Human Capital

UC3M, Labor Economics
Jan Stuhler

January 28, 2020

Human capital: Introduction

Human capital

- Set of skills and abilities that contribute to productivity
- Inherent (e.g. genes and family environment) or acquired (e.g. schooling) and multi-dimensional
- May generate *private* and *social* returns (externalities)

Human capital investment:

- Activity which increases the stock of human capital

Human capital: Introduction

Alternative views of human capital:

Becker (1964)

- Human capital investment improves productivity

Nelson and Phelps (1966)

- Ability to adapt to changing environments and technological innovations

Bowles-Gintis (1976)

- "Soft skills" to adapt to organizations and capitalist society

Spence (1974)

- Education may "signal" inherent ability

Human capital: Introduction

Human capital has important individual and aggregate implications:

Micro-level implications:

- Education explains large part of individual earnings variation
- Individuals explicitly choose their level of education

Macro-level implications:

- Education is a key determinant of economic growth and inequality
- Expenditures for educational system important policy choice

General and specific human capital

Human capital theory formalized by Gary Becker and others (in 1960s).

- Education and training raises workers' productivity
- Individuals invest in human capital just as firms invest in physical capital

We focus on two aspects:

1. General vs. specific human capital
2. Optimal investment into schooling and other forms of human capital over the life cycle

General and specific human capital

Workers invest into formal schooling but also participate in on-the-job training in a diverse set of skills.

Becker distinguished two types of HC:

- **General HC**: general skills effective at all or most firms
- **Specific HC**: effective only at one firm

Simple model

- Period 1: worker produces y_1 and can be trained at cost H
- Period 2: worker produces $y_2(H)$, with $y_2' > 0$ and $y_2'' < 0$

Zero-profit condition for firm (without discounting)

$$w_1 + w_2 + H = y_1 + y_2(H) \tag{1}$$

General human capital

Assume training is **general**

- Trained worker produces $y_2(H)$ at *all* firms. Assuming competitive markets (MP=w) they offer wage $w_2 = y_2(H)$
- Firm that provided training must follow suit to retain the worker. Equation (1) simplifies to

$$w_1 = y_1 - H$$

Implications:

- Worker must bear costs of general training, directly or indirectly (e.g. apprenticeship at reduced wage)
- Rationalizes why educational system are privately or publicly funded

Key difference to physical capital: HC attached to worker and may move to another firm

Specific human capital

Assume training is **specific**

- Trained worker produces $y_2(H)$ at firm that provided training and $y_2(0)$ at other firms
- To retain worker, firm has to pay at least $w_2 = y_2(0) < y_2(H)$ and would make zero profits if

$$w_1 = y_1 - H + (y_2(H) - y_2(0))$$

Implications:

- Firm might bear costs of specific training
- Post-training wage < marginal productivity (but above marginal productivity at other firms) $\rightarrow$ neither worker nor firm wants to terminate employment contract
- Can explain "last hired, first fired" rule during economic downturns (older workers have more specific job training)
- Fired workers experience wage loss ($\rightarrow$ "scarring" literature)

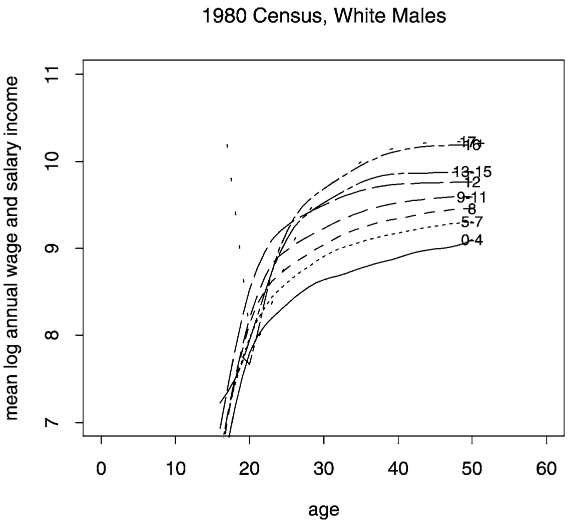
Human capital over the life cycle

Earnings and wage profiles follow a particular pattern over the life cycle

- Schooling and no or only low earnings in early life
- Earnings increase throughout career, but are flat or declining before retirement

