Purdue University Economics Department Working Paper No **1309**

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

July 9, 2018

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 2.7 percent relative to ordinary least squares and two stage least squares estimates which lie in the range 10.7-44.4 percent. A decomposition of the aggregate least squares bias shows that the omitted ability component is responsible

*We are grateful to Timothy N. Bond, Christopher Kilby, Junsoo Lee, Miguel Sarzosa, Justin Tobias, and seminar participants at the Census Economic Research workshop and the Villanova School of Business for helpful comments. 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. Data from the SIPP Gold Standard File are confidential. All results have been formally reviewed to ensure that no confidential Census Bureau data have been disclosed. All statistics have been rounded in accordance with disclosure review standards.

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

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.

Keywords: Returns to Schooling; Interactive Fixed Effects; Multidimensional Skills; Individual Heterogeneity

JEL Classification: C38, I26

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 (Pedace and Bates, 2000; Cristia and Schwabish, 2009; Chenevert et al., 2016). 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; Koop and Tobias, 2004; Ashworth et al., 2017; 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 either the quarter of birth or its interaction 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 a model with heterogeneous coefficients yields a point estimate of the average marginal returns to schooling of about 2.7 percent relative to OLS and two stage least squares (2SLS) estimates which lie in the range 10.7-44.4 percent. While both omitted ability and heterogeneity biases contribute to the overall OLS bias, a decomposition of the aggregate

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

least 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" rather than diminishing marginal returns to years of schooling. Finally, we uncover considerable within-group 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.

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. 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. 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.[3,4] 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.

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

[3] 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.

[4] 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.

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). 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.[5] 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 (see Heckman, Stixrud, and Urzua, 2006). 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 interactive fixed effects estimates suggest a reduction in the average marginal returns to schooling between 64-95 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 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

[5] 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.

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 earnings ability (Lazear, 1977; Card, 1995). That said, we believe these concerns are mitigated somewhat by the fact that we do replicate well-established results from the cross-section literature; the fact that 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 the fact that other research has shown that the student population who works during school is both large and growing (Hotz et al., 2002; Bacolod and Hotz, 2006; Bound et al., 2012), 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 highlights the intuition of 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

$$y_{it} = c_i + x_{it}\beta_i + w_{it}'\gamma_i + v_{it} \quad (1)$$

$$v_{it} = \lambda_i' f_t + u_{it} \quad (2)$$

where y_{it} and x_{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.[6] Both loadings and the factors are potentially correlated with the observables (x_{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.[7] 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 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.

[6]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.

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

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.

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.

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

$$\hat{v_{it}} = y_{it} - \hat{c}_i - x_{it}\hat{\beta}_i - w_{it}'\hat{\gamma}_i$$

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

$$y_{it} = c_i + x_{it}\beta_i + w_{it}'\gamma_i + \lambda_i'\hat{f}_t + \xi_{it} \tag{3}$$

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 (3) 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.[8]

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

[8]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 (2-4, depending on the specification).

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 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).[9]

4 Data

4.1 Linked Survey-Administrative Data

Linked survey-administrative data come from the U.S. Census Bureau Gold Standard File (GSF). The dataset is based on respondents from the SIPP linked with tax and benefit data from the IRS and SSA.[10] The linked dataset includes respondents' SIPP survey information 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. The SIPP provides detailed social and economic information.[11] This includes the respondent's educational history, including not just the highest level of schooling completed, but also the year each level was completed. Annual earnings comes from the SSA's detailed earnings record, 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.

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

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

[10]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 here: https://www.census.gov/programs-surveys/sipp/guidance/sipp-synthetic-beta-data-product.html.

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

work has shown that earnings data from surveys appears to be overstated at the bottom of the earnings distribution and understated at the top (Pedace and Bates, 2000; Cristia and Schwabish, 2009; Chenevert et al., 2016). 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.

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 and Heckman (2002); Carneiro et al. (2003); Cunha et al. (2005); Carneiro et al. (2011)). 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); Koop and Tobias (2004); Ashworth et al. (2017)). 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;[12] 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 econometric approaches. The use of linked survey-administrative data addresses the former of those recommendations by providing a large, high-quality panel of earnings and education. 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).

[12]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).

4.2 Sample Selection and Summary Statistics

The sample of individuals in the analysis was selected based on seven selection criteria: (1) males; (2) with variation in their years of schooling during 1978-2011; (3) with earnings observations in each year from 1978-2011; (4) without any missing data for the other variables included in the analysis; (5) between the ages of 16-65 during the entirety of 1978-2011; (6) at least 27 years of age at the time of their SIPP survey; and (7) not currently enrolled in school at the time of their SIPP survey. The sample is restricted to males in order to analyze a population that historically is consistently and strongly attached to the labor market, and to be consistent with the majority of the literature. Restricting the sample to those with variation in schooling allows for the estimation of person fixed effects models. The other restrictions generate a balanced sample of prime working years for individuals least likely to have incomplete educational histories in the survey data.

Table 1 shows summary statistics for the variables included in the analysis.[13] The final sample includes 6,300 individuals.[14] Each column corresponds to a different sample used in the analysis below. The full balanced panel sample is shown in column (3). Columns (1) and (2) show two point-in-time cross-sections of the full balanced panel sample used to replicate cross-section estimates from the literature. Column (1) is based on the 6,300 individuals in the year 1990. Column (2) is based on the 6,300 individuals at age 40.[15]

Years of school is a longitudinal variable based on survey responses indicating 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

[13]Survey weights are not used in the sample statistics or regression analysis to re-weight to a nationally representative sample. SIPP survey weights would need to be adjusted not only for the linkage rate to administrative data and missing data, but also for the combination of many SIPP panels into one sample.

[14]All counts are rounded according to U.S. Census Bureau disclosure avoidance rules.

[15]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.

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. Collectively, these variables were used to build a longitudinal schooling variable.

First, individuals were assigned highest-level-completed for each year. 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 college degree were assigned 'some college' beginning in the year their post-high school education began. 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.[16] Then, based on highest level completed at each year, individuals were assigned a years of school variable. 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.[17]

To alleviate the concern that the sample of analysis may not be representative, we investigate a comparison sample that does not require earnings while enrolled in school in Appendix C1. Nonetheless, much of this work while enrolled in school may be part-time work. Thus, earnings

[16]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. 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.

[17]Assigning years of school based on highest level completed is common in the literature (e.g., Heckman, Lochner, and Todd, 2006; Henderson, Polacheck, and Wang, 2011). Another approach is to measure actual years spent in school, regardless of completed education 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. Another set of results not shown here were based on a schooling variable that smoothed the discrete jumps in years of school described above by attempting to impute actual years spent in school. However, doing so was difficult based on available information in some instances, such as for individuals with long periods of time between the beginning and ending of post-high school education; while the educational history variables do report when post-high school education began and ended, it is not possible to know if or when individuals took breaks from school during college or between college and graduate school.

may be artificially low during school, which could lead to larger estimates of the return to schooling from panel datasets. For this reason, the analysis below first estimates cross-section specifications in order to replicate the well-known pattern of OLS/2SLS results from the literature. Furthermore, the full 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 effect of experience is accounted for, whether the regression parameters are allowed to be heterogeneous and whether interactive fixed effects are incorporated.[18] 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 or its interaction with year of birth as instruments (Angrist and Krueger, 1991). 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. Finally, Section 5.5 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.

