a_{it}\rho_1 + a_{it}^2\rho_2 + \lambda_i' f_t + u_{it}$$

  We also estimate specifications with time instead of person fixed effects.

- **Group 4 [Specifications 6-7]:** This group consists of specifications where the slope parameters are allowed to be individual specific. The general specification is given by

$$y_{it} = c_i + s_{it}\beta_i + a_{it}\rho_{1i} + a_{it}^2\rho_{2i} + \lambda_i' f_t + u_{it} \tag{7}$$

  We also estimate specifications when the interactive fixed effects are excluded (OLSMG). We consider the specification shown in (7) with person and interactive fixed effects, age controls, and individual specific parameters to be our preferred specification because it is the most flexible version and also accounts for experience as discussed in Appendix B.

Our empirical findings are robust to the inclusion of quadratic schooling and/or quartic age variables (Murphy and Welch, 1990; Cho and Phillips, 2018) for all specifications described above. See Appendix D1 for details. We have also estimated specifications 1 and 3 with demographic controls including race, Hispanic status, foreign born status, marital status, state of residence during the SIPP survey and birth year. The results are very similar and available upon request.

5.2 Cross-Section Estimates

Columns (1)-(4) of Table 3 present the cross-section OLS and 2SLS estimation results. Columns (1)-(2) report findings based on the final sample as summarized in Table 1, Panel B, column (5). OLS yields an estimated effect of schooling of about 9.2 percent while the corresponding 2SLS point estimate, using the interactions of quarter of birth with year of birth as instruments, is about 13.4 percent. A similar pattern of results is observed in columns (3)-(4) for the comparison sample without the earnings-in-school restriction as described in Table 1, Panel B, column (2), with the 2SLS point estimate exceeding the OLS point estimate by about 45 percent.

Overall, these findings are in accordance with the literature summarized in Card (2001) which indicates OLS estimates generally range from 5 to 10 percent, while 2SLS estimates generally range from 10 to 16 percent; and demonstrates the robustness of the "IV>OLS" result across different datasets as well as different instrument sets. For instance, the seminal study by Angrist and Krueger (1991) finds, based on the 1920-29 birth cohort using data on men from the 1970 Census, an OLS estimate of about 7 percent and a 2SLS estimate of about 10 percent when controlling for age and its square, race, marital status and urban residence.

5.3 Pooled Estimates

The results from OLS and 2SLS estimation using panel data over 1978-2011 are presented in Table 3, columns (5)-(7). Similar in spirit to the cross-section analysis, the OLS point estimates are apparently smaller than the 2SLS point estimates across specifications. In addition to parameter estimates, column (5)-(7) also report the results of Pesaran's (2015) CD test for the presence of cross-section dependence applied to the residuals for each estimated specification.[24] In all cases,

[24]The test is based on estimated pairwise correlation coefficients between the pooled OLS/2SLS residuals for each pair of cross-section units. The test has a standard normal asymptotic distribution under the null hypothesis of no cross-section dependence.

the test provides evidence against no cross-section dependence (at the 1% level) which further motivates the use of the interactive fixed effects estimators.

Table 4 reports the results from estimating pooled specifications with interactive fixed effects. The estimators included are the IFE, CCEP and CCEP-2 estimators. Irrespective of whether one controls for interactive effects using principal components or cross-section averages of the observed variables, the point estimates are smaller in magnitude than the OLS and 2SLS estimates reported in Table 3. For instance, the IFE point estimate with time fixed effects is about 2.6 percent while the corresponding OLS and 2SLS estimates are about 10.5 percent and 12.7 percent, respectively. Under the assumption that the interactive effects specification represents the true model, the pattern of results suggests that the OLS and 2SLS estimates are both upward biased, with the magnitude of the 2SLS bias exceeding the OLS bias. This is consistent with the premise that the IV approach suffers from flawed instruments that are correlated with unobserved abilities or skills, which the interactive fixed effects specifications can account for. The CCEP point estimates are slightly larger than the IFE estimates reflecting the difference in how the unobserved common factors are accounted for in the two approaches. However, the CCEP-2 estimates that employ the estimated factors are closer to the IFE estimates.

5.4 Mean Group Estimates

Table 5 presents results from estimating the specifications 6-7 in Table 2 that allow the slope parameters to be individual-specific. In addition to the CCEMG, CCEMG-2 and IFEMG estimators, we also include the OLSMG estimator that entails taking the average of the individual specific time series OLS regressions of log earnings on a constant, age controls, and schooling. Note that a mean group 2SLS estimate cannot be computed since the instruments are time-invariant. To confirm the presence of heterogeneity, Table 5 also reports the results from conducting two slope homogeneity tests recently proposed by Ando and Bai (2015) and Su and Chen (2013).[25] Both tests provide evidence against the null of slope homogeneity at the 1% significance level.

Consistent with the foregoing pooled results, when the interactive effects are ignored, the OLSMG point estimate of the return to schooling is larger (7.8 percent). The one-step and two-step CCE approaches yield similar estimates of about 4.4 and 4.1 percent, respectively, which is larger relative to the IFE point estimate of the average marginal returns to schooling (2.8 percent). Given that the mean group estimates exceed the corresponding pooled estimates for all three approaches, we should expect a negative correlation between the individual level estimate $\hat{\beta}_i$ and the weight on individual $i's$ return ω_i according to the heterogeneity bias analysis in Appendix A2. Indeed, the IFE-based correlation was estimated to be -0.003, while the corresponding one and two-step CCE correlation was estimated as -0.005 and -0.017, respectively. The pattern of findings for the estimated schooling effect obtained from the heterogeneous factor models therefore suggest that ignoring potential heterogeneity is likely to induce an downward bias in the parameter estimates. We also computed the CD test for cross-section dependence based on the OLSMG estimate and found evidence against no cross-section dependence for both specifications at the 1% level.[26]

5.5 Bias Estimates

As discussed in Appendix A1, the interactive fixed effects estimates can be used to obtain estimates of the OLS and 2SLS biases associated with each of the skill components.[27] The top three panels of Table 6 show the biases corresponding to the first four common factors for each of the IFE, CCEP and CCEP-2 estimates. The contribution of additional factors to the total bias (reported in column 5) is marginal in all cases. For all three estimation approaches, the aggregate 2SLS bias

[25]The Ando and Bai (2015) test is based on the (scaled) difference between the individual level estimates and the IFEMG estimate while the Su and Chen (2013) test is based on the Lagrange Multiplier (LM) principle that utilizes IFE residuals computed under the null of slope homogeneity. Both tests have a standard normal asymptotic null distribution. We refer the reader to the original articles for details.

[26]The results are available upon request.

[27]We investigate an alternative interpretation of the factor structure in Appendix D2.

exceeds the OLS bias across specifications, consistent with the findings reported in Tables 3 (panel results) and 4. For instance, the aggregate OLS bias with year fixed effects using the IFE approach accounts for about 75 percent of the estimated OLS effect of schooling (Table 3, column 6). The corresponding aggregate 2SLS bias accounts for about 80 percent of the estimated 2SLS effect of schooling (Table 3, column 7). Similar magnitudes are obtained using the CCE approaches.

The disaggregate bias estimates reveal some interesting patterns. First, the leading common component is the major contributor to the aggregate OLS bias, accounting for at least 56 percent of the bias across specifications/estimators and nearly all of the bias when year fixed effects are used to control for unobserved heterogeneity. In contrast, the first two common components are important contributors to the 2SLS bias, with the first component being relatively more important. Notably, there is some evidence suggesting negative bias components emerged in both OLS and 2SLS cases. The negative bias components can be interpreted as the presence of mechanical skills that are negatively correlated with schooling but make a positive contribution to earnings (Heckman, Lochner, and Todd, 2006; Prada and Urzua, 2017). Finally, it is useful to note that the 2SLS estimator is mostly successful at ameliorating the bias associated with the common components which only make a negligible contribution to the total bias (i.e., components other than the first two), at the expense of aggravating the bias in the two leading components (the exception being factor skill 1 and 4 using the IFE estimates with year fixed effects). These results show that, assuming an underlying interactive factor structure, the consistent "IV $>$OLS" finding in the literature could be due to the use of instruments that actually worsen the ability bias.

As in the pooled case, we compute the biases associated with the OLSMG estimate using the CCEMG, CCEMG-2 and IFEMG estimates of the common structure. The findings are reported in the bottom three panels of Table 6. The first common component is responsible for roughly 50 percent of the aggregate bias using the CCE mean group estimates and nearly all of the bias using the IFEMG estimate. In any case, the aggregate bias is very large: the aggregate OLSMG bias accounts for up to 64 percent of the estimated OLSMG effect of schooling in Table 5, depending on the particular estimator used to control for interactive effects.

Finally, since the interactive effects framework allows for both individual slope heterogeneity and cross-sectional dependence modeled through a common factor structure, it is possible to obtain estimates of the biases emanating from each of the two sources. We can use the decomposition $\hat{\beta}_{POLS} - \hat{\beta}_{IFEMG} = \left(\hat{\beta}_{POLS} - \hat{\beta}_{IFE}\right) + \left(\hat{\beta}_{IFE} - \hat{\beta}_{IFEMG}\right)$, where $\hat{\beta}_{POLS}$ denotes the OLS estimate assuming a homogeneous slope parameter. The first term in the decomposition may be interpreted as the bias arising from ignoring the common factor structure while the second term denotes the bias from ignoring potential parameter heterogeneity. The results are shown in Figure 1. Based on the IFE results, we find $\hat{\beta}_{POLS} - \hat{\beta}_{IFEMG} \simeq 4.9$ percentage points, $\hat{\beta}_{POLS} - \hat{\beta}_{IFE} \simeq 5.7$ percentage points, $\hat{\beta}_{IFE} - \hat{\beta}_{IFEMG} \simeq -0.8$ percentage points. A similar calculation using the one-step CCE estimates yields $\hat{\beta}_{POLS} - \hat{\beta}_{CCEMG} \simeq 3.3$ percentage points, $\hat{\beta}_{POLS} - \hat{\beta}_{CCEP} \simeq 3.9$ percentage points, $\hat{\beta}_{CCEP} - \hat{\beta}_{CCEMG} \simeq -0.6$ percentage points. For the two-step CCE method, we obtain $\hat{\beta}_{POLS} - \hat{\beta}_{CCEMG-2} \simeq 3.6$ percentage points, $\hat{\beta}_{POLS} - \hat{\beta}_{CCEP-2} \simeq 5.4$ percentage points, $\hat{\beta}_{CCEP-2} - \hat{\beta}_{CCEMG-2} \simeq -1.8$ percentage points. For all three estimation approaches, the omitted ability bias captured using the interactive fixed effects structure appears to be the more important contributor to the total bias of the least squares estimator.