These empirical observations can be rationalized with a life-cycle model of human capital accumulation

Figure: Age-earnings profile



Source: Heckman, Lochner and Todd (2006)

Human capital over the life cycle

These pattern can be rationalized with a life-cycle model of human capital accumulation, such as Ben-Porath (1967).

Assumptions (see Section 2.2.1 in CCZ):

- Human capital is one-dimensional object
- Individuals live in continuous time from $t = 0$ until T
- Spend fraction of time $\sigma(t)$ on HC accumulation and $1 - \sigma(t)$ on work
- Law of motion for HC

$$\dot{h}(t) = \theta \sigma(t) h(t) \tag{2}$$

  where θ is the efficiency in HC accumulation ("learning speed")
- Wage = productivity = $Ah(t)(1 - \sigma(t))$

Human capital over the life cycle

Assume individuals maximize total wage gain over the life cycle (i.e. no credit constraints or non-pecuniary benefits of education)

$$\Omega = \int_0^T e^{-rt} \left[A h(t)(1-\sigma(t)) \right] dt \tag{3}$$

where r is the interest rate. Marginal returns to education at t

$$\begin{aligned}
\frac{\partial \Omega}{\partial \sigma(t)} &= \underbrace{-e^{-rt} A h(t)}_{\text{opportunity costs}} + \int_0^T e^{-rz} A \left[1-\sigma(z) \right] \frac{\partial h(z)}{\partial \sigma(t)} dz \\
&= \underbrace{-e^{-rt} A h(t)}_{\text{opportunity costs}} + \int_t^T e^{-rz} A \left[1-\sigma(z) \right] \theta h(z) dz
\end{aligned} \tag{4}$$

since (see CCZ 2.2.1) $\frac{\partial h(z)}{\partial \sigma(t)} = 0$ if $z < t$ and $\frac{\partial h(z)}{\partial \sigma(t)} = \theta h(z)$ if $z \geq t$

Human capital over the life cycle

The derivative of these marginal returns to education with respect to t is

$$\frac{d}{dt}\left[\frac{\partial \Omega}{\partial \sigma(t)}\right] = -e^{-rt} A\dot{h}(t) + e^{-rt} rAh(t) - e^{-rt} A\left[1-\sigma(t)\right]\theta h(t)$$

Plugging in equation (2) yields

$$\frac{d}{dt}\left[\frac{\partial \Omega}{\partial \sigma(t)}\right] = Ah(t)e^{-rt}(r-\theta)$$

- If $r > \theta$ ($r < \theta$) the marginal returns to education increase (decrease) over time

Human capital over the life cycle

Moreover, the marginal returns to education are always negative at T. From equation (4)

$$\frac{\partial \Omega}{\partial \sigma(T)} = -e^{-rT} Ah(T) < 0$$

Therefore, if $r > \theta$ the marginal returns to education are negative over the whole life cycle $[0, T]$.

Individuals will invest into education only if (i) they are patient enough (as measured by r) and (ii) they are sufficiently efficient in acquiring education.

Human capital over the life cycle

If $r < \theta$ the marginal return to effort decreases over time and is negative at T

- Individuals stop accumulating HC at date s, defined by $\frac{\partial \Omega}{\partial \sigma(s)} = 0$
- $\sigma(t) = 1$ for $t < s$ and $\sigma(t) = 0$ for $t > s$
- $h(t) = h_0 e^{\theta s}$ for $t \geq s$.