[18]Note that no tests to determine statistical significance have been performed except where indicated explicitly in the text or tables. Estimates of the return to schooling across the different sets of specifications listed here have not been tested to determine whether they are statistically different from one another.

5.1 Estimated Specifications

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

- **Group 1 [Specifications 1-2]:** Cross-section OLS and 2SLS regressions of log hourly earnings on schooling to verify the "IV > OLS" result commonly found in empirical studies. When age controls are included, we estimate the specification

  $$y_i = c + x_i\beta + w_i'\gamma + a_i\rho_1 + a_i^2\rho_2 + u_i$$

  where w_i is a vector of demographic controls and 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). Demographic controls include race, Hispanic status, foreign born status, marital status, state of residence during the SIPP survey and birth year. We also explore the sensitivity of the results to the omission of age controls.

- **Group 2 [Specifications 3-7]:** Standard panel data specifications that include time and/or person fixed effects to control for unobserved heterogeneity. Here, we estimate five different specifications depending on the type of fixed effects included as well as whether age and demographic controls are included. The most general specification in this group takes the form

  $$y_{it} = \delta_t + x_{it}\beta + w_{it}'\gamma + a_{it}\rho_1 + a_{it}^2\rho_2 + u_{it} \quad (4)$$

  where a_{it} denotes the age of individual i at period t. We consider the following variants of (4): (a) the age and demographic controls are excluded; (b) the age and demographic controls are replaced by a person fixed effect; (c) the demographic controls are excluded; (d) the age controls are excluded. Angrist and Newey (1991) consider a specification of the form

$$y_{it} = c_i + \delta_t + x_{it}\beta + w_{it}'\gamma + pe_{it}\rho_1 + pe_{it}^2\rho_2 + u_{it} \tag{5}$$

where pe_{it} denotes potential experience and is computed as $pe_{it} = a_{it} - x_{it} - 6$, where they define x_{it} as the highest grade completed. They estimate a reduced form schooling effect (expressed as a function of x_{it} and a_{it}) based on the observation that the effect of schooling conditional on potential experience is not identified.[19] 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 (4).

- **Group 3 [Specifications 8-10]:** This group contains specifications that include interactive fixed effects while assuming that the regression coefficients are homogeneous. The nesting model takes the form

$$y_{it} = \delta_t + x_{it}\beta + a_{it}\rho_1 + a_{it}^2\rho_2 + \lambda_i' f_t + u_{it} \tag{6}$$

The following variants of (6) are considered: (a) the age controls are excluded; (b) the age controls are replaced by a person fixed effect.

- **Group 4 [Specifications 11-14]:** This group consists of specifications where the slope parameters are allowed to be individual specific. The general specification is given by

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

We estimate three variants of (7): (a) the interactive fixed effects and age controls are

[19] The returns to schooling literature often controls for potential experience, measured as (age−years of school−6), by assuming that individuals do not work while in school but do work during every other work-age year. This provides a proxy for experience when no direct measure exists. We account for actual experience rather than potential experience because we observe the accumulation of experience based on the presence of annual earnings and we specifically limit our sample to individuals who continue to work and earn income while increasing their schooling.

replaced by a person fixed effect; (b) the person and interactive fixed effects are excluded; (c) the age controls are replaced by a person fixed effect. We consider the specification shown in (7) with 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.

5.2 Cross-Section Estimates

Table 3 presents the cross-section OLS and 2SLS estimation results. Columns (1)-(6) report findings based on the cross-section at year 1990 while columns (7)-(9) report findings based on the cross-section at age 40. OLS results using the first cross-section indicate that the age controls have little impact on the estimated effect of schooling with a point estimate of about 9 percent in either case. The corresponding 2SLS point estimates are much larger when quarter of birth indicators are used as instruments. The estimated schooling effect depends crucially on the instruments used: when age controls are excluded, using the birth quarter as instruments results in a point estimate of about 28.6 percent while using the interactions of quarter of birth with year of birth as instruments yields an estimate of only 13.9 percent. The same pattern of results is observed for the cross-section at age 40 with the 2SLS point estimates exceeding the OLS estimate with the extent of the excess determined by the set of instruments employed.[20]

Overall, these findings are in accordance with the literature summarized in Card (2001) which 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.[21]

[20] The estimates using the comparison sample without earnings-in-school restriction are shown in Appendix C2.

[21] Our 2SLS estimates are slightly larger than those in the literature when only birth quarter is used as an instrument; OLS estimates generally range from 5 to 10 percent, while 2SLS estimates generally range from 10 to 16 percent

5.3 Pooled Estimates

The results from OLS and 2SLS estimation using panel data over 1978-2011 are presented in Table 4. Columns (1)-(3) report the OLS estimates while the 2SLS estimates are reported in columns (4)-(7). The following observations are readily apparent from these findings: (a) similar in spirit to the cross-section analysis, the OLS point estimates are smaller than the 2SLS point estimates across specifications; (b) the age controls only have a minor effect on the OLS estimate, while the 2SLS point estimates are noticeably smaller when age controls are included.[22] In addition to parameter estimates, Table 4 also reports the results of Pesaran's (2015) CD test for the presence of cross-section dependence for each estimated specification.[23] In all cases, 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 5 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 4. For instance, the IFE point estimate when age controls are included is about 4.2 percent while the corresponding OLS and 2SLS estimates are about 11.6 percent and 15.5 percent (or 21.5 percent if quarter of birth indicators are used as instruments), respectively. Under the assumption that the interactive effects specification represents the true model, the pattern of results

(Card, 2001). The slightly larger 2SLS estimates could be due to the fact that we have a sample of continuous earners, which could make returns to schooling appear larger (Lazear, 1977; Card 1995); the fact that our sample is based on individuals born in more recent years than most of the literature and there is evidence that returns to schooling have been increasing over time (e.g., Card and Lemieux, 2001); the fact that our results are based on a different data source than most of the literature, and the only paper of which we are aware that has estimated returns to schooling based on SIPP and administrative data finds slightly larger OLS estimates than the literature (Chenevert et al., 2016); or an odd LATE interpretation for this sample.

[22]See Appendix C3 for the robustness check of Table 4 with demographic controls.

[23]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.

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 assumption that the IV approach suffers from poor instruments that are correlated with unobserved abilities or skills, which the interactive fixed effects specifications can account for. The CCEP point estimates are 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 much closer to the IFE estimates, especially so when age controls are included. Finally, an interesting pattern emerges across the three specifications (columns 1-3) when using the IFE estimate - the estimated number of common factors corresponding to the most general specification (column 3) is one less than that when a person and time fixed effect are included and two less than that when only a time fixed effect is included. This is precisely the pattern that one would expect a priori if the factors are able to pick up the components that are not controlled for in a particular specification. For example, the difference between the specifications in columns 2 and 3 amounts to the presence of a person-specific trend in the latter which is accounted for by the additional factor estimated for the specification with only time and person fixed effects.

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.[24] Table 6 shows 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 exceeds the OLS bias across specifications with the magnitude of the excess being relatively greater for the IFE and two-step CCE approaches, consistent with the findings reported in Tables 4 and 5. While including age controls mitigates the biases to some extent, the magnitudes remain considerable even in this case. For instance, the aggregate OLS bias with age controls using the IFE approach accounts for about 64 percent of the estimated OLS effect of schooling (Table 4, column 3). The

[24] We investigate an alternative interpretation of the factor structure in Appendix C4.

corresponding aggregate 2SLS bias accounts for about 73 percent of the estimated 2SLS effect of schooling (Table 4, column 7). Similar magnitudes are obtained using the two-step CCE approach.