5.6 Heterogeneity Analysis

This section examines the extent of heterogeneity in individual-level returns. We focus on the differences between the OLS and factor model (FM, henceforth) estimates pertaining to distributional characteristics of the individual returns, and differences in mean returns across and within subgroups.[28] Most studies in the literature assume that the return to schooling is the same for all individuals, but there are exceptions (Harmon et al., 2003; Henderson et al., 2011; Koop and Tobias, 2004; Li and Tobias, 2011). The results for heterogeneity across and within subgroups discussed below are most comparable to the results from Henderson et al. (2011). They use cross-

[28]We also investigate the characteristics associated with extreme returns, and the results are available upon request.

section nonparametric kernel regression methods to study heterogeneity in returns and summarize the heterogeneity across and within subgroups, but their method does not address omitted ability bias.

5.6.1 Distribution of Individual Returns

Figure 2 plots the distribution of individual returns for each estimator based on kernel density plots.[29] There are clearly large differences in returns across individuals. Most of the density associated with the heterogeneous OLS model falls between approximately a negative 50 percent return and a positive 50 percent return. The FMs clearly shift the distribution to the left, which is consistent with evidence that the FMs are removing positive ability bias from the OLS estimates. The FM estimates place greater density immediately around the modal return, which is illustrated by the height of the density plots compared to OLS.

The most striking result from the figure is that each of the estimators shows a considerable fraction of individuals with negative returns to schooling. Overall, 39.2 percent of individuals have negative returns in the heterogeneous OLS model, 46.2 percent in the heterogeneous IFE model, 42.8 percent in the heterogeneous CCE model, and 44.2 percent in the heterogeneous CCE-2 model as shown in Table 5. These percentages are larger than the estimate in Henderson et al. (2011), who find that 15.2 percent of individuals who are White have negative returns to schooling. Heckman et al. (2017) and Prada and Urzua (2017), who use ability proxies from the NLSY to address ability bias, appear to find fractions of negative returns between the estimate in Henderson et al. (2011) and our estimates.

5.6.2 Across-Group Heterogeneity

Table 7 reports the mean and variance of the individual returns separately by several subgroups: race (White, Black, other race), Hispanic status, foreign born status, birth cohort (born before 1950,

[29]The kernel density plots are based on a standard normal (Gaussian) kernel with a bandwidth of 0.5.

born 1950-1954, born 1955-1959, born after 1959), and highest education level completed.[30,31] Mean returns for individuals who are non-White are statistically tested against the mean for individuals who are White. For the other subgroups, the mean for each group is statistically tested against the mean for the group listed directly above it within each panel of the table. Based on the OLSMG model, the mean return to schooling is statistically larger for (1) individuals who are White compared to Black; (2) for individuals born in later birth cohorts; (3) individuals with a high school degree compared to some college; (4) individuals with a bachelor's degree compared to some college; (5) for individuals with a bachelor's degree compared to a graduate degree.

Mean individual returns from the FMs are generally smaller than those from the OLSMG model for every subgroup, which is consistent with the main results discussed in the previous sections. Further, the FM estimates show there are no statistically significant differences by race. Similar to the OLSMG results, the FM estimates indicate the largest mean return to schooling for individuals who ultimately stop at high school, and the next largest mean returns for individuals whose highest achievement is a bachelor's, and then a graduate degree.

The statistically larger returns for more recent birth cohorts, found across all four heterogeneous models, is consistent with evidence that returns to schooling have risen over time (Card and Lemieux, 2001). In contrast to Henderson et al. (2011), which suggests diminishing marginal returns to years of schooling at least until graduate school, both OLSMG and FM results are more suggestive of "sheepskin effects", i.e., 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 (Hungerford and Solon, 1987; Jaeger and Page, 1996; Layard and Psacharopoulos,

[30]The CCE model from Pesaran (2006) makes a random coefficients assumption on the individual-level returns. This assumption only affects the CCE standard errors and therefore analysis of the mean and variance of individual-level returns by particular characteristics is feasible without violating assumptions of the model.

[31]Due to the limited sample size, the results of non-white, Hispanic, and foreign born individuals should be interpreted with caution. It is possible that these groups in our sample have unique attributes and are not representative of the rest of the population or that we lack the statistical power to detect significant differences.

1974).

5.6.3 Within-Group Heterogeneity

The heterogeneous models also allow for the analysis of heterogeneity within subgroups. Table 7 shows the variance of the individual returns within each subgroup. The FM estimates show mostly larger variance than OLSMG for every subgroup. Both FM and OLSMG estimates show interesting patterns of the relative variance across subgroups that are worth noting: (1) The results generally show larger variance for more recent birth cohorts; (2) Evidence also suggests larger variance for individuals who only obtain a high school degree than any other education level.

Table 8 reports the 25th, 50th, and 75th percentiles of the distribution of individual returns by subgroup. The FM estimators generally show smaller returns at each percentile, again consistent with previous results. The difference between the 25th and 75th percentiles is often larger for OLSMG than the FMs. This appears inconsistent with the larger variance associated with the FMs in Table 7 . However, this can be reconciled by analyzing the distributional plot in Figure 2. The FM estimators place relatively more density immediately around the mode than OLS, which produces a smaller range between the 25th and 75th percentiles than OLS. But the FM estimators also have longer right tails than OLS, which increases the overall variance.

6 Conclusion

This study explores the viability of an interactive fixed effects approach to estimating the returns to schooling employing a large panel dataset that links survey data with tax and benefit information obtained from administrative records. SIPP provides longitudinal education information, while administrative records from the IRS and SSA provide a long history of high-quality earnings data. The generality of the interactive fixed effects approach over most existing approaches is apparent in at least three dimensions: (1) 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; (2) The endogeneity of schooling is accounted for through estimation of or proxying for the skill prices that is made possible by the high-dimensional nature of the the panel without the need to resort to external instruments or proxies for ability; (3) Individual-level heterogeneity in the returns to schooling can be accommodated that allows us to simultaneously address the twin sources of bias that can arise due to unmeasured skills (the omitted variable bias) and assuming that the marginal returns to schooling are homogeneous across individuals.

The estimates from our preferred specification indicate considerably lower average marginal returns to schooling compared to traditional methods such as OLS or 2SLS. While both aforementioned sources of bias contribute to the aggregate least squares bias, our estimates point to a relatively more important role for the bias induced by omission of time-varying returns to skills. The two biases operate in the opposite direction serving to explain the gap in the heterogeneous interactive fixed effects estimates and the homogeneous panel OLS estimates. Our subgroup heterogeneity analysis suggests larger returns for individuals born in more recent years, the presence of "sheepskin effects" so that degree attainment can have an important impact in determining the value of additional years of schooling, and considerable within-group heterogeneity.

Several extensions of our analysis are in order. First, it would be interesting to investigate the extent of heterogeneity in returns at different quantiles of the earnings distribution using the quantile interactive effects approach recently developed by Harding and Lamarche (2014). Second, while our results indicate important differences both across and within subgroups, our sample only includes men. Analysis of heterogeneity from a gender standpoint is a promising avenue for future research. Third, our paper only considers cross-sectional heterogeneity but as the nonparametric analysis of Henderson et al. (2011) documents, returns vary not only across individuals but also across time. A limitation of our analysis in this context is that splitting the sample by time periods would leave us with relatively few observations in each subsample (splitting by, say, half would imply a time series dimension of seventeen for each subsample) to estimate the individual specific parameters. Fourth, our analysis assumes that the skill prices are homogeneous across individuals although they are allowed to vary over time. Heckman and Scheinkman (1987) find evidence in favor of a model where skill prices are sector-specific which suggests the presence of a grouped factor structure for earnings which allows heterogeneity in skill prices across sectors of the economy but possibly homogeneous for individuals within a particular sector. We leave analyses of these and related issues as possible directions for further research.

References

Abowd, J. M. and M. H. Stinson (2013). Estimating measurement error in annual job earnings: A comparison of survey and administrative data. *The Review of Economics and Statistics 95*(5), 1451–1467.

Altonji, J. G. (2010). Multiple skills, multiple types of education, and the labor market: A research agenda. Note AEA: Ten Years and Beyond: Economists Answer NSF's Call for Long-Term Research Agendas, September.

Ando, T. and J. Bai (2015). A simple new test for slope homogeneity in panel data models with interactive effects. *Economics Letters 136*, 112–117.

Angrist, J. D. and A. B. Krueger (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics 106*(4), 979–1014.

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.

Ashenfelter, O. and A. Krueger (1994). Estimates of the economic return to schooling from a new sample of twins. *The American Economic Review 84*(5), 1157–1173.

Ashworth, J., V. J. Hotz, A. Maurel, and T. Ransom (2017). Changes across cohorts in wage returns to schooling and early work experiences. NBER Working Paper No. 24160.

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

Barua, R. and K. Lang (2016). School entry, educational attainment, and quarter of birth: A cautionary tale of a local average treatment effect. *Journal of Human Capital 10*, 347–376.

Becker, G. (1962). Investment in human capital: a theoretical analysis. *Journal of Political Economy 70*(5), 9–49.