Plug $\frac{\partial \Omega}{\partial \sigma(s)} = 0$, $\sigma(t) = 0$ for $t > s$ and $h(t) = h_0 e^{\theta s}$ for $t \geq s$ into equation (4)

$$\frac{\partial \Omega}{\partial \sigma(s)} = 0 = -e^{-rs} A h(s) + \int_{s}^{T} e^{-rz} A \theta h(z) dz$$

Human capital over the life cycle

It follows (check)

$$s = \begin{cases} T + \frac{1}{r} ln(\frac{\theta - r}{\theta}) & \text{if } \theta \geq \frac{1}{1-e^{-rT}} \\ 0 & otherwise \end{cases}$$

Implications. Duration of schooling ...

- Increases with life duration T
  - Individuals/populations with longer expected life-span will acquire more human capital.
- Increases with efficiency θ
  - Most efficient learner spends longest time in education
- Decreases with interest rate r

Extended model

In an extended model, consider the law of motion (CCZ 2.3.1),

$$\dot{h}(t) = \theta g(\sigma(t)h(t)) - \delta h(t),$$

where $\delta \geq 0$ is the rate of depreciation of knowledge and the function g is concave ($g' > 0$ and $g'' < 0$)

With these modifications the model can explain

- part-time education
- education at older age

and yields realistic hump-shape profiles for education and wages

Extended model: Three phases of HC accumulation

Phase 1: $\sigma(t) = 1$

- Full-time education
- Opportunity costs (foregone earnings) small
- Returns to education large (many working years left)

Phase 2: $\sigma(t) \in (0, 1)$

- Some education and some work
- e.g. on-the-job learning or training

Phase 3: $\sigma(t) = 0$

- No education
- HC and earnings decline due to depreciation

Human capital over the life cycle

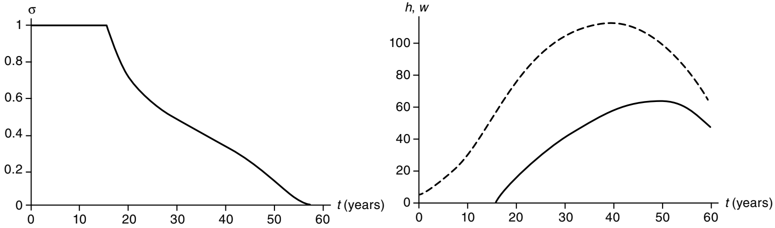


FIGURE 4.9
The law of motion of time dedicated to education (graph on the left), stock of human capital (dotted line in the graph on the right), and wage gains (solid line in the graph on the right) in the human capital model for an efficiency coefficient $\theta = 0.5$.

Human capital over the life cycle

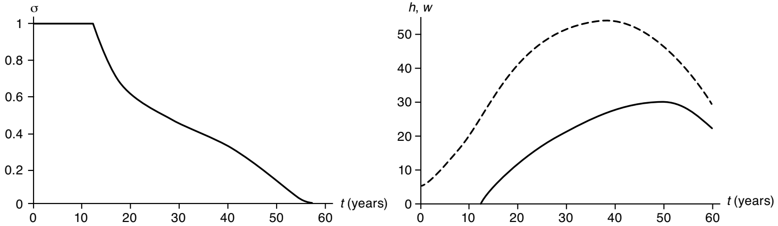


FIGURE 4.10
The law of motion of time dedicated to education (graph on the left), stock of human capital (dotted line in the graph on the right), and wage gains (solid line in the graph on the right) in the human capital model for an efficiency coefficient $\theta = 0.4$.