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 50 percent of the bias across specifications/estimators and nearly all of the bias when only 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 when age controls are included and vice-versa. Notably, a sizeable negative bias component (corresponding to the second common factor) emerges in both OLS and 2SLS cases when age controls are included (the exception being the case where the 2SLS bias is computed using the IFE estimates). In the 2SLS case, this component makes a substantial contribution in the (one-step) CCE approach serving to reduce the resulting aggregate estimated bias sufficiently to a value smaller than the estimated bias using the IFE approach, even though the positive bias emanating from the first common component is larger using the former approach. The negative bias component 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 only 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. 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.

5.4 Mean Group Estimates

Table 7 presents results from estimating the specifications 11-14 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 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 7 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.

When the interactive effects are ignored, point estimates of the return to schooling are considerably larger - when age controls are included, the OLSMG point estimate is about 35 percent, which is more than ten times as large as the IFEMG (2.7 percent) and CCEMG-2 (2.9 percent) estimates and larger than three times the CCEMG estimate (9.7 percent). Consistent with the foregoing pooled results, the one-step CCE approach yields a larger point estimate of the average marginal returns to schooling relative to the IFE and two-step approaches which yield very similar estimates. Given that the pooled estimates exceed the corresponding mean group estimates for both the IFE and two-step CCE approaches, we should expect a positive 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 correlations were estimated to be .009 and .005 with and without age controls, respectively, while the corresponding two-step CCE correlations were estimated as .019 and .022, respectively. The one-step CCE results were also in agreement with the predicted signs, except when age controls are included, although in this case the difference between the pooled and mean group estimate was rather small (.3%) [indeed smaller (in absolute value) than the difference between any other pair of estimates in Tables 5 and 7].[26] The pattern of findings for the estimated schooling effect obtained from the IFE and two-step CCE

[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]When age controls are excluded, $Corr(\hat{\beta}_i, \hat{\omega}_i) \approx -.054$ based on the one-step CCE approach, in accordance with $\hat{\beta}_{CCEP} < \hat{\beta}_{CCEMG}$. With age controls, $Corr(\hat{\beta}_i, \hat{\omega}_i) \approx -.008$ but $\hat{\beta}_{CCEP} > \hat{\beta}_{CCEMG}$.

approaches therefore suggest that ignoring potential heterogeneity is likely to induce an upward 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.[27]

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 Table 8. When age controls are excluded, the first common component is responsible for at least 80 percent of the aggregate bias across three estimation approaches. Consistent with the pooled results, the inclusion of age controls only alleviate the aggregate bias to a limited extent: the bias reduction using both the IFE and two-step approaches is about 22 percent while that based on the one-step CCE procedure is about 23 percent. In either case, the aggregate bias is very large: the aggregate OLSMG bias accounts for up to 95 percent of the estimated OLSMG effect of schooling in Table 7, depending on the specification. An important difference with the pooled bias results in Table 6 is that the biases associated with each of the skill components are now positive, regardless of whether one controls for experience. This finding suggests that the emergence of a negative bias component in the pooled case might be a consequence of the failure to incorporate cross-sectional heterogeneity in the returns to schooling.

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 results for our preferred specification that includes age controls, we

[27] The results are available upon request.

find $\hat{\beta}_{POLS} - \hat{\beta}_{IFEMG} \simeq 8.9$ percentage points, $\hat{\beta}_{POLS} - \hat{\beta}_{IFE} \simeq 7.4$ percentage points, $\hat{\beta}_{IFE} - \hat{\beta}_{IFEMG} \simeq 1.5$ percentage points. A similar calculation using the two-step CCE estimate yields $\hat{\beta}_{POLS} - \hat{\beta}_{CCEMG-2} \simeq 8.7$ percentage points, $\hat{\beta}_{POLS} - \hat{\beta}_{CCEP-2} \simeq 7.4$ percentage points, $\hat{\beta}_{CCEP-2} - \hat{\beta}_{CCEMG-2} \simeq 1.3$ percentage points. For the one-step CCE method, we obtain $\hat{\beta}_{POLS} - \hat{\beta}_{CCEMG} \simeq 1.9$ percentage points, $\hat{\beta}_{POLS} - \hat{\beta}_{CCEP} \simeq 1.6$ percentage points, $\hat{\beta}_{CCEP} - \hat{\beta}_{CCEMG} \simeq 0.3$ 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.

5.5 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, 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; Koop and Tobias, 2004; Henderson et al., 2011; 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-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.

5.5.1 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

[28]We also investigate the characteristics associated with extreme returns in Appendix D.

[29]The kernel density plots are based on a standard normal (Gaussian) kernel with a bandwidth of 0.1.

return and a positive 100 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 OLS returns are somewhat left-skewed compared to the FM returns. Finally, despite the relatively large difference between the CCEMG estimate and the IFEMG and CCEMG-2 estimates reported in Table 7, the three FM distributions look very similar; the difference in the CCEMG return appears to be due to relatively small differences along the left tail and right side of the distribution.

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, 13.0 percent of individuals have negative returns in the heterogeneous OLS model, 45.2 percent in the heterogeneous IFE model, 38.4 percent in the heterogeneous CCE model, and 45.9 percent in the heterogeneous CCE-2 model. The OLS fraction is similar to 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 OLS and FM estimates.

5.5.2 Across-Group Heterogeneity

Table 9 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, 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 in-

[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.

dividuals 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 Black compared to White; (2) for individuals born in later birth cohorts; (3) individuals with a high school degree compared to some college; (4) individuals with some college compared to a bachelor's degree; (5) for individuals with a graduate degree compared to a bachelor's degree. These results are broadly consistent with those reported in Henderson et al. (2011).

Mean individual returns from the FMs are 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 three differences in the relative size of returns across subgroups compared to the OLSMG estimates: (1) There are no statistically significant differences by race; (2) CCEMG shows statistically smaller returns for Hispanic and foreign-born individuals; (3) There is a different pattern across highest level of education completed. Both OLSMG and the FM estimates indicate the largest mean return to schooling for individuals who ultimately stop at high school, but show different relative returns for other education levels: FMs show the next largest mean returns for individuals whose highest achievement is a bachelor's or a graduate degree, whereas OLSMG shows the second largest returns for individuals who begin college but do not finish. The FM returns are statistically smaller for some college compared to high school (IFEMG and CCEMG), statistically larger for a bachelor's compared to some college (CCEMG), and statistically smaller for a graduate degree compared to a bachelor's (CCEMG).

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).[32] The OLSMG results, similar to the evidence from Henderson et al. (2011), suggest diminishing marginal returns to years of schooling, at least until graduate school, while

[32]The differences across birth cohorts for CCEMG-2 do not show statistically significant differences when a given birth cohort is only tested against the cohort directly above it in the table, but statistically larger returns for more recent birth cohorts do exist when the 1955-1959 or after-1959 cohorts are tested against the before-1950 cohort.

the FM results are more suggestive of "sheepskin effects" (Layard and Psacharopoulos, 1974; Hungerford and Solon, 1987; Jaeger and Page, 1996); 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.