Benedetto, G., J. C. Stanley, and E. Totty (2018). The creation and use of the sipp synthetic beta v7.0. CES Technical Notes Series 18-03, Center for Economic Studies, U.S. Census Bureau.

Bhuller, M., M. Mogstad, and K. G. Salvanes (2017). Life-cycle earnings, education premiums, and internal rates of return. *Journal of Labor Economics 35*, 993–1030.

Bound, J., D. A. Jaeger, and R. M. Baker (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American statistical association 90*, 443–450.

Buckles, K. S. and D. M. Hungerman (2013). Season of birth and later outcomes: Old questions, new answers. *Review of Economics and Statistics 95*, 711–724.

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. In S. Polschek (Ed.), *Research in Labor Economics*, pp. 23–48. JAI Press.

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

Card, D. and T. Lemieux (2001). Can falling supply explain the rising return to college for younger men? a cohort-based analysis. *The Quarterly Journal of Economics 116*, 705–746.

Carneiro, P., K. Hansen, and J. Heckman (2003). 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.

Carneiro, P. and J. Heckman (2002). The evidence on credit constraints in post-secondary schooling. *Economic Journal 112*(482), 705–734.

Carneiro, P., J. Heckman, and E. Vytlacil (2011). Estimating marginal returns to education. *The American Economic Review 101*(16), 2754–2781.

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*.

Chenevert, R., M. Klee, and K. Wilkin (2016). Do imputed earnings earn their keep? evaluating sipp earnings and nonresponse with administrative records. U.S. Census Bureau Working Paper Number: SEHSD-WP2016-18, SIPP-WP-275.

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

Cristia, J. and J. Schwabish (2009). Measurement error in the sipp: Evidence from administrative matched records. *Journal of Economic and Social Measurement 34*, 1–17.

Cunha, F., J. Heckman, and S. Navarro (2005). Separating uncertainty from heterogeneity in life cycle earnings. *Oxford Economic Papers 57*(2), 191–261.

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

Harding, M. and C. Lamarche (2014). Estimating and testing a quantile regression model with interactive effects. *Journal of Econometrics 178*, 101–113.

Harmon, C., V. Hogan, and I. Walker (2003). Dispersion in the economic return to schooling. *Labour Economics 10*, 205–214.

Hause, J. C. (1980). The fine structure of earnings and the on-the-job training hypothesis. *Econometrica 48*(4), 1013–1029.

Heckman, J. and J. Scheinkman (1987). The importance of bundling in a gorman-lancaster model of earnings. *The Review of Economic Studies 54*(2), 243–255.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2017). Returns to education: The causal effects of education on earnings, health and smoking. *Journal of Political Economy forthcoming*.

Heckman, J. J., L. L. Lochner, and P. E. Todd (2006). Earnings functions, rates of return and treatment effects: The mincer equation and beyond. In E. Hanushek and F. Welsch (Eds.), *Hankbook of Economics of Education*, Volume 1, pp. 307–458. New York: Elsview.

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*, 411–482.

Heckman, J. J., S. Urzua, and E. Vytlacil (2006). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics 88*, 389–432.

Henderson, D. J., S. W. Polacheck, and L. Wang (2011). Heterogeneity in schooling rates of return. *Economics of Education Review 30*(6), 1202–1214.

Humphries, J. E. and F. Kosse (2017). On the interpretation of non-cognitive skills–what is being measured and why it matters. *Journal of Economic Behavior & Organization 136*, 174–185.

Hungerford, T. and G. Solon (1987). Sheepskin effects in the returns to education. *The Review of Economics and Statistics 69*, 175–177.

Imbens, G. and J. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Jaeger, D. A. and M. E. Page (1996). Degrees matter: New evidence on sheepskin effects in the returns to education. *The Review of Economics and Statistics 78*, 733–740.

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

Layard, R. and G. Psacharopoulos (1974). The screening hypothesis and the returns to education. *Journal of Political Economy 82*, 985–998.

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

Li, M. and J. L. Tobias (2011). Bayesian inference in a correlated random coefficients model: Modeling causal effect heterogeneity with an application to heterogeneous returns to schooling. *Journal of Econometrics 162*, 345–361.

Meyer, B., W. Mok, and J. Sullivan (2015). Household surveys in crisis. *Journal of Economic Perspectives 29*, 199–226.

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

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

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

Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *The American Economic Review 96*(1), 152–175.

Pedace, R. and N. Bates (2000). Using administrative records to assess earnings reporting error in the survey of income and program participation. *Journal of Economic and Social Measurement 26*, 173–192.

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

Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. *Econometric Reviews 34*, 1089–1117.

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*, 1410–1455.

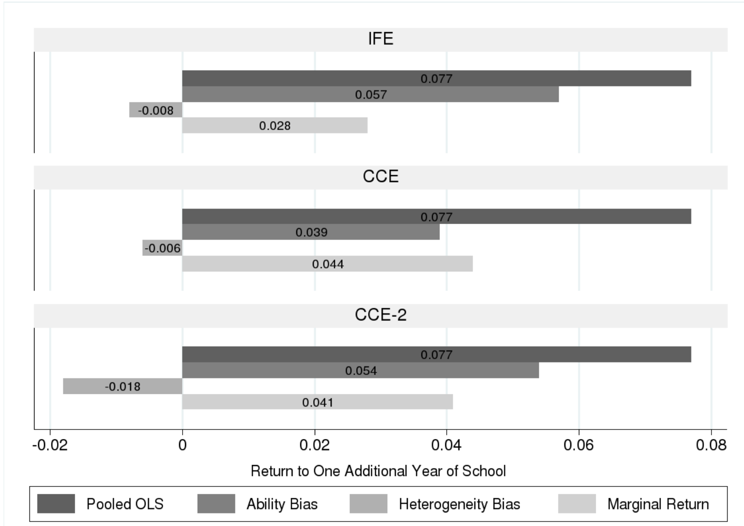
Prada, M. F. and S. Urzua (2017). One size does not fit all: Multiple dimensions of ability, college attendance, and earnings. *Journal of Labor Economics 35*(4), 953–991.

Song, M. (2013). Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications. *Unpublished Manuscript*.

Su, L. and Q. Chen (2013). Testing homogeneity in panel data models with interactive fixed effects. *Econometric Theory 29*(6), 1079–1135.

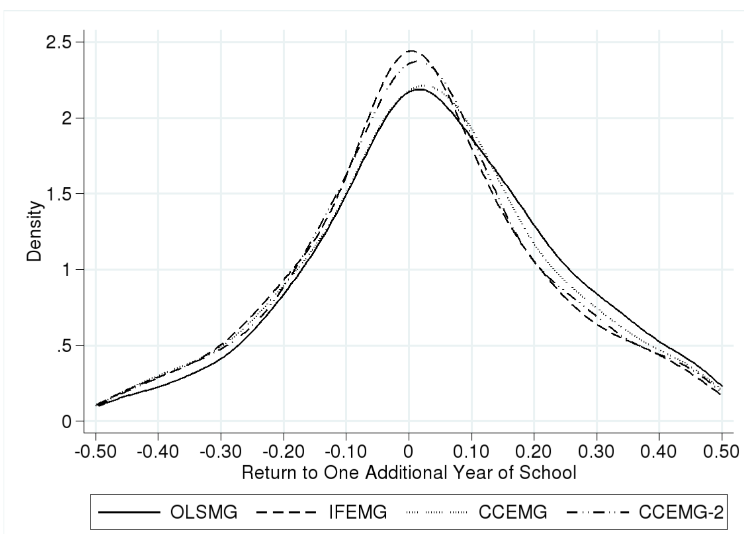
Westerlund, J. and Y. Petrova (2017). Asymptotic collinearity in cce estimation of interactive effects models. *Economic Modeling forthcoming*.

Figure 1: Bias Decomposition of Pooled OLS Estimate



Source: SIPP respondents linked to IRS and SSA data in the U.S. Census Bureau Gold Standard File.
Note: The darkest shaded bar for each estimator represents the pooled OLS estimate with person fixed effects and age controls, corresponding to column (5) in Table 3. We use the decomposition $\hat{\beta}_{POLS} - \hat{\beta}_{IFEMG} = \left(\hat{\beta}_{POLS} - \hat{\beta}_{IFE}\right) + \left(\hat{\beta}_{IFE} - \hat{\beta}_{IFEMG}\right)$ to estimate the ability and heterogeneity bias, where $\hat{\beta}_{POLS}$ denotes the OLS estimate assuming a homogeneous slope parameter. The same calculations are applied using CCE and CCE-2 estimates.

Figure 2: Distribution of Marginal Returns to Schooling



Note: Each line is a kernel density plot of individual returns based on the heterogeneous model for the given estimator, based on a Gaussian kernel with a bandwidth of 0.5. Results are based on the specification in Table 5.