Other determinants of life cycle profiles

This model can helps us understand life-cycle profiles in earnings and working hours. But other factors may matter:

Age effects, experience effects, and "learning-by-doing"

- Work hours peak earlier than the hourly wage
- Cannot be explained by standard models of intertemporal labor supply or HC accumulation
- But if work experience increases future wages, individuals may want to accumulate experience early in life

Search capital and "job ladder". Older individuals had more time to find a better job ($\rightarrow$ Equilibrium Search and Monopsony Models)

The Mincer regression

Our life-cycle model of human capital accumulation implies an earnings function of the form (CCZ 4.1.1)

$$logw(H_i) = logw(0) + \rho H_i$$

Mincer (1958, 1974) considers two different models of schooling choices that motivate the so-called **Mincer regression**

$$logw_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 X_i^2 + \varepsilon_i$$

distinguishing between formal schooling S_i and labor market experience X_i (e.g. representing on-the-job learning)

The Mincer regression

The **Mincer regression**

$$logw_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 X_i^2 + \varepsilon_i$$

where S_i is schooling, X_i is labor market experience

Key implications (Heckman, Lochner and Todd, 2006):

1. Log earnings are linear in schooling
2. Experience-earnings (log) profiles are parallel across schooling levels
3. Age-earnings (log) profiles diverge with age
4. Variance of earnings over life cycle has U-shaped pattern

plus others (see Bhuller, Mogstad and Salvanes 2017)

Figure: Log earnings are linear in schooling? US data

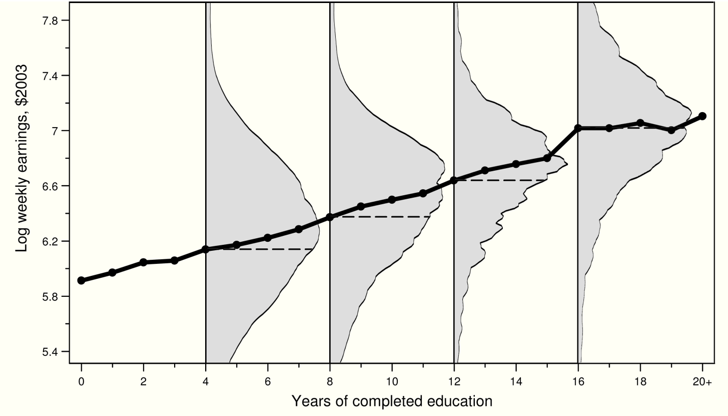
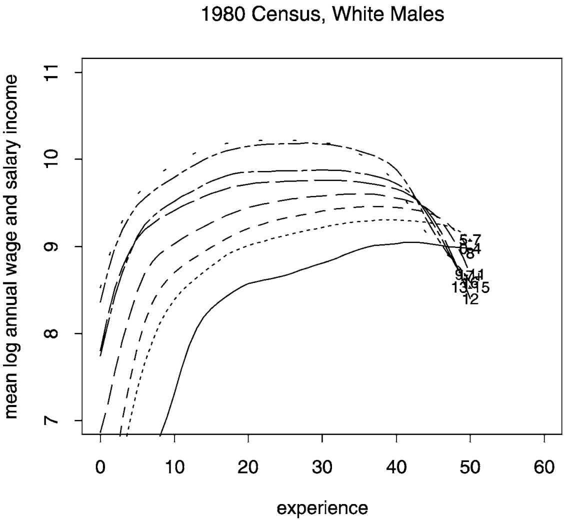


Figure 3.1.1: Raw data and the CEF of average log weekly wages given schooling. The sample includes white men aged 40-49 in the 1980 IPUMS 5 percent file.

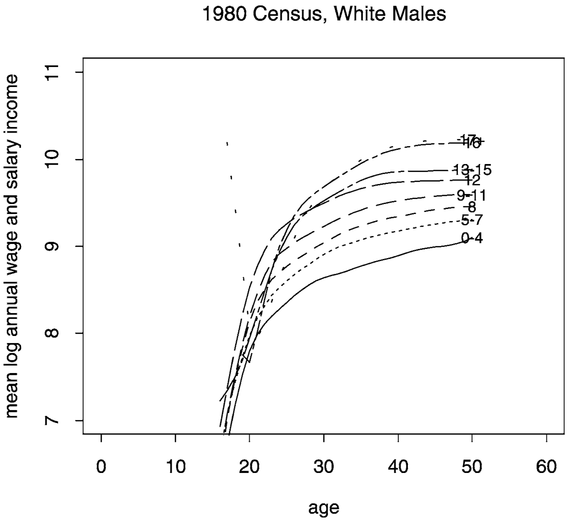
Source: Angrist and Pischke (2008). Cross-sectional data.