5.5.3 Within-Group Heterogeneity

The heterogeneous models also allow for the analysis of heterogeneity within subgroups. Table 9 shows the variance of the individual returns within each subgroup. The FM estimates show larger variance than OLSMG for every subgroup except for individuals born before 1950 (CCEMG). The FMs also change the relative variance across subgroups compared to OLSMG in two cases: (1) The FMs generally show larger variance for more recent birth cohorts, whereas OLSMG shows smaller variance;[33] (2) FMs show much larger variance for individuals who only obtain a high school degree than any other education level, whereas OLSMG shows similar variance between these individuals and those with higher levels of education.

Table 10 reports the 25th, 50th, and 75th percentiles of the distribution of individual returns by subgroup. The FM estimators 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 is inconsistent with the larger variance associated with the FMs in Table 9 . 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 tails than OLS, which increases the overall variance.

[33]The one notable exception is that IFEMG shows much larger variance for individuals born before 1950 than any other cohort. But IFEMG still shows increasing variance over time for the subsequent cohorts.

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 same direction serving to explain the gap in the heterogeneous interactive fixed effects estimates and the homogeneous panel OLS estimates. Our subgroup heterogeneity analysis suggests interesting differences among methods both within and across subgroups. For example, OLS or standard nonparametric regressions suggest the presence of diminishing marginal returns to schooling, at least until graduate school. In contrast, our preferred estimates are suggestive of “sheepskin effects” so that degree attainment can have an important impact in determining the value of additional years of schooling.

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.

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.

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

Becker, G. S. (1964). *Human Capital, A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: Columbia University Press.

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

Caplan, B. (2018). *The Case Against Education: Why the Education System is a Waste of Time and Money*. Princeton University Press.

Card, D. (1995). Earnings, schooling, and ability revisited. 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.

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.

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.

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

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

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.

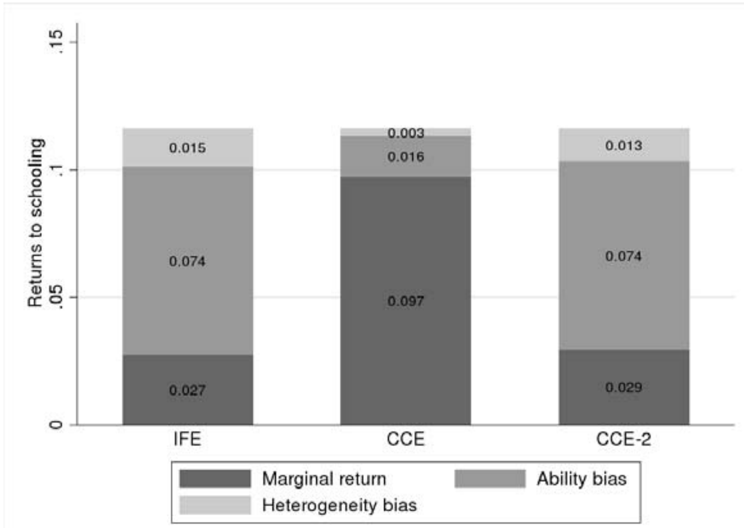
Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. *Econometric Reviews 34*, 1089–1117.

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*.

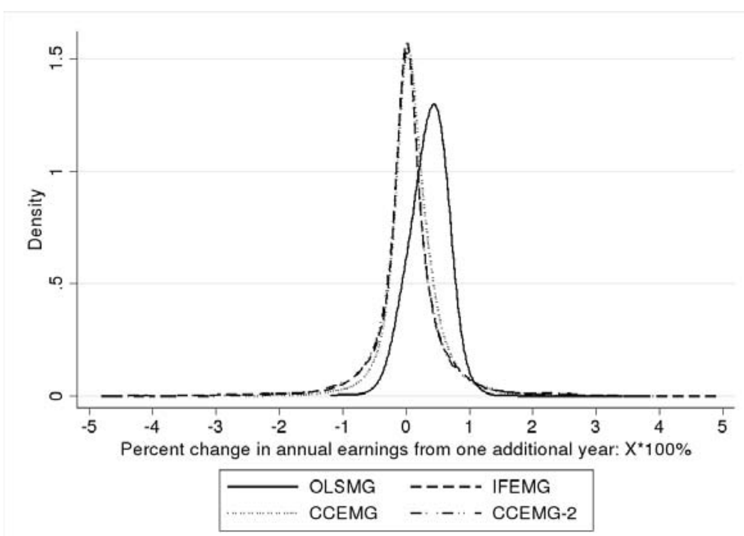
Westerlund, J. and Y. Petrova (2017). Asymptotic collinearity in cce estimation of interactive effects models. *Economic Modeling forthcoming*.

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 total stacked bar for each estimator represents the pooled OLS estimate of the returns to schooling with year fixed effects and age controls, corresponding to column (3) in Panel B of Table 4. We use the decomposition $\hat{\beta}_{POLS} - \hat{\beta}_{IFEMG} = \left(\hat{\beta}_{POLS} - \hat{\beta}_{IFE}\right) + \left(\hat{\beta}_{IFE} - \hat{\beta}_{IFEMG}\right)$, 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 same calculations are applied using CCE and CCE-2 estimates.

Figure 2: Distribution of Marginal Returns to Schooling



Source: SIPP respondents linked to IRS and SSA data in the U.S. Census Bureau Gold Standard File.
Note: Each line is a kernel density plot of individual returns based on the heterogeneous model for the given estimator. Results are based on the specification with age controls, which corresponds to columns (2), (4), (6), and (8) in Table 7.

Table 1: Summary Statistics

	(1) Year 1990 sample	(2) Age 40 sample	(3) Panel sample
Annual earnings	43,920	59,270	50,660
	(35,540)	(61,390)	(76,890)
Years of school	14.60	14.83	14.46
	(2.076)	(2.004)	(2.138)
Age (in quarters)	133.2	160.3	151.2
	(16.65)	(1.715)	(42.63)
Married	0.750	0.805	0.698
	(0.433)	(0.396)	(0.459)
Black	0.052	0.052	0.052
	(0.222)	(0.222)	(0.222)
Other race	0.024	0.024	0.024
	(0.152)	(0.152)	(0.152)
Hispanic	0.039	0.039	0.039
	(0.193)	(0.193)	(0.193)
Foreign born	0.026	0.026	0.026
	(0.159)	(0.159)	(0.159)
Birth year	1957	1957	1957
	(4.160)	(4.160)	(4.160)
Observations	6,300	6,300	213,000

Source: SIPP respondents linked to IRS and SSA data in the U.S. Census Bureau Gold Standard File.
Note: Each column reports averages and the standard deviations in parentheses for the given sample. Columns (1) and (2) report averages at given points in time from the panel sample. The panel sample used for analysis includes males with earnings observations in each year from 1978-2011, variation in level of education during 1978-2011, between the ages of 16-65 during the entirety of 1978-2011, age 27 or older at the time of the SIPP survey, not currently enrolled in school at the time of the SIPP survey, and without any missing data. The sample includes a balanced panel of N=6,300 individuals over T=34 years. Annual earnings are adjusted for inflation to 1999 dollars.