Table 1: Summary Statistics

	(1) Baseline Sample	(2) Plus Positive Earnings After Schooling	(3) Plus Positive DER Earnings 1978-2011	(4) Plus Annual Minimum Earnings	(5) Plus ≥ One Schooling Change During 1978-2011
A. Balanced Panel Sample					
Annual Earnings	36,620	50,420	50,740	55,030	55,000
	(109,200)	(95,450)	(97,170)	(110,200)	(80,490)
Years of School	13.55	14.28	14.19	14.19	14.81
	(2.31)	(2.11)	(2.06)	(2.05)	(2.05)
Age	40.16	39.53	40.21	41.15	39.32
	(10.94)	(10.86)	(10.79)	(10.61)	(10.6)
Black	0.085	0.060	0.059	0.052	0.049
	(0.280)	(0.239)	(0.236)	(0.222)	(0.216)
Other Race	0.045	0.032	0.023	0.022	0.024
	(0.208)	(0.175)	(0.151)	(0.147)	(0.160)
Hispanic	0.073	0.038	0.035	0.032	0.037
	(0.260)	(0.190)	(0.184)	(0.178)	(0.191)
Foreign Born	0.089	0.040	0.026	0.027	0.033
	(0.285)	(0.195)	(0.158)	(0.164)	(0.177)
Married	0.684	0.727	0.748	0.793	0.760
	(0.465)	(0.445)	(0.434)	(0.406)	(0.428)
Birth Year	1954	1955	1954	1953	1955
	(4.83)	(4.66)	(4.49)	(4.04)	(4.011)
Observations	1,609,000	744,000	624,000	401,000	123,000
B. 1990 Cross Section Sample					
Annual Earnings	34,800	44,340	45,520	49,400	47,230
	(36,610)	(38,750)	(39,420)	(43,860)	(38,610)
Years of School	13.61	14.36	14.25	14.21	14.87
	(2.32)	(2.09)	(2.05)	(2.05)	(2.03)
Age	35.66	35.03	35.71	36.65	34.82
	(4.83)	(4.66)	(4.49)	(4.04)	(4.01)
Black	0.084	0.059	0.059	0.050	0.056
	(0.280)	(0.239)	(0.236)	(0.222)	(0.216)
Other Race	0.044	0.032	0.024	0.021	0.028
	(0.208)	(0.175)	(0.151)	(0.147)	(0.160)
Hispanic	0.074	0.036	0.035	0.033	0.042
	(0.260)	(0.190)	(0.184)	(0.178)	(0.191)
Foreign Born	0.088	0.039	0.024	0.029	0.028
	(0.285)	(0.195)	(0.158)	(0.164)	(0.178)
Married	0.716	0.773	0.784	0.792	0.833
	(0.452)	(0.423)	(0.410)	(0.380)	(0.392)
Birth Year	1954	1955	1954	1953	1955
	(4.83)	(4.66)	(4.49)	(4.04)	(4.01)
Observations	47,500	22,000	18,500	12,000	3,600

Source: SIPP respondents linked to IRS and SSA data in the U.S. Census Bureau Gold Standard File.
Note: Each column reports averages and standard deviations (in parentheses) for the sample specified in the column and table panel. Column (1) starts with a baseline sample. Columns (2)-(5) sequentially add additional sample criteria until the final sample is shown in column (5). Table Panel A shows the balanced panel samples. Panel B shows cross-section samples based on the 1990 cross-section values of the balanced panel samples. The number of observations and all other statistics are rounded according to U.S. Census Bureau disclosure avoidance rules. Annual earnings are adjusted for inflation to 1999 dollars. See section 4.2 for additional details.

Table 2: Summary of Estimated Specifications

	Specification	Controls	Estimator
1.	$y_i = c + s_i\beta + a_i\rho_1 + a_i^2\rho_2 + u_i$	age controls	CSOLS, CS2SLS
2.	$y_{it} = c_i + s_{it}\beta + a_{it}\rho_1 + a_{it}^2\rho_2 + u_{it}$	person fixed effects, age controls	POLS
3.	$y_{it} = \delta_t + s_{it}\beta + a_{it}\rho_1 + a_{it}^2\rho_2 + u_{it}$	time fixed effects, age controls	POLS, P2SLS
4.	$y_{it} = c_i + s_{it}\beta + a_{it}\rho_1 + a_{it}^2\rho_2 + \lambda_i' f_t + u_{it}$	person and interactive fixed effects, age controls	IFE, CCEP, CCEP-2
5.	$y_{it} = \delta_t + s_{it}\beta + a_{it}\rho_1 + a_{it}^2\rho_2 + \lambda_i' f_t + u_{it}$	time and interactive fixed effects, age controls	IFE, CCEP, CCEP-2
6.	$y_{it} = c_i + s_{it}\beta_i + a_{it}\rho_{1i} + a_{it}^2\rho_{2i} + u_{it}$	person fixed effects, age controls	OLSMG
7.	$y_{it} = c_i + s_{it}\beta_i + a_{it}\rho_{1i} + a_{it}^2\rho_{2i} + \lambda_i' f_t + u_{it}$	person and interactive fixed effects, age controls	IFEMG, CCEMG, CCEMG-2

Note: The estimators are abbreviated as follows: (1) CSOLS: Cross-section ordinary least squares; (2) CS2SLS: Cross-section two stage least squares; (3) POLS: Panel ordinary least squares; (4) P2SLS: Panel two stage least squares; (5) IFE: pooled interactive fixed effects estimator [Bai, 2009]; (6) IFEMG: mean group interactive fixed effects estimator [Song, 2013]; (7) CCEP: common correlated effects pooled estimator [Pesaran, 2006]; (8) CCEMG: common correlated effects mean group estimator [Pesaran, 2006]; (9) OLSMG: mean group ordinary least squares estimator; (10) CCEP-2: two-step CCEP estimator [Pesaran, 2006]; (11) CCEMG-2: two-step CCEMG estimator [Pesaran, 2006].

Table 3: Cross-Section and Panel OLS and 2SLS Estimates of the Return to Schooling for Males

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Cross-Section		Comparative Sample			Panel	
	OLS	2SLS	OLS	2SLS	OLS	OLS	2SLS
Years of School	0.092***	0.134***	0.095***	0.138***	0.077***	0.105***	0.127***
	(0.004)	(0.025)	(0.002)	(0.035)	(0.005)	(0.003)	(0.016)
Age & Age-Squared	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Person Fixed Effects					Yes	No	No
Year Fixed Effects					No	Yes	Yes
First-State F-Stat		9.19		6.10			184.9
CD Test Stat.					130	7.26	5.66
Observations	3,600	3,600	22,000	22,000	123,000	123,000	123,000

Note: The dependent variable is the log of annual W-2 earnings and self-employment earnings. Columns (1)-(4) are based on a cross-section in 1990. The comparative sample used in columns (3)-(4) is shown in Table 1 Panel B column (2). Years of school is instrumented for with quarter of birth indicator variables interacted with year of birth indicator variables in columns (2), (4) and (7). Standard errors are shown in parentheses and heteroskedasticity-robust for cross-section and clustered at the person level for panel. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*.

Table 4: Common Factor Model Estimates of the Return to Schooling for Males - Pooled Model

	(1) IFE	(2) IFE	(3) CCEP	(4) CCEP	(5) CCEP-2	(6) CCEP-2
Years of School	0.020***	0.026**	0.038***	0.037***	0.023***	0.024***
	(0.003)	(0.003)	(0.004)	(0.006)	(0.004)	(0.004)
Age & Age-Squared	Yes	Yes	Yes	Yes	Yes	Yes
Person Fixed Effects	Yes	No	Yes	No	Yes	No
Year Fixed Effects	No	Yes	No	Yes	No	Yes
Observations	123,000	123,000	123,000	123,000	123,000	123,000

Note: Columns (1)-(2) are based on Interactive Fixed Effects (IFE) (Bai, 2009) with 8 and 7 factors, respectively, selected by the IC_{p1} procedure in Bai and Ng (2002). Columns (3)-(4) are based on Common Correlated Effects pooled (CCEP) (Pesaran, 2006). Columns (5)-(6) are based on the two-step CCE procedure with 7 and 8 factors, respectively, selected by the IC_{p1} procedure in Bai and Ng (2002) applied to residuals based on the CCEP estiamtes. IFE and CCE standard errors are calculated following Bai (2009) and Pesaran (2006), respectively. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*.

Table 5: Common Factor Model Estimates of the Return to Schooling for Males - Heterogeneous Model

	(1) OLSMG	(2) IFEMG	(3) CCEMG	(4) CCEMG-2
Years of School	0.078***	0.028***	0.044***	0.041***
	(0.006)	(0.003)	(0.006)	(0.006)
Age & Age-Squared	Yes	Yes	Yes	Yes
Person Fixed Effects	Yes	Yes	Yes	Yes
Su-Chen Slope Test		34.19	21.6	21.48
Ando-Bai Slope Test		5324	1422	-73.63
Observations	123,000	123,000	123,000	123,000
Percent of individuals with negative returns	0.392	0.462	0.428	0.442

Note: The panel OLS, IFE, CCEP, and CCEP-2 estimators from Table 3 and 4 are replaced with versions that allow regression coefficients to vary across individuals (Pesaran, 2006; Pesaran and Smith, 1996; Song, 2013). The individual-level regression coefficients are then averaged across all individuals to produce a "mean group" (MG) result. IFEMG estimates are based on 4 factors for columns (2), selected by the IC_{p1} procedure in Bai and Ng (2002). Two-step CCE estimates are based on 4 factors for columns (4), selected by the IC_{p1} procedure in Bai and Ng (2002) applied to residuals based on the CCEMG estimates. The Su-Chen and Ando-Bai slope homogeneity tests are based on Su and Chen (2013) and Ando and Bai (2015).

Table 6: Bias Associated with OLS (Pooled and Heterogeneous) and 2SLS (Pooled) Estimates, Due to Common Factor Structure

	(1) Factor skill 1	(2) Factor skill 2	(3) Factor skill 3	(4) Factor skill 4	(5) All others	(6) Total
IFE						
A. Covariates: Age controls and person fixed effects						
OLS	0.037	0.004	0.011	0.002	0.002	0.057
B. Covariates: Age controls and year fixed effects						
OLS	0.080	-0.001	-0.001	0.0005	0.0002	0.079
2SLS	0.075	0.025	0.001	-0.001	-0.0001	0.101
CCEP						
A. Covariates: Age controls and person fixed effects						
OLS	0.023	0.013	0.003	0.001	0.001	0.041
B. Covariates: Age controls and year fixed effects						
OLS	0.068	-0.0005	0.0005	0.0004	0.0001	0.068
2SLS	0.078	0.012	0.0003	-0.0001	-0.00003	0.090
CCEP-2						
A. Covariates: Age controls and person fixed effects						
OLS	0.042	0.002	0.006	0.002	0.002	0.054
B. Covariates: Age controls and year fixed effects						
OLS	0.079	0.0005	0.001	0.001	0.0003	0.081
2SLS	0.084	0.019	0.0004	-0.0001	-0.000008	0.103
IFEMG						
A. Covariates: Age controls and intercept						
OLSMG	0.047	0.003	-0.0008	0.0006		0.050
CCEMG						
A. Covariates: Age controls and intercept						
OLSMG	0.017	0.006	0.006	0.005		0.034
CCEMG-2						
A. Covariates: Age controls and intercept						
OLSMG	0.018	0.008	0.006	0.005		0.037

Note: In the top three panels with pooled specifications, bias estimates for OLS are based on the part of years of school that is unexplained by the covariates listed in the panel title. Similarly, bias estimates for 2SLS are based on the part of quarter of birth indicators interacted with year of birth indicators that is unexplained by the other covariates listed in the panel title. The OLS and 2SLS estimates correspond to the specifications in Table 3 that include the covariates listed in the panel title. The common factors in the IFE table panels are based on the IFE results, and in the CCE panels are based on the principal components procedure applied to residuals from the CCEP estimates in Table 4 that correspond to the specifications in the panel titles. Column (5) includes common factors up to the number indicated in Table 4 and 5. In the bottom three panels with heterogeneous models, bias estimates are based on the common factor estimates from the heterogeneous factor model results in Table 5. The CCE common component estimates are based on applying principal components to $\eta_{it} = y_{it} - s_{it}\hat{\beta}_{CCE}$.

