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

Online Supplemental Appendices

Online Supplemental Appendices

A Heterogeneity Bias Derivations

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

$$Y_i = S_i\beta_i + F\lambda_i + U_i \tag{A.1}$$

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

The IFE estimator is given by

$$\hat{\beta}_{IFE} = \left(\sum_{i=1}^{N} S_i' M_{\hat{F}} S_i\right)^{-1} \left(\sum_{i=1}^{N} S_i' M_{\hat{F}} Y_i\right) \tag{A.2}$$

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 $(A.1)$, we can write $(A.2)$ as

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

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

$$\hat{\beta}_{IFE} \underset{N,T\ large}{\simeq} \sum_i \omega_i\beta_i \tag{A.3}$$

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

B Accounting for Experience

Consider the pooled specification

$$y_{it} = c_i + s_{it}\beta + e_{it}\rho_1 + e_{it}^2\rho_2 + \lambda_i{}' f_t + u_{it} \tag{A.4}$$

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

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

or,

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

or,

$$y_{it} = \widetilde{\rho}_{1i} + \widetilde{\rho}_{2i}t + \widetilde{\delta}_t + s_{it}\beta + \lambda_i{}' f_t + u_{it} \tag{A.5}$$

where

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

Thus, from $(A.5)$ in the pooled model, besides time fixed effect, we should include person fixed effects and person-specific linear trend, which is equivalent to a pooled model that includes person fixed effects, age and age-squared terms instead of the person-specific linear trend.

In the heterogeneous model,

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

or

$$y_{it} = \breve{\rho}_{1i} + \breve{\rho}_{2i}t + \rho_{2i}t^2 + s_{it}\beta_i + \lambda_i{}' f_t + u_{it} \tag{A.6}$$

where

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

From $(A.6)$, we should include person fixed effects and person-specific quadratic trend, which is equivalent to a heterogeneous specification that includes person fixed effects, age and age-squared terms instead of the person-specific quadratic trend.

C Data: Schooling Variable Construction

We construct a longitudinal years of schooling variable based on the SIPP education information that includes highest education level completed (‘no high school degree’, ‘high school degree’, ‘some college’, ‘college degree’, and ‘graduate degree’), the year during which high school was completed, the year during which post-high school education began, the year during which post-high school education ended, and the year during which a bachelor’s degree was earned. First, individuals were assigned one of the five highest-level-completed values for each year.[1] All individuals were assigned ‘no high school degree’ before the year they graduated high school and ‘high school degree’ beginning in their graduation year. Individuals whose highest completed level was ‘some college’ and thus did not obtain a bachelor’s degree were assigned ‘some college’ beginning in the year their post-high school education ended. Individuals who obtained at least a college degree were assigned ‘college degree’ beginning in the year they obtained their bachelor’s degree. Individuals who obtained a graduate degree were assigned ‘graduate degree’ beginning in the year their post-high school education ended.[2] Then, based on highest level completed at each year, individuals

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

[2] 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

were assigned a years of schooling value. Individuals with 'no high school degree' in a given year were assigned 10 years of school, individuals with 'high school degree' were assigned 12 years, individuals with 'some college' were assigned 14 years, individuals with 'college degree' were assigned 16 years, and individuals with 'graduate degree' were assigned 18 years.[3]

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 into the variable. However, we did want to attempt to smooth the discrete jumps described above for two reasons. First, the scheme introduces measurement error by explicitly missing some variation in years of school. For example, it misses the transition through high school by only assigning 10 years for any year before high school degree completion. It also misses the distinction between individuals working with a high school degree with versus without college experience, because the years of schooling variable does not increase until the individual either finishes their post-high school schooling or obtains a bachelor's degree. Second, because we have to limit the main sample to individuals with at least one change in schooling (and further limit to individuals with at least two changes in schooling in Appendix D.1), this allows us to retain more individuals. We therefore make the following two adjustments to smooth the years of schooling variable: (1) we change years of schooling from 10 to 11 the year before a high school degree was finished, which captures progression from 10th grade through 12th[4]; and (2) we change years of schooling from 12 to 13 beginning the year when an individual begins their post-high school education, which captures the distinction between an individual working with a high school degree with versus without college experience.[5] Our main sample of analysis in Panel A column (5) of Table 1 has the following distribution of within-person changes in years of schooling: 250 changes from 10 years to 11; 500 changes from 11 to 12; 900 changes from 12 to 13; 1,500 changes from 13 to 14; 1,100

or obtained a shorter degree. If a person obtained a bachelor's and then stopped, then it is the same as the bachelor's year variable. If the person obtained a graduate degree, then it indicates when they finished graduate school.

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

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

[5]We also conducted our analysis using the non-smoothed version of the schooling variable and the results obtained were qualitatively similar to those reported in the paper based on the smoothed version.

changes from 13 to 16; and 900 changes from 16 to 18.

D Robustness Checks

D.1 Robustness to Alternative Specifications

In this section we discuss robustness of the main results to alternative specifications. Our main results are based on a linear years of schooling and quadratic age specification. This specification is the traditional model originating from Mincer (1974). However, numerous studies have indicated that this specification may not be flexible enough and higher-order terms in schooling and/or experience may be needed (Murphy and Welch, 1990; Heckman, Lochner, and Todd, 2006; Cho and Phillips, 2018). These papers provide evidence supporting the use of up to a quadratic term in years of schooling and a quartic term in experience.[6]

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

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

Table D1: OLS and 2SLS Specification Robustness

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

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

Table D2: Common Factor Pooled Model Robustness

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

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

Table D3: Common Factor Heterogeneous Model Robustness

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

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

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

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

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

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

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

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

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

References

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

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.

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

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

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