Table 2: Summary of Estimated Specifications

	Specification	Controls	Estimator
1.	$y_i = c + x_i\beta + w_i'\gamma + u_i$	Demographic controls	CSOLS, CS2SLS
2.	$y_i = c + x_i\beta + w_i'\gamma + a_i\rho_1 + a_i^2\rho_2 + u_i$	Demographic controls	CSOLS, CS2SLS
3.	$y_{it} = \delta_t + x_{it}\beta + u_{it}$	time fixed effects	POLS, P2SLS
4.	$y_{it} = c_i + \delta_t + x_{it}\beta + u_{it}$	time and person fixed effects	POLS
5.	$y_{it} = \delta_t + x_{it}\beta + a_{it}\rho_1 + a_{it}^2\rho_2 + u_{it}$	time fixed effects, age controls	POLS, P2SLS
6.	$y_{it} = \delta_t + x_{it}\beta + w_{it}'\gamma + u_{it}$	time fixed effects, Demographic controls	POLS, P2SLS
7.	$y_{it} = \delta_t + x_{it}\beta + w_{it}'\gamma + a_{it}\rho_1 + a_{it}^2\rho_2 + u_{it}$	time fixed effects, Demographic and age controls	POLS, P2SLS
8.	$y_{it} = \delta_t + x_{it}\beta + \lambda_i' f_t + u_{it}$	time and interactive fixed effects	IFE, CCEP, CCEP-2
9.	$y_{it} = c_i + \delta_t + x_{it}\beta + \lambda_i' f_t + u_{it}$	time, person and interactive fixed effects	IFE, CCEP, CCEP-2
10.	$y_{it} = \delta_t + x_{it}\beta + a_{it}\rho_1 + a_{it}^2\rho_2 + \lambda_i' f_t + u_{it}$	time and interactive fixed effects, age controls	IFE, CCEP, CCEP-2
11.	$y_{it} = c_i + x_{it}\beta_i + u_{it}$	person fixed effects	OLSMG
12.	$y_{it} = x_{it}\beta_i + a_{it}\rho_{1i} + a_{it}^2\rho_{2i} + u_{it}$	age controls	OLSMG
13.	$y_{it} = c_i + x_{it}\beta_i + \lambda_i' f_t + u_{it}$	person and interactive fixed effects	IFEMG, CCEMG, CCEMG-2
14.	$y_{it} = x_{it}\beta_i + a_{it}\rho_{1i} + a_{it}^2\rho_{2i} + \lambda_i' f_t + u_{it}$	interactive fixed effects and 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 OLS and 2SLS Estimates of the Return to Schooling for Males

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	At year 1990						At age 40		
	OLS	OLS	2SLS	2SLS	2SLS	2SLS	OLS	2SLS	2SLS
Years of school	0.090***	0.090***	0.286***	0.231*	0.139***	0.137**	0.116***	0.312***	0.194***
	(0.005)	(0.005)	(0.102)	(0.126)	(0.043)	(0.045)	(0.006)	(0.088)	(0.040)
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age & age-squared	No	Yes	No	Yes	No	Yes	No	No	No
Instrument			Quarter	Quarter	Quarter x year	Quarter x year		Quarter	Quarter x year
First stage F-stat			12.4	12.3	8.2	8.1		11.6	7.6
Observations	6,300	6,300	6,300	6,300	6,300	6,300	6,300	6,300	6,300

Source: SIPP respondents linked to IRS and SSA data in the U.S. Census Bureau Gold Standard File.
Note: The dependent variable is the log of annual W-2 earnings or self-employment earnings. Columns (1)-(6) are based on a cross-section in 1990. Columns (7)-(9) are based on a cross-section at age 40. Demographic controls include race, Hispanic status, foreign born status, marital status, state of residence during the SIPP survey, and birth year. Years of school is instrumented for with quarter of birth indicator variables in columns (3), (4), and (8) and with quarter of birth indicator variables interacted with year of birth indicator variables in columns (5), (6), and (9). Standard errors are clustered at the state level and shown in parentheses. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*.

Table 4: Panel OLS and 2SLS Estimates of the Return to Schooling for Males

	(1) OLS	(2) OLS	(3) OLS	(4) 2SLS	(5) 2SLS	(6) 2SLS	(7) 2SLS
Years of school	0.130***	0.142***	0.116***	0.235**	0.215**	0.237***	0.155***
	(0.004)	(0.005)	(0.004)	(0.105)	(0.102)	(0.008)	(0.016)
Age & age-squared	No	No	Yes	No	Yes	No	Yes
Person FE	No	Yes	No	No	No	No	No
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Instrument				Quarter	Quarter	Quarter x year	Quarter x year
First stage F-stat				854.9	1,009	376.7	410.7
CD test stat	114.1	124.0	130.8	98.8	109.1	98.4	127.8
Observations	213,000	213,000	213,000	213,000	213,000	213,000	213,000

Source: SIPP respondents linked to IRS and SSA data in the U.S. Census Bureau Gold Standard File.
Note: Table 4 replicates the results from Table 3, except with panel data on earnings and education from 1978-2011 rather than point-in-time cross-section data.

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

	(1) IFE	(2) IFE	(3) IFE	(4) CCEP	(5) CCEP	(6) CCEP	(7) CCEP-2	(8) CCEP-2	(9) CCEP-2
Years of school	0.033***	0.033***	0.042***	0.087***	0.076***	0.100***	0.041***	0.041***	0.042***
	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.006)	(0.005)	(0.005)
Age & age-squared	No	No	Yes	No	No	Yes	No	No	Yes
Person FE	No	Yes	No	No	Yes	No	No	Yes	No
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	213,000	213,000	213,000	213,000	213,000	213,000	213,000	213,000	213,000

Source: SIPP respondents linked to IRS and SSA data in the U.S. Census Bureau Gold Standard File.
Note: Columns (1)-(3) are based on Interactive Fixed Effects (IFE) estimates (Bai, 2009) with 9, 8 and 7 factors, respectively, selected by the IC_{p1} procedure in Bai and Ng (2002). Columns (4)-(6) are based on Common Correlated Effects pooled (CCEP) estimates (Pesaran, 2006). Columns (7)-(9) are based on the two-step CCE procedure with 8, 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 6: Bias Associated with OLS and 2SLS 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: Year fixed effects						
OLS	0.081	0.008	0.004	0.002	0.002	0.097
2SLS	0.100	0.107	-0.002	-0.0002	-0.001	0.204
B. Covariates: Person and year fixed effects [1]						
OLS	0.062	0.020	0.014	0.006	0.008	0.110
C. Covariates: Year fixed effects and age controls						
OLS	0.073	-0.008	0.005	0.002	0.002	0.074
2SLS	0.083	0.025	0.003	0.00003	0.001	0.113
CCEP						
A. Covariates: Year fixed effects						
OLS	0.040	0.001	0.002	0.001	0.0001	0.044
2SLS	0.050	0.102	-0.003	0.0001	0.0001	0.150
B. Covariates: Person and year fixed effects						
OLS	0.045	0.013	0.008	0.003	0.002	0.071
C. Covariates: Year fixed effects and age controls						
OLS	0.035	-0.021	0.002	0.001	0.0004	0.017
2SLS	0.216	-0.162	0.0003	0.00003	-0.0004	0.054
CCEP-2						
A. Covariates: Year fixed effects						
OLS	0.074	0.007	0.004	0.002	0.001	0.089
2SLS	0.088	0.110	-0.003	0.0001	0.0001	0.196
B. Covariates: Person and year fixed effects						
OLS	0.058	0.020	0.013	0.005	0.006	0102
C. Covariates: Year fixed effects and age controls						
OLS	0.081	-0.015	0.005	0.002	0.002	0.075
2SLS	0.267	-0.152	0.0001	-0.00005	-0.0003	0.113

Source: SIPP respondents linked to IRS and SSA data in the U.S. Census Bureau Gold Standard File.