Table 7: Mean and Variance of Heterogeneous Model Estimates by Characteristic Group

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OLSMG		IFEMG		CCEMG		CCEMG-2		Group
	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance	Size
A. Race									
White	0.080	0.110	0.028	0.143	0.045	0.124	0.041	0.122	3,400
Black	0.032***	0.085	-0.001	0.147	0.013	0.118	0.011	0.108	200
Other Race	0.084	0.146	0.085	0.181	0.068	0.177	0.082	0.171	100
B. Hispanic Status									
Non-Hispanic	0.079	0.108	0.027	0.143	0.043	0.125	0.040	0.122	3,500
Hispanic	0.063	0.134	0.067	0.170	0.060	0.134	0.065	0.154	150
C. Foreign Born Status									
Native	0.078	0.108	0.027	0.142	0.045	0.124	0.041	0.121	3,500
Foreign Born	0.072	0.149	0.049	0.189	0.009	0.145	0.037	0.178	100
D. Birth Cohort									
Born before 1950	0.002	0.061	-0.001	0.112	0.001	0.087	0.003	0.078	400
Born 1950-1954	0.022	0.079	-0.008	0.110	0.021	0.086	0.018	0.091	1,000
Born 1955-1959	0.091***	0.105	0.031***	0.129	0.05**	0.118	0.042*	0.124	1,700
Born after 1959	0.207***	0.196	0.113***	0.274	0.102***	0.249	0.111***	0.211	500
E. Highest Education Level Completed									
High School	0.198	0.304	0.141	0.401	0.107	0.356	0.116	0.284	300
Some College	0.055***	0.141	-0.001***	0.193	0.014***	0.160	0.016***	0.172	1,500
Bachelor's	0.092***	0.036	0.041***	0.045	0.074***	0.044	0.056***	0.042	850
Graduate Degree	0.064***	0.052	0.027	0.061	0.044***	0.060	0.042	0.059	900

Note: Means and variances reported in the table are characteristic-specific summary statistics of the individual-level coefficients from the heterogeneous model results in Table 5. For Panel A, mean returns for individuals who are Black and other races are both tested against the mean for individuals who are White. For the remaining panels, each subgroup mean is tested against the subgroup mean directly above it, beginning with the second subgroup listed. For example, in Panel D, the mean return for individuals born in 1950-1954 is tested against the mean for individuals born before 1950, the mean for individuals born in 1955-1959 is tested against the mean for individuals born in 1950-1954, and the mean for individuals born after 1959 is tested against the mean for individuals born in 1955-1959.

Table 8: Quantiles of Heterogeneous Model Estimates by Characteristic Group

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
		OLSMG			IFEMG			CCEMG			CCEMG-2	
	p25	p50	p75	p25	p50	p75	p25	p50	p75	p25	p50	p75
All Individuals	-0.076	0.049	0.215	-0.111	0.016	0.169	-0.102	0.034	0.185	-0.103	0.025	0.171
A. Race												
White	-0.075	0.049	0.215	-0.111	0.017	0.168	-0.101	0.035	0.185	-0.104	0.025	0.173
Black	-0.087	0.040	0.169	-0.123	0.001	0.149	-0.114	0.005	0.144	-0.094	-0.006	0.121
Other Race	-0.064	0.075	0.290	-0.091	0.018	0.282	-0.099	0.041	0.254	-0.054	0.071	0.240
B. Hispanic Status												
Non-Hispanic	-0.076	0.049	0.217	-0.111	0.016	0.168	-0.102	0.033	0.184	-0.104	0.025	0.172
Hispanic	-0.081	0.065	0.181	-0.114	0.008	0.204	-0.109	0.047	0.201	-0.094	0.050	0.141
C. Foreign Born Status												
Native	-0.077	0.049	0.216	-0.111	0.016	0.168	-0.102	0.034	0.186	-0.102	0.025	0.172
Foreign Born	-0.061	0.048	0.183	-0.112	0.001	0.204	-0.100	0.030	0.153	-0.120	0.033	0.153
D. Birth Cohort												
Born before 1950	-0.115	0.007	0.113	-0.092	-0.001	0.113	-0.104	0.011	0.113	-0.094	0.011	0.108
Born 1950-1954	-0.106	0.009	0.145	-0.132	-0.009	0.117	-0.101	0.022	0.140	-0.112	0.007	0.121
Born 1955-1959	-0.066	0.071	0.230	-0.101	0.025	0.173	-0.093	0.043	0.190	-0.100	0.036	0.179
Born after 1959	0.002	0.159	0.384	-0.112	0.079	0.336	-0.132	0.079	0.312	-0.111	0.069	0.317
E. Highest Education Level Completed												
High School	-0.056	0.135	0.411	-0.178	0.087	0.487	-0.165	0.064	0.389	-0.136	0.063	0.400
Some College	-0.130	0.028	0.220	-0.176	0.003	0.182	-0.178	0.014	0.188	-0.159	0.011	0.180
Bachelor's	-0.028	0.070	0.208	-0.062	0.021	0.143	-0.048	0.055	0.186	-0.056	0.031	0.155
Graduate Degree	-0.057	0.046	0.168	-0.071	0.017	0.135	-0.061	0.033	0.147	-0.067	0.027	0.142

Note: Quantiles reported in the table are characteristic-specific summary statistics of the individual-level coefficients from the heterogeneous model results in Table 5. p25 = 25th percentile, p50 = 50th percentile, p75 = 75th percentile. Sample sizes for each group are shown in Table 7.

Appendices

A Bias Derivations

A.1 Omitted Ability Bias

We employ the IFE and CCE estimators to derive analytical expressions for and estimates of the biases induced by the OLS and IV estimators assuming that the true model is given by (1) and (2) in section 3.1. The interactive effects framework allows us to not only obtain estimates of the aggregate ability bias but also the bias attributable to each of the ability components. To simplify the exposition, we consider a setup where ability is two-dimensional ($r = 2$) and the regression coefficients are homogeneous.[32] The model is given by

$$y_{it} = s_{it}\beta + \lambda_{1i} f_{1t} + \lambda_{2i} f_{2t} + u_{it} \tag{A.1}$$

where y_{it} (s_{it}) is the residual obtained by regressing log wages (schooling) on the set of controls and a full set of time and person dummies. Note that given the set of dummies included, the means of y_{it} and s_{it} across i and t as well as their overall means (over i and t) are all zero. Let $c_{j,it} = \lambda_{ji} f_{jt}$ be the common component associated with factor j ($j = 1, 2$).

The probability limit of the OLS estimator can be expressed as

$$\begin{aligned} p\lim \hat{\beta}_{OLS} &= [Var(s_{it})]^{-1} Cov(s_{it}, y_{it}) \\ &= \beta + [Var(s_{it})]^{-1} Cov(s_{it}, c_{1,it}) + [Var(s_{it})]^{-1} Cov(s_{it}, c_{2,it}) \\ &= \beta + B1_{ols} + B2_{ols} \end{aligned} \tag{A.2}$$

[32]The ability bias associated with the OLSMG estimator is derived in the Appendix A3.

where

$$Var(s_{it}) = plim_{N,T\to\infty}(NT)^{-1}\sum_t\sum_i s_{it}^2 \tag{A.3}$$

$$Cov(s_{it}, c_{j,it}) = plim_{N,T\to\infty}(NT)^{-1}\sum_t\sum_i s_{it}\lambda_{ji}f_{jt} \tag{A.4}$$

In $(A.2)$, $B1_{ols}$ can be interpreted as the bias in the OLS estimator induced by f_1 and $B2_{ols}$ the bias induced by f_2. The aggregate OLS bias is given by

$$Bias(\hat{\beta}_{OLS}) = p\lim\hat{\beta}_{OLS} - \beta = B1_{ols} + B2_{ols} = B_{ols}$$

Now consider a 2SLS estimator based on a set of K instruments z_{it} (as before, z_{it} is the residual from regressing the instruments on the set of controls and a full set of time and person dummies.) where $Cov(z_{it,k}, s_{it}) \neq 0$ where $k = 1,...,K$. Define the $(T\times 1)$ vector $Y_i = (Y_{i1},...,Y_{iT})'$, the $(T\times K)$ matrix $Z_i = (z_{i1},...,z_{iT})'$ and the $(NT\times K)$ matrix $Z = (Z_1',...,Z_N')'$. The first stage estimate is $\hat{\Pi} = \left(\sum_{i=1}^N Z_i'Z_i\right)^{-1}\sum_{i=1}^N Z_i'S_i$. The 2SLS estimate is

$$\hat{\beta}_{2SLS} = \left(\hat{\Pi}'\sum_{i=1}^N Z_i'Z_i\hat{\Pi}\right)^{-1}\left(\hat{\Pi}'\sum_{i=1}^N Z_i'Y_i\right)$$

Denote $\hat{S}_i = Z_i\hat{\Pi}$. Then we have

$$\begin{aligned} p\lim\hat{\beta}_{2SLS} &= \beta + [p\lim(NT)^{-1}\sum_{i=1}^N \hat{S}_i'\hat{S}_i]^{-1}\left\{\begin{array}{c} p\lim(NT)^{-1}\left[\sum_i \hat{S}_i'F_1\lambda_{1i}\right] \\ +p\lim(NT)^{-1}\left[\sum_i \hat{S}_i'F_2\lambda_{2i}\right]\end{array}\right\} \\ &= \beta + B1_{iv} + B2_{iv} \end{aligned} \tag{A.5}$$

In $(A.5)$, $B1_{iv}$ can be interpreted as the bias in the 2SLS estimator induced by f_1 and $B2_{iv}$ the bias induced by f_2. The aggregate 2SLS bias is given by

$$Bias(\hat{\beta}_{2SLS}) = p\lim \hat{\beta}_{2SLS} - \beta = B1_{iv} + B2_{iv} = B_{iv}$$

The 2SLS estimator has a larger aggregate bias than the OLS estimator if $B_{iv} > B_{ols}$ or

$$B2_{iv} - B2_{ols} > B1_{ols} - B1_{iv} \tag{A.6}$$

In accordance with our empirical results, we assume that $B1_{ols} + B2_{ols} = B_{ols} > 0$. We consider the following two cases depending on the magnitude and direction of the component-specific biases that turn out to be relevant in our context:

- **Case A:** $B1_{ols} > 0$, $B2_{ols} < 0$ such that $B1_{ols} > |B2_{ols}|$. Then $\hat{\beta}_{OLS}$ is upward biased with the positive bias induced by f_1 dominating the negative bias induced by f_2:

$$Bias(\hat{\beta}_{OLS}) = p\lim \hat{\beta}_{OLS} - \beta = B1_{ols} + B2_{ols} = B_{ols} > 0$$

The inequality $(A.6)$ is consistent with any of the following four scenarios:

1. IV is effective in reducing the magnitude of the bias from *both* components: $|B2_{iv}| < |B2_{ols}|$, $|B1_{iv}| < B1_{ols}$.

2. IV is effective in reducing the magnitude of the bias from component 1 *only*: $|B2_{iv}| > |B2_{ols}|$, $|B1_{iv}| < B1_{ols}$.

3. IV is effective in reducing the magnitude of the bias from component 2 *only*: $|B2_{iv}| < |B2_{ols}|$, $B1_{iv} > B1_{ols}$.

4. IV is completely ineffective: $|B2_{iv}| > |B2_{ols}|$, $B1_{iv} > B1_{ols}$.

In general, if ability is multidimensional and one of its components is negatively correlated with schooling, it is possible for the aggregate 2SLS bias to exceed the aggregate OLS bias *regardless* of whether the instruments are fully, partially or not effective in reducing the magnitude of the bias in any or all of its components.

- **Case B:** $B1_{ols} > 0$, $B2_{ols} > 0$

The inequality $(A.6)$ is consistent with any of the following three scenarios:

1. IV is effective in reducing the magnitude of the bias from component 1 *only*: $B2_{iv} > B2_{ols}$, $B1_{iv} < B1_{ols}$.

2. IV is effective in reducing the magnitude of the bias from component 2 *only*: $B2_{iv} < B2_{ols}$, $B1_{iv} > B1_{ols}$.

3. IV is completely ineffective: $B2_{iv} > B2_{ols}$, $B1_{iv} > B1_{ols}$.

In contrast to case A, if each of the ability components induce a positive bias in the OLS estimates, the instruments can be (at most) effective at reducing the bias associated with only a *subset* of the components at the expense of exacerbating the bias associated with the remaining components, for $(A.6)$ to hold.

Under the factor model framework $(A.1)$, each of the bias terms in $(A.2)$ and $(A.5)$ can be consistently estimated. This is because even though the factors and their loadings are not separately identified, their product, i.e., the common components $(c_{j,it})$ are. The estimated biases can be obtained as follows:

$$
\begin{aligned}
\widehat{B1}_{ols} &= [SVar(s_{it})]^{-1} SCov(s_{it}, \hat{c}_{1,it}) \\
\widehat{B2}_{ols} &= [SVar(s_{it})]^{-1} SCov(s_{it}, \hat{c}_{2,it}) \\
\widehat{B1}_{iv} &= [SVar(\hat{S}_i)]^{-1} SCov(\hat{S}_i, \hat{F}_1 \hat{\lambda}_{1i}) \\
\widehat{B2}_{iv} &= [SVar(\hat{S}_i)]^{-1} SCov(\hat{S}_i, \hat{F}_2 \hat{\lambda}_{2i})
\end{aligned}
$$

where, for $j = 1,2,\ \hat{c}_{j,it} = \hat{\lambda}_{ji}\hat{f}_{jt}$ are the Bai (2009) estimates of the common components and $SVar(s_{it}),\ SVar(\hat{S}_i),\ SCov(\hat{S}_i, \hat{F}_j\hat{\lambda}_{ji}),\ SCov(s_{it}, \hat{c}_{j,it})$ denote the sample variance and sample covariances respectively, which are the sample analogs of the quantities defined in $(A.3)$ and $(A.4)$. Specifically, these quantities are computed as follows:

$$SVar(s_{it}) = (NT)^{-1}\sum_t\sum_i s_{it}^2 \tag{A.7}$$

$$SVar(\hat{S}_i) = (NT)^{-1}\sum_{i=1}^{N}\hat{S}_i'\hat{S}_i \tag{A.8}$$

$$SCov(\hat{S}_i, \hat{F}_j\hat{\lambda}_{ji}) = (NT)^{-1}\left[\sum_i \hat{S}_i'\hat{F}_j\hat{\lambda}_{ji}\right] \tag{A.9}$$

$$SCov(s_{it}, \hat{c}_{j,it}) = (NT)^{-1}\sum_t\sum_i s_{it}\hat{\lambda}_{ji}\hat{f}_{jt} = T^{-1}\sum_t\left\{N^{-1}\sum_i s_{it}\hat{\lambda}_{ji}\hat{f}_{jt}\right\} \tag{A.10}$$

Note that in $(A.7\text{-}A.10)$, we do not need to subtract the means since the variables already have mean zero. Note that $SCov(s_{it}, \hat{c}_{j,it})$ is the average (over time) of the cross-sectional correlation between s_{it} and $\hat{c}_{j,it}$. Each of the terms in $(A.7\text{-}A.10)$ can be computed based on our data and factor model estimates to examine the extent to which the component-specific biases offset or reinforce each other.

The CCE approach does not directly estimate the factors so we employ the following two-step procedure to estimate the ability bias components: (1) Obtain the residuals $\eta_{it} = y_{it} - s_{it}\hat{\beta}_{CCE}$, where $\hat{\beta}_{CCE}$ is either the CCEP or CCEMG estimator depending on whether one estimates a pooled or heterogeneous model; (2) Given a choice of the number of factors, estimate the common factor model $\eta_{it} = \lambda_i' f_t + u_{it}$ by principal components. Once the factor structure estimates are obtained, the biases attributable to each of the skill components can be estimated as discussed for the IFE estimator. Note that since the CCE procedure proxies for the factors using cross-section averages of the variables, the aggregate bias estimated using the two-step procedure will not necessarily equal the difference between the OLS and the CCEP (or CCEMG). Our empirical results indicate that the difference is, however, minimal. For the CCEP-2 and CCEMG-2 estimates, the biases can be computed in the same way as for IFE and IFEMG, respectively.

A.2 Heterogeneity Bias

Heterogeneity bias arises when one estimates a pooled specification when the regression coefficients are in fact heterogeneous across the cross-section units. To analyze this source of bias, we consider the IFE estimator of Bai (2009). We can write (1) as

$$Y_i = S_i\beta_i + F\lambda_i + U_i \tag{A.11}$$

with Y_i, S_i, U_i being $(T\times 1)$ vectors defined as $Y_i = (y_{i1},...,y_{iT})'$, $S_i = (s_{i1},...,s_{iT})'$, $U_i = (u_{i1},...,u_{iT})'$ and $F = (f_1,...,f_T)'$ being the $(T\times r)$ matrix of common factors. Here we interpret y_{it} (s_{it}) as the part of log wages (schooling) unexplained by the controls w_{it} and person/time fixed effects.

The IFE estimator is given by

$$\hat{\beta}_{IFE} = \left(\sum_{i=1}^{N} S_i' M_{\hat{F}} S_i\right)^{-1}\left(\sum_{i=1}^{N} S_i' M_{\hat{F}} Y_i\right) \tag{A.12}$$

where $M_{\hat{F}} = I_T - \hat{F}\left(\hat{F}'\hat{F}\right)^{-1}\hat{F}'$, and $\hat{F}$ is the principal components (PC) estimate of F.