Note: 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 4 Panel B that include the covariates listed in the panel title. The common factors in the IFE panels are based on the IFE results, and in the CCE panels are based on the principal components procedure applied to residuals based on the CCEP estiamtes in Table 5 that correspond to the specifications in the panel titles. Column (5) includes common factors up to 9 in the IFE panel (8 in the CCE panels) for the specifications with only year effects, 8 in the IFE panel (7 in the CCE panels) for the specifications with person and year effects and 7 in the IFE panel (8 in the CCE panels) for the specifications with age controls.

[1] The discrepancy of the total bias (0.110 versus 0.109 which is the difference between OLS estimate in Table 4 Panel B column (2) and IFE estimate in Table 5 column (2)) is due to rounding.

Table 7: Common Factor Model Estimates of the Return to Schooling for Males - Heterogeneous Model

	(1) OLSMG	(2) OLSMG	(3) IFEMG	(4) IFEMG	(5) CCEMG	(6) CCEMG	(7) CCEMG-2	(8) CCEMG-2
Years of school	0.444***	0.353***	0.025**	0.027***	0.105***	0.097***	0.028***	0.029***
	(0.005)	(0.004)	(0.011)	(0.008)	(0.005)	(0.005)	(0.007)	(0.007)
Age & age-squared	No	Yes	No	Yes	No	Yes	No	Yes
Su-Chen slope test			21.9	36.9	24.6	20.7	24.3	19.6
Ando-Bai slope test			741.0	64,760.0	334.9	778.8	631.5	782.0
Observations	213,000	213,000	213,000	213,000	213,000	213,000	213,000	213,000

Source: SIPP respondents linked to IRS and SSA data in the U.S. Census Bureau Gold Standard File.
Note: Each estimator is replaced with a version that allows regression coefficients to vary across individuals (Pesaran and Smith, 1996; Pesaran, 2006; Song, 2013). The individual-level regression coefficients are then averaged across all individuals to produce a "mean group" (MG) estimate. IFEMG estimates are based on 9 and 7 factors for columns (3) and (4), respectively, selected by the IC_{p1} procedure in Bai and Ng (2002). Two-step CCE estiamtes are based on 7 and 8 factors for columns (7) and (8), respectively, 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 8: Bias Associated with Heterogeneous Model OLS 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
IFEMG						
A.Covariates: Person fixed effect						
OLSMG	0.389	0.019	0.003	0.003	0.005	0.419
B. Covariates: Age controls						
OLSMG	0.283	0.019	0.018	0.004	0.003	0.326
CCEMG						
A. Covariates: Person fixed effect						
OLSMG	0.304	0.019	0.017	0.005	0.001	0.345
B. Covariates: Age controls						
OLSMG	0.188	0.043	0.027	0.007	0.001	0.265
CCEMG-2						
A. Covariates: Person fixed effect						
OLSMG	0.343	0.026	0.030	0.008	0.009	0.416
B. Covariates: Age controls						
OLSMG	0.201	0.064	0.025	0.015	0.018	0.324

Source: SIPP respondents linked to IRS and SSA data in the U.S. Census Bureau Gold Standard File.
Note: Bias estimates are based on the common factor estimates from the heterogeneous factor model results in Table 7. The CCE common component estimates are based on applying principal components to $\eta_{it} = y_{it} - x_{it}\hat{\beta}_{CCE}$. Column (5) includes common factors up to 9 in the IFE panel (7 in the CCE panels) for the specifications with person fixed effects and 7 in the IFE panel (8 in the CCE panels) for the specifications with age controls.

Table 9: 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	0.350	0.095	0.028	0.386	0.098	0.170	0.030	0.313	5,800
Black	**0.395**	0.100	0.024	0.397	0.089	0.149	0.035	0.320	350
Other race	0.364	0.092	-0.008	0.276	0.066	0.115	0.010	0.256	150
B. Hispanic status									
Non-Hispanic	0.353	0.095	0.028	0.387	0.099	0.167	0.031	0.309	6,000
Hispanic	0.345	0.098	0.002	0.312	**0.051**	0.178	-0.015	0.389	250
C. Foreign born status									
Native	0.354	0.095	0.026	0.385	0.098	0.168	0.030	0.314	6,100
Foreign Born	0.327	0.092	0.054	0.353	**0.041**	0.159	0.021	0.228	150
D. Birth cohort									
Born before 1950	0.160	0.136	-0.033	0.640	0.005	0.131	-0.010	0.207	500
Born 1950-1954	**0.299**	0.102	0.003	0.313	**0.044**	0.158	0.020	0.263	1,200
Born 1955-1959	**0.374**	0.084	0.013	0.341	**0.089**	0.143	0.026	0.284	2,600
Born after 1959	**0.405**	0.080	**0.073**	0.413	**0.160**	0.206	0.049	0.401	2,000
E. Highest education level completed									
High school	0.538	0.068	0.120	0.762	0.202	0.357	0.049	0.678	850
Some college	**0.458**	0.093	**0.001**	0.418	**0.042**	0.164	0.013	0.351	2,500
Bachelor's	**0.191**	0.057	0.023	0.216	**0.129**	0.076	0.034	0.139	1,700
Graduate	**0.240**	0.077	0.022	0.285	**0.091**	0.160	0.042	0.225	1,200

Source: SIPP respondents linked to IRS and SSA data in the U.S. Census Bureau Gold Standard File.
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 7. Estimates are based on the specifications from Table 7 that include age and age-squared. Bold numbers indicate means that are statistically different than within-panel counterparts at at least the ten-percent level. 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 10: 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	0.160	0.378	0.563	-0.126	0.021	0.208	-0.079	0.056	0.263	-0.138	0.020	0.202
A. Race												
White	0.158	0.375	0.561	-0.127	0.022	0.210	-0.079	0.056	0.260	-0.139	0.020	0.203
Black	0.213	0.426	0.616	-0.113	0.011	0.196	-0.082	0.061	0.272	-0.140	0.015	0.199
Other race	0.180	0.394	0.549	-0.097	0.016	0.152	-0.071	0.052	0.248	-0.112	0.014	0.167
B. Hispanic status												
Non-Hispanic	0.161	0.378	0.563	-0.126	0.022	0.208	-0.078	0.057	0.264	-0.136	0.020	0.204
Hispanic	0.122	0.371	0.575	-0.143	0.018	0.191	-0.105	0.046	0.218	-0.226	-0.002	0.173
C. Foreign born status												
Native	0.161	0.378	0.564	-0.126	0.022	0.208	-0.079	0.057	0.265	-0.139	0.020	0.202
Foreign Born	0.104	0.363	0.545	-0.120	0.008	0.145	-0.083	0.020	0.189	-0.114	0.011	0.181
D. Birth cohort												
Born before 1950	-0.032	0.158	0.371	-0.084	0.007	0.112	-0.098	0.008	0.126	-0.121	0.001	0.126
Born 1950-1954	0.099	0.337	0.515	-0.118	0.004	0.134	-0.098	0.022	0.148	-0.112	0.009	0.145
Born 1955-1959	0.195	0.391	0.568	-0.131	0.018	0.187	-0.079	0.052	0.245	-0.143	0.018	0.189
Born after 1959	0.215	0.435	0.607	-0.140	0.052	0.303	-0.058	0.130	0.361	-0.160	0.040	0.275
E. Highest education level completed												
High school	0.424	0.578	0.694	-0.198	0.082	0.481	-0.105	0.157	0.489	-0.282	0.032	0.397
Some college	0.311	0.489	0.636	-0.142	0.014	0.186	-0.125	0.032	0.210	-0.159	0.011	0.194
Bachelor's	0.025	0.209	0.359	-0.107	0.019	0.185	-0.032	0.089	0.276	-0.105	0.018	0.185
Graduate	0.058	0.260	0.435	-0.096	0.019	0.157	-0.062	0.048	0.209	-0.091	0.029	0.175

Source: SIPP respondents linked to IRS and SSA data in the U.S. Census Bureau Gold Standard File.
Note: Quantiles reported in the table are characteristic-specific summary statistics of the individual-level coefficients from the heterogeneous model results in Table 7. Estimates are based on the specifications from Table 7 that include age and age-squared. p25 = 25th percentile, p50 = 50th percentile, p75 = 75th percentile. Sample sizes for each group are shown in Table 9.