Under the heterogeneous model $(A.11)$, we can write $(A.12)$ as

$$\begin{aligned}
\hat{\beta}_{IFE} &= \left(\sum_i S_i' M_{\hat{F}} S_i\right)^{-1}\sum_i S_i' M_{\hat{F}}\left(S_i\beta_i + F\lambda_i + U_i\right) = \left(\sum_i S_i' M_{\hat{F}} S_i\right)^{-1}\sum_i S_i' M_{\hat{F}}\left(S_i\beta_i + (F-\hat{F})\lambda_i + \hat{F}\lambda_i + U_i\right)\\
&= \left(\sum_i S_i' M_{\hat{F}} S_i\right)^{-1}\sum_i S_i' M_{\hat{F}} S_i\beta_i + \left(\sum_i S_i' M_{\hat{F}} S_i\right)^{-1}\left(\sum_i S_i' M_{\hat{F}}(F-\hat{F})\lambda_i + \sum_i S_i' M_{\hat{F}} U_i\right)\\
&\underset{N,T\ large}{\simeq} \left(\sum_i S_i' M_{\hat{F}} S_i\right)^{-1}\sum_i S_i' M_{\hat{F}} S_i\beta_i
\end{aligned}$$

where the approximation in the last line holds since the other terms are negligible for large N,T [Bai, 2009]. This gives

$$\hat{\beta}_{IFE} \underset{N,T\ large}{\simeq} \sum_i \omega_i \beta_i \tag{A.13}$$

where $\omega_i = \left(\sum_i S_i' M_{\hat{F}} S_i\right)^{-1} S_i' M_{\hat{F}} S_i$ is the weight on the individual i's return (note that $\sum_i \omega_i = 1$). This suggests that $\hat{\beta}_{IFE}$ is likely to exceed $\hat{\beta}_{IFEMG}$ (since $\hat{\beta}_{IFEMG}$ is an estimate of $N^{-1}\sum_i \beta_i$) if there exists positive correlation between β_i and ω_i, i.e., marginal returns are higher for those individuals who have higher time variation in the unexplained portion of schooling. This can be verified empirically by computing the cross-sectional correlation between $\hat{\beta}_i$ (the individual-specific IFE estimate) and ω_i.

A.3 Bias in the OLS Mean Group [OLSMG] Estimator

The aggregate bias in the OLSMG estimator (based on the IFE approach) can be expressed as

$$\begin{aligned} \hat{\beta}_{OLSMG} - \hat{\beta}_{IFEMG} &= N^{-1}\sum_i \left\{ \left(\sum_t S_{it}^2\right)^{-1} \sum_t S_{it} \hat{\lambda}_i' \hat{f}_t \right\} \\ &= \sum_{j=1}^{r} \left[N^{-1} \sum_i \left\{ \left(\sum_t S_{it}^2 \sum_t S_{it} \hat{\lambda}_{ji} \hat{f}_{jt} \right)^{-1} \right\} \right] \end{aligned} \tag{A.14}$$

assuming r common factors. In $(A.14)$, $\hat{\lambda}_i = (\hat{\lambda}_{1i}, \hat{\lambda}_{2i}, ..., \hat{\lambda}_{ri})'$ so that $\hat{\lambda}_{ji}$ represents the j-th factor loading for individual i. The contribution of the j-th factor to the aggregate bias is therefore

$$N^{-1}\sum_i \left\{ \left(\sum_t S_{it}^2 \sum_t S_{it} \hat{\lambda}_{ji} \hat{f}_{jt} \right)^{-1} \right\}$$

For the CCE approach, since the factors are not directly estimated, we follow a two-step procedure to estimate the component-specific biases as described in Appendix A1. The only difference is that the residuals in the first step are now computed using $\hat{\beta}_{CCEMG}$.

A.4 Bias with Quadratic Schooling Terms

The general interactive fixed effects model with quadratic schooling terms is given by

$$y_{it} = s_{it}\beta_1 + s_{it}^2\beta_2 + v_{it} \tag{A.15}$$

$$v_{it} = \lambda_i' F_t + u_{it} \tag{A.16}$$

Let $B = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ and $W = \begin{bmatrix} s & s^2 \end{bmatrix}$, therefore $\hat{B}_{OLS} = (W'W)^{-1} W'Y = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$. The OLS marginal effect of schooling $= \hat{\beta}_1 + 2\hat{\beta}_2\bar{s} = \hat{B}'_{OLS}\bar{S}_s$, where $\bar{S}_s = \begin{pmatrix} 1 \\ 2\bar{s} \end{pmatrix}$. Thus, the OLS bias equals to

$$\left[E(\hat{B}_{OLS}) - B\right]' \bar{S}_s \tag{A.17}$$

and $\left[E(\hat{B}_{OLS}) - B\right]$ can be estimated by $(W'W)^{-1} W'\hat{C}$ where $W'\hat{C} = \sum w_{it}\hat{c}_{it}$ and $\hat{c}_{it} = \hat{\lambda}_i'\hat{F}_t$.

B Accounting for Experience

Consider the pooled specification

$$y_{it} = c_i + s_{it}\beta + e_{it}\rho_1 + e_{it}^2\rho_2 + \lambda_i{}' f_t + u_{it} \tag{A.18}$$

where e_{it} denotes actual experience and s_{it} denotes schooling. Let $e_{it} = e_{i0} + t$, where e_{i0} is initial experience and t is the time trend. Therefore,

$$y_{it} = c_i + s_{it}\beta + (e_{i0} + t)\rho_1 + (e_{i0} + t)^2\rho_2 + \lambda_i{}' f_t + u_{it}$$

or,

$$y_{it} = (c_i + e_{i0}\rho_1 + e_{i0}^2\rho_2) + (2e_{i0}\rho_2)t + (\rho_1 t + \rho_2 t^2) + s_{it}\beta + \lambda_i{}' f_t + u_{it}$$

or,

$$y_{it} = \widetilde{\rho}_{1i} + \widetilde{\rho}_{2i}t + \widetilde{\delta}_t + s_{it}\beta + \lambda_i{}' f_t + u_{it} \tag{A.19}$$

where

$$\begin{aligned} \widetilde{\rho}_{1i} &= c_i + e_{i0}\rho_1 + e_{i0}^2\rho_2,\ \widetilde{\rho}_{2i} = 2e_{i0}\rho_2 \\ \widetilde{\delta}_t &= \rho_1 t + \rho_2 t^2 \end{aligned}$$

Thus, from $(A.19)$ in the pooled model, besides time fixed effect, we should include person fixed effects and person-specific linear trend, which is equivalent to a pooled model that includes person fixed effects, age and age-squared terms instead of the person-specific linear trend.

In the heterogeneous model,

$$y_{it} = c_i + s_{it}\beta_i + e_{it}\rho_{1i} + e_{it}^2\rho_{2i} + \lambda_i{}' f_t + u_{it}$$

or

$$y_{it} = \breve{\rho}_{1i} + \breve{\rho}_{2i}t + \rho_{2i}t^2 + s_{it}\beta_i + \lambda_i{}' f_t + u_{it} \tag{A.20}$$

where

$$\breve{\rho}_{1i} = c_i + e_{i0}\rho_{1i} + e_{i0}^2\rho_{2i},\ \breve{\rho}_{2i} = \rho_{1i} + 2e_{i0}\rho_{2i}$$

From $(A.20)$, we should include person fixed effects and person-specific quadratic trend, which is equivalent to a heterogeneous specification that includes person fixed effects, age and age-squared terms instead of the person-specific quadratic trend.

C Data: Schooling Variable Construction

We construct a longitudinal years of schooling variable based on the SIPP education information that includes highest education level completed ('no high school degree', 'high school degree', 'some college', 'college degree', and 'graduate degree'), the year during which high school was completed, the year during which post-high school education began, the year during which post-high school education ended, and the year during which a bachelor's degree was earned. First, individuals were assigned one of the five highest-level-completed values for each year.[33] All individuals were assigned 'no high school degree' before the year they graduated high school and 'high school degree' beginning in their graduation year. Individuals whose highest completed level was 'some college' and thus did not obtain a bachelor's degree were assigned 'some college' beginning in the year their post-high school education ended. Individuals who obtained at least a college degree were assigned 'college degree' beginning in the year they obtained their bachelor's degree. Individuals who obtained a graduate degree were assigned 'graduate degree' beginning in the year their post-high school education ended.[34] Then, based on highest level completed at each year, individuals were assigned a years of schooling value. Individuals with 'no high school degree' in a given year were assigned 10 years of school, individuals with 'high school degree' were assigned 12 years, individuals with 'some college' were assigned 14 years, individuals with 'college degree' were assigned 16 years, and individuals with 'graduate degree' were assigned 18 years.[35]

Another approach is to measure actual years spent in school, regardless of completed educa-

[33] 'Some college' includes anything less than a bachelor's degree. Thus it includes both individuals with some years of college but no degree and individuals with an associate's degree.

[34] Note that the variable for the year post-high school education ended could be before, the same as, or after the year a bachelor's degree was earned. If a person started college but did not obtain a bachelor's degree, then it indicates when the person dropped out or obtained a shorter degree. If a person obtained a bachelor's and then stopped, then it is the same as the bachelor's year variable. If the person obtained a graduate degree, then it indicates when they finished graduate school.

[35] Assigning years of school based on highest level completed is common in the literature (e.g., Heckman, Lochner, and Todd, 2006; Henderson et al., 2011).

tion levels. This is not feasible in the U.S. Census Bureau GSF as it is in some other datasets such as the NLSY, although it is not obvious that this approach would be preferable: variation in years of school that is independent of completed education levels (e.g., individuals who complete college in three versus five years) might introduce more measurement error or bias into the variable. However, we did want to attempt to smooth the discrete jumps described above for two reasons. First, the scheme introduces measurement error by explicitly missing some variation in years of school. For example, it misses the transition through high school by only assigning 10 years for any year before high school degree completion. It also misses the distinction between individuals working with a high school degree with versus without college experience, because the years of schooling variable does not increase until the individual either finishes their post-high school schooling or obtains a bachelor's degree. Second, because we have to limit the main sample to individuals with at least one change in schooling (and further limit to individuals with at least two changes in schooling in Appendix D), this allows us to retain a few more individuals. We therefore make the following two adjustments to smooth the years of schooling variable: (1) we change years of schooling from 10 to 11 the year before a high school degree was finished, which captures progression from 10th grade through 12th[36]; and (2) we change years of schooling from 12 to 13 beginning the year when an individual begins their post-high school education, which captures the distinction between an individual working with a high school degree with versus without college experience. Our main sample of analysis in Panel A column (5) of Table 1 has the following distribution of within-person changes in years of schooling: 250 changes from 10 years to 11; 500 changes from 11 to 12; 900 changes from 12 to 13; 1,500 changes from 13 to 14; 1,100 changes from 13 to 16; and 900 changes from 16 to 18.