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.[34] The model is given by

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

where y_{it} (x_{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 x_{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

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

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

where

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

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

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 $Cov(z_{it,k},x_{it}) \neq 0$ 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 $\hat{\Pi} = \left(\sum_{i=1}^N Z_i'Z_i\right)^{-1}\sum_{i=1}^N Z_i'X_i$. 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 $\hat{X}_i = Z_i\hat{\Pi}$. Then we have

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

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:

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

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 $SVar(x_{it})$, $SVar(\hat{X}_i)$, $SCov(\hat{X}_i, \hat{F}_j\hat{\lambda}_{ji})$, $SCov(x_{it}, \hat{c}_{j,it})$ 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:

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

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

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

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

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 $SCov(x_{it}, \hat{c}_{j,it})$ is the average (over time) of the cross-sectional correlation between x_{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 $\eta_{it} = y_{it} - x_{it}\hat{\beta}_{CCE}$, 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

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

with Y_i, X_i, U_i being $(T \times 1)$ vectors defined as $Y_i = (y_{i1}, \ldots, y_{iT})'$, $X_i = (x_{i1}, \ldots, x_{iT})'$, $U_i = (u_{i1}, \ldots, u_{iT})'$ and $F = (f_1, \ldots, f_T)'$ being the $(T \times r)$ matrix of common factors. Here we interpret y_{it} (x_{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

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

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

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

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

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

where $\omega_i = \left(\sum_i X_i' M_{\hat{F}} X_i\right)^{-1} X_i' M_{\hat{F}} X_i$ 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

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

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

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

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 $\hat{\beta}_{CCEMG}$.

B Accounting for Experience

Consider the pooled specification

$$y_{it} = \delta_t + x_{it}\beta + e_{it}\rho_1 + e_{it}^2\rho_2 + \lambda_i' f_t + u_{it} \tag{A.15}$$

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

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

or,

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

or,

$$y_{it} = \widetilde{\rho}_{1i} + \widetilde{\rho}_{2i}t + \widetilde{\delta}_t + x_{it}\beta + \lambda_i' f_t + u_{it} \tag{A.16}$$

where

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

Thus, from $(A.16)$ in the pooled model, besides time fixed effect, we should include a person fixed effect and person-specific linear trend, which is equivalent to a pooled model that includes age and age-squared terms instead of the person fixed effect and person-specific linear trend.

In the heterogeneous model,

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

or

$$y_{it} = \breve{\rho}_{1i} + \breve{\rho}_{2i}t + \rho_{2i}t^2 + x_{it}\beta_i + \lambda_i{}' f_t + u_{it} \tag{A.17}$$

where

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

From $(A.17)$, we should include a person fixed effect, person-specific quadratic trend, which is equivalent to a heterogeneous specification that includes age and age-squared terms instead of the person fixed effect and person-specific quadratic trend.

C Robustness Checks

C.1 Comparative Sample - without earnings-in-school restriction

Table C1 below shows summary statistics for a comparison sample with the same individuals, except also including individuals who have missing earnings data while enrolled in school. As discussed earlier in the paper, using a panel approach necessitates analysis of individuals who have earnings data while increasing their education, which could introduce a selected sample of individuals who are different with respect to observable or unobservable characteristics. Comparing Table 1 and Table C1, one can see that 62.38 percent of individuals in the comparison sample remain in the sample of analysis. This is generally consistent with evidence that as much as 92 percent of individuals gain at least some work experience during high school and 88 percent during college (Hotz et al., 2002). Employment during college, in particular, has been on the rise and is one of the main variables related to increased time-to-graduation (Bound et al., 2012). The frequent occurrence of work while enrolled in school, both in our data and in the literature, helps eliminate concerns that the sample of analysis is a small, non-representative group of individuals.

Table C1: Summary Statistics - Comparative Sample

	(1) Year 1990 sample	(2) Age 40 sample	(3) Panel sample
Annual earnings	42,570	62,300	50,410
	(35,730)	(71,820)	(81,400)
Years of school	14.75	15.01	14.56
	(2.164)	(2.093)	(2.249)
Age (in quarters)	130.8	160.3	148.8
	(16.51)	(1.716)	(42.58)
Married	0.720	0.792	0.669
	(0.449)	(0.406)	(0.471)
Black	0.057	0.057	0.057
	(0.232)	(0.232)	(0.232)
Other race	0.039	0.039	0.039
	(0.194)	(0.194)	(0.194)
Hispanic	0.041	0.041	0.041
	(0.198)	(0.198)	(0.198)
Foreign born	0.052	0.052	0.052
	(0.222)	(0.222)	(0.222)
Birth year	1957	1957	1957
	(4.117)	(4.117)	(4.117)
Observations	10,100	10,100	342,000

Source: SIPP respondents linked to IRS and SSA data in the U.S. Census Bureau Gold Standard File.
Note: The comparative samples shown above are the same as those in Table 1, except they also include individuals without earnings data while enrolled in school.

C.2 Robustness Check of Table 3

Table C2 below presents cross-section estimates based on the comparative sample from Table C1. The OLS and 2SLS estimates are very similar to those in Table 3 when individuals who have missing earnings data during school are included. This helps alleviate concerns that our panel results are driven largely by sample selection effects.

Table C2: Cross-Section OLS and 2SLS Estimates of the Return to Schooling for Males - Comparative Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	At year 1990						At age 40		
	OLS	OLS	2SLS	2SLS	2SLS	2SLS	OLS	2SLS	2SLS
Years of school	0.097***	0.097***	0.297***	0.234	0.169***	0.157**	0.136***	0.277***	0.175***
	(0.005)	(0.005)	(0.110)	(0.155)	(0.043)	(0.037)	(0.004)	(0.082)	(0.040)
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age & age-squared	No	Yes	No	Yes	No	Yes	No	No	No
Instrument			Quarter	Quarter	Quarter x year	Quarter x year		Quarter	Quarter x year
First stage F-stat			15.9	15.6	10.09	10.0		15.5	9.8
Observations	9,900	9,900	9,900	9,900	9,900	9,900	10,000	10,000	10,000

Source: SIPP respondents linked to IRS and SSA data in the U.S. Census Bureau Gold Standard File.

Note: The analysis in Table C2 is the same as in Table 3, except based on the comparative sample shown in Table C1. The number of observations is slightly less than the number reported in Table C1 because some individuals in the comparative sample had not yet finished schooling and had missing earnings data at the point of the cross-section time period.

C.3 Robustness Check of Table 4

The Table below presents a robustness check of Table 4 when demographic controls are included. The general patterns are very similar to those in Table 4. In addition, the demographic controls have little impact on the estimated schooling coefficient (OLS and 2SLS), once the age controls are included.

Table C3: Panel OLS and 2SLS Estimates of the Return to Schooling for Males with Demographic Controls

	(1) OLS	(2) OLS	(3) 2SLS	(4) 2SLS	(5) 2SLS	(6) 2SLS
Years of school	0.108***	0.107***	0.262***	0.213**	0.185***	0.167***
	(0.004)	(0.004)	(0.082)	(0.096)	(0.030)	(0.033)
Age & age-squared	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Instrument			Quarter	Quarter	Quarter x year	Quarter x year
First stage F-stat			353.9	371.5	246.6	259.8
CD test stat	121.5	124.2	85.7	102.9	107.9	116.1
Observations	213,000	213,000	213,000	213,000	213,000	213,000

Source: SIPP respondents linked to IRS and SSA data in the U.S. Census Bureau Gold Standard File.
Note: The analysis in Table C3 is the same as in Table 4, but with demographic controls. A specification with person fixed effects is not available when demographic controls are included.

C.4 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:

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

The estimates, reported in columns (1)-(3) in Table C4 below, are only marginally smaller than those reported in columns (1)-(3) in Table 4, with a reduction of about 10 percent for our preferred specification with age controls. These findings lend support to our interpretation of the factor loadings as skills and that interactive fixed effects are needed to fully model these skills.

Table C4: Time-Varying Returns to Demographics as Proxies for Interactive Fixed Effects

	(1)	(2)	(3)
	OLS	OLS	OLS
Years of School	0.105***	0.116***	0.104***
	(0.004)	(0.004)	(0.004)
Age & age-squared	No	No	Yes
Person FE	No	Yes	No
Year FE	Yes	Yes	Yes
Demographics-by-Year FE	Yes	Yes	Yes
CD test stat	112.1	110.8	112.2
Observations	213,000	213,000	213,000

Source: SIPP respondents linked to IRS and SSA data in the U.S. Census Bureau Gold Standard File.
Note: Columns (1)-(3) are identical to columns (1)-(3) in Table 4, 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 $y_{it} = \delta_t + x_{it}\beta + w_{it}'\gamma + \lambda_i' f_t + u_{it}$, here we estimate $y_{it} = \delta_t + x_{it}\beta + w_{it}'\gamma + d_i'\theta_t + v_{it}$. 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.

D Extreme Returns

Given that the results discussed in Section 5.5 show significant heterogeneity in returns both across and within subgroups, Table D1 below shows characteristics that are associated with being in the top 5 percent and bottom 5 percent of returns for each estimator. In addition to the characteristics discussed above, this table also includes whether the individual finished high school late (at age 20 or later), began college late (three or more years after finishing high school), and finished college late (did not obtain their bachelor's until age 26 or later).

Some aspects of the OLSMG results are intuitive. For example, individuals born in more recent birth cohorts being statistically more likely to have a top 5 percent return and statistically less likely to have a bottom 5 percent return is consistent with the literature that returns to schooling have increased over time. Other results are less intuitive, such as individuals who hold a bachelor's or graduate degree being statistically less likely to have a top 5 percent return and statistically more likely to have a bottom 5 percent return, compared to the opposite pattern for individuals with only some college.

The FM results seem to correct some of the less intuitive results from OLSMG. They show that individuals with a bachelor's or graduate degree are statistically less likely to end up in the bottom 5 percent of returns while those with only some college are statistically more likely to end up in the bottom 5 percent. This is the opposite of the OLSMG result and is consistent with the potential "sheepskin effects" discussed previously. The result that individuals born after 1959 are statistically more likely to end up in both the top and bottom 5 percent according to the FM estimates, despite having larger mean returns in Table 9, is consistent with the larger variance in returns for more recent birth cohorts.

Finally, all three FM estimators find evidence that individuals who begin college late or finish college late are more likely to be in both the top and bottom 5 percent of returns. In addition to suggesting large within-group heterogeneity for these individuals, this result is also consistent with multiple potential selection biases that have different predictions for the return to schooling: (1) Some individuals who begin and finish college late may do so because of poor grades or lack of motivation, which could also be related to lower earnings after college; (2) Other individuals who begin and finish college late may do so because their realized outcomes in the labor market without a college degree suggested that they had the most to gain from continuing their education; (3) Still more individuals who begin and finish college late may have delayed because of large financial or psychic costs associated with attending college, in which case those who eventually went to and completed college must have had large potential returns in order to take on the large costs (Becker, 1964; Heckman, Lochner, and Todd, 2006).

Table D1: Percentage Difference in Characteristics Associated with Extreme Returns

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Top 5 percent of returns				Bottom 5 percent of returns			
	OLSMG	IFEMG	CCEMG	CCEMG-2	OLSMG	IFEMG	CCEMG	CCEMG-2
Married at 40	**-0.077**	-0.033	-0.023	-0.002	-0.006	**-0.043**	**-0.049**	**-0.060**
Finished HS late	0.026	-0.008	-0.018	**0.032**	**0.093**	-0.008	**0.042**	-0.018
Began college late	**0.306**	**0.132**	**0.127**	**0.096**	**-0.061**	0.035	**0.085**	**0.094**
Finished college late	**0.283**	**0.069**	0.041	**0.066**	**0.081**	**0.049**	**0.126**	**0.105**
Black	**0.030**	-0.001	-0.021	-0.011	0.003	0.006	0.012	-0.001
Other race	-0.001	-0.008	**-0.015**	-0.008	-0.005	0.005	-0.005	0.005
Hispanic	0.003	-0.004	**-0.020**	-0.004	-0.001	**0.026**	0.016	0.017
Foreign born	-0.014	-0.012	**-0.017**	-0.004	-0.001	-0.011	-0.004	-0.004
Born 1950-1954	**-0.067**	-0.034	-0.014	**-0.038**	**0.151**	-0.028	0.033	-0.027
Born 1955-1959	0.025	**-0.054**	**-0.073**	-0.023	**-0.130**	-0.023	**-0.067**	**-0.056**
Born after 1959	**0.093**	**0.129**	**0.124**	**0.099**	**-0.149**	**0.062**	0.008	**0.090**
Some college	**0.256**	-0.034	**-0.070**	0.010	**-0.144**	**0.033**	**0.147**	**0.055**
Bachelor's	**-0.271**	**-0.123**	**-0.160**	**-0.143**	**0.143**	**-0.096**	**-0.217**	**-0.156**
Grad degree	**-0.151**	**-0.060**	0.001	**-0.030**	**0.095**	**-0.054**	-0.020	**-0.063**

Source: SIPP respondents linked to IRS and SSA data in the U.S. Census Bureau Gold Standard File.
Notes: Each entry shows the difference in the mean value for the row characteristic between individuals in the top (bottom) 5 percent and the bottom (top) 95 percent of individual-level returns for the estimator listed in the column heading. Specifically, the average of each characteristic for individuals in the bottom (top) 95 percent is subtracted from the average of each characteristic for individuals in the top (bottom) 5 percent and a t-test for statistical significance is performed. Bold numbers indicate statistical significance at at least the ten-percent level. The results for each estimator correspond to the specifications from Table 7 with age controls included.