[36]Our sample is limited to individuals at least 16 years of age, so we do not expect to capture many individuals in grades earlier than 10th.

D Robustness Checks

D.1 Robustness to Alternative Specifications

In this section we discuss robustness of the main results to alternative specifications. Our main results are based on a linear years of schooling and quadratic age specification. This specification is the traditional model originating from Mincer (1974). However, numerous studies have indicated that this specification may not be flexible enough and higher-order terms in schooling and/or experience may be needed (Murphy and Welch, 1990; Heckman, Lochner, and Todd, 2006; Cho and Phillips, 2018). These papers provide evidence supporting the use of up to a quadratic term in years of schooling and a quartic term in experience.[37]

The robustness of our main results in Tables 3-5 to the inclusion of a quadratic years of schooling term and/or a quartic age term are shown in Tables D1-D3, respectively. The sample for specifications that include a quadratic in years of schooling is further restricted to individuals with at least two changes in years of schooling so that we can estimate quadratic terms for the individual-level regressions associated with the heterogeneous models. The marginal returns shown in the tables for specifications with a quadratic years of schooling term are evaluated at the mean level of schooling in the whole sample for the pooled models. For the heterogeneous models, we compute each individual's return based on their mean schooling, and then average the returns across individuals. The bias estimates according to the derivation in Appendix A4 are similar to the results in Table 6 and available upon request. All of the results are very similar to those in the main text, suggesting that our findings are not sensitive to the assumption of a linear relationship between schooling and earnings or a quadratic relationship between age and earnings.

[37]Notably, however, Cho and Phillips (2018) find that the original Mincer specification is appropriate when no additional explanatory variables are included beyond years of school and experience, as is the case in our specifications.

Table D1: OLS and 2SLS Specification Robustness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Cross-Section		Comparative Sample			Panel	
	OLS	2SLS	OLS	2SLS	OLS	OLS	2SLS
A. Quadratic Schooling and Quadratic Age							
Years of School	0.078***	0.124***	0.092***	0.112***	0.069***	0.093***	0.101***
	(0.006)	(0.021)	(0.002)	(0.038)	(0.007)	(0.005)	(0.017)
Person Fixed Effects					Yes	No	No
Year Fixed Effects					No	Yes	Yes
First-State F-Stat		3.12		6.10			188.3
CD Test Stat.					71.43	-2.13	-2.12
Observations	1,300	1,300	22,000	22,000	45,000	45,000	45,000
B. Linear Schooling and Quartic Age							
Years of School	0.091***	0.125***	0.095***	0.151***	0.073***	0.105***	0.127***
	(0.004)	(0.027)	(0.002)	(0.035)	(0.005)	(0.003)	(0.016)
Person Fixed Effects					Yes	No	No
Year Fixed Effects					No	Yes	Yes
First-State F-Stat		7.89		1.13			182.4
CD Test Stat.					136.9	7.18	5.68
Observations	3,600	3,600	22,000	22,000	123,000	123,000	123,000
C. Quadratic Schooling and Quartic Age							
Years of School	0.077***	0.122***	0.092***	0.134***	0.060***	0.092***	0.100***
	(0.006)	(0.023)	(0.002)	(0.038)	(0.007)	(0.005)	(0.017)
Person Fixed Effects					Yes	No	No
Year Fixed Effects					No	Yes	Yes
First-State F-Stat		2.58		1.13			184.19
CD Test Stat.					71.25	-2.31	-2.29
Observations	1,300	1,300	22,000	22,000	45,000	45,000	45,000

Note: Each table panel shows robustness of the results in Table 3 to extending the specification to include a quadratic in years of schooling and/or a quartic in age. See Table 3 for additional details.

Table D2: Common Factor Pooled Model Robustness

	(1) IFE	(2) IFE	(3) CCEP	(4) CCEP	(5) CCEP-2	(6) CCEP-2
A. Quadratic Schooling and Quadratic Age						
Years of School	0.026***	0.023***	0.031***	0.035***	0.025***	0.026***
	(0.006)	(0.006)	(0.008)	(0.008)	(0.008)	(0.007)
Person Fixed Effects	Yes	No	Yes	No	Yes	No
Year Fixed Effects	No	Yes	No	Yes	No	Yes
Observations	45,000	45,000	45,000	45,000	45,000	45,000
B. Linear Schooling and Quartic Age						
Years of School	0.020***	0.026***	0.037***	0.036***	0.024***	0.024***
	(0.003)	(0.003)	(0.004)	(0.010)	(0.004)	(0.005)
Person Fixed Effects	Yes	No	Yes	No	Yes	No
Year Fixed Effects	No	Yes	No	Yes	No	Yes
Observations	123,000	123,000	123,000	123,000	123,000	123,000
C. Quadratic Schooling and Quartic Age						
Years of School	0.026***	0.029***	0.031***	0.034***	0.026***	0.025***
	(0.005)	(0.005)	(0.008)	(0.009)	(0.009)	(0.007)
Person Fixed Effects	Yes	No	Yes	No	Yes	No
Year Fixed Effects	No	Yes	No	Yes	No	Yes
Observations	45,000	45,000	45,000	45,000	45,000	45,000

Note: Each table panel shows robustness of the results in Table 4 to extending the specification to include a quadratic in years of schooling and/or a quartic in age. Columns (1)-(2) are based on 7 and 7 factors in Panel A; 8 and 7 factors in Panel B; and 7 and 6 factors in Panel C, selected by the IC_{p1} procedure in Bai and Ng (2002). Columns (5)-(6) are based on 7 and 8 factors in Panel A; 7 and 8 factors in Panel B; and 7 and 8 factors in Panel C, selected by the IC_{p1} procedure in Bai and Ng (2002) applied to residuals based on the CCEP estimates. See Table 4 for additional details.

Table D3: Common Factor Heterogeneous Model Robustness

	(1) OLSMG	(2) IFEMG	(3) CCEMG	(4) CCEMG-2
A. Quadratic Schooling and Quadratic Age				
Years of School	0.101***	0.030**	0.035**	0.025*
	(0.013)	(0.013)	(0.014)	(0.015)
Person Fixed Effects	Yes	Yes	Yes	Yes
Su-Chen Slope Test		15.39	12.26	12.16
Ando-Bai Slope Test		4,009	1,331	-51.37
Observations	45,000	45,000	45,000	45,000
Percent of individuals with negative returns	0.432	0.489	0.482	0.471
B. Linear Schooling and Quartic Age				
Years of School	0.067***	0.023***	0.030***	0.029***
	(0.006)	(0.007)	(0.006)	(0.006)
Person Fixed Effects	Yes	Yes	Yes	Yes
Su-Chen Slope Test		27.49	21.5	21.39
Ando-Bai Slope Test		7,536	1,724	-84.91
Observations	123,000	123,000	123,000	123,000
Percent of individuals with negative returns	0.410	0.466	0.464	0.472
C. Quadratic Schooling and Quartic Age				
Years of School	0.087**	0.034***	0.032**	0.030*
	(0.013)	(0.013)	(0.014)	(0.017)
Person Fixed Effects	Yes	Yes	Yes	Yes
Su-Chen Slope Test		20.35	12.26	12.21
Ando-Bai Slope Test		5,734	1,545	-57.41
Observations	45,000	45,000	45,000	45,000
Percent of individuals with negative returns	0.461	0.483	0.494	0.492

Note: Each table panel shows robustness of the results in Table 5 to extending the specification to include a quadratic in years of schooling and/or a quartic in age. Column (2) is based on 3 factors in Panel A; 4 factors in Panel B; and 2 factors in Panel C, selected by the IC_{p1} procedure in Bai and Ng (2002). Column (4) is based on 3 factors in Panel A; 3 factors in Panel B; and 3 factors in Panel C, selected by the IC_{p1} procedure in Bai and Ng (2002) applied to residuals based on the CCEMG estimates. Specifications with a quadratic in years of schooling are based on a sample of individuals with at least two changes in schooling, in order to identify quadratic terms from individual-level regressions. See Table 5 for additional details.

D.2 Time-Varying Returns to Demographics as Proxies for Interactive Fixed Effects

Our interpretation of the interactive fixed effects structure as capturing unobserved skills or abilities hinges on the assumption that there are no suitable proxies to fully account for their effects. Alternatively, such a structure could be potentially capturing time-varying returns to time invariant individual-specific characteristics such as demographics, or these characteristics could serve as useful proxies for individual skills or abilities. To investigate this possibility, we estimated the following specification with demographic-by-year fixed effects, denoted $d_i'\theta_t$, by OLS:

$$y_{it} = \delta_t + s_{it}\beta + w_{it}'\gamma + d_i'\theta_t + v_{it}$$

The estimates, reported in columns (1)-(2) in Table D4 below, are only marginally smaller than those reported in columns (5)-(6) in Table 3, which strengthens our interpretation that the interactive fixed effects models capture unobservable skills/abilities that cannot be accounted for using observable characteristics.

Table D4: Time-Varying Returns to Demographics as Proxies for Interactive Fixed Effects

	(1) OLS	(2) OLS
Years of School	0.066*** (0.003)	0.098*** (0.005)
Age & Age-Squared	Yes	Yes
Person Fixed Effects	Yes	No
Year Fixed Effects	No	Yes
Demo-by-Year Fixed Effects	Yes	Yes
CD test stat	23.37	7.79
Observations	123,000	123,000

Note: Columns (1)-(2) are identical to columns (5)-(6) in Table 3, except with demographic-by-year fixed effects included. These additional fixed effects are intended to proxy for the interactive fixed effects structure. That is, whereas a general version of the pooled interactive fixed effects approach estimates $y_{it} = \delta_t + s_{it}\beta + w_{it}'\gamma + \lambda_i' f_t + u_{it}$, here we estimate $y_{it} = \delta_t + s_{it}\beta + w_{it}'\gamma + d_i'\theta_t + v_{it}$. The demographic variables included in d_i are race, Hispanic status, foreign born status, marital status, birth year, and state of residence in the SIPP survey.