Figure: Experience-earnings (log mean) profile



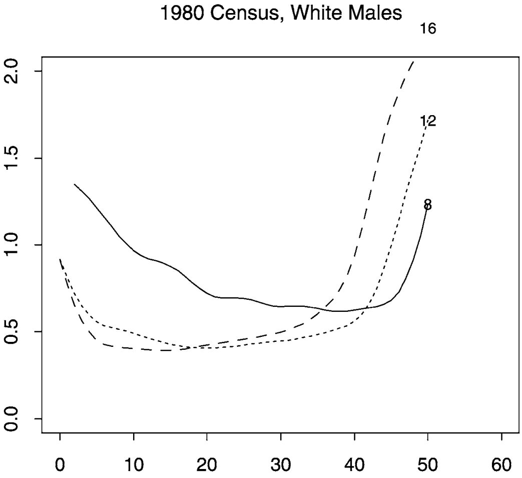
Source: Heckman, Lochner and Todd (2006)

Figure: Age-earnings (log mean) profile



Source: Heckman, Lochner and Todd (2006)

Figure: Experience-earnings (log variance) profile



Source: Heckman, Lochner and Todd (2006)

The Mincer regression

The *Mincer regression*

$$logw_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 X_i^2 + \varepsilon_i$$

where S_i is schooling, X_i is labor market experience.

Typical findings:

- Positive returns to schooling
  (e.g. OLS estimates of $\beta_1 \approx 0.12$ in Heckman et al. 2006)
- Concave returns to experience ($\beta_2 > 0$ and $\beta_3 < 0$)

Internal rate of return

Internal rate of return (IRR):

- Defined as the rate of return that equates the net present value of all benefits and costs from an investment
- Under certain conditions (see Heckman, Lochner and Todd 2006) β_1 from the Mincer regression is the IRR to schooling

Complications:

- ex-ante vs. ex-post returns?
- uncertainty?

A puzzle

OLS estimates of the Mincer equation imply that in the US, (real) returns to an additional year of schooling are ten percent or more.

A puzzle:

- Few other investments have such high returns, so why are not more people attending university?

Two possible answers:

1. OLS estimates of β_1 are biased ("selection bias")
2. Heterogeneity in returns or costs to education ("sorting on gains")

Selection bias

Assume error term partial reflects unobserved "ability" a_i, such that $\varepsilon_i = \gamma a_i + \tilde{\varepsilon}_i$ and (abstracting from experience)

$$logw_i = \beta_0 + \beta_1 S_i + \gamma A_i + \tilde{\varepsilon}_i$$

where $\tilde{\varepsilon}_i$ is uncorrelated with S_i or A_i

Probability limit of OLS estimator is then (omitted variable bias)

$$plim\hat{\beta}_1 = \frac{Cov(logw_i, S_i)}{Var(S_i)} = \beta_1 + \gamma \frac{Cov(A_i, S_i)}{Var(S_i)}$$

so if ability affects earnings ($\gamma > 0$) and individuals with high ability tend to have more schooling ($Cov(A_i, S_i) > 0$) then the OLS estimator is upward biased ($E[\hat{\beta}_1] > \beta_1$)

Selection bias

How to deal with selection bias? Common approaches:

(1) **Control approach**

- Control for "ability" in the Mincer regression
- e.g. include IQ scores (Grilliches 1977)

(2) **Twin approach**.

- Try to eliminate ability from Mincer regression

(3) **Instrumental variable approach.** A valid IV predicts schooling (rank condition) and affects wages only via schooling (exclusion restriction). Example IVs:

- Date of birth interacting with compulsory schooling age
- Compulsory school reforms
- Distance to college (Card 1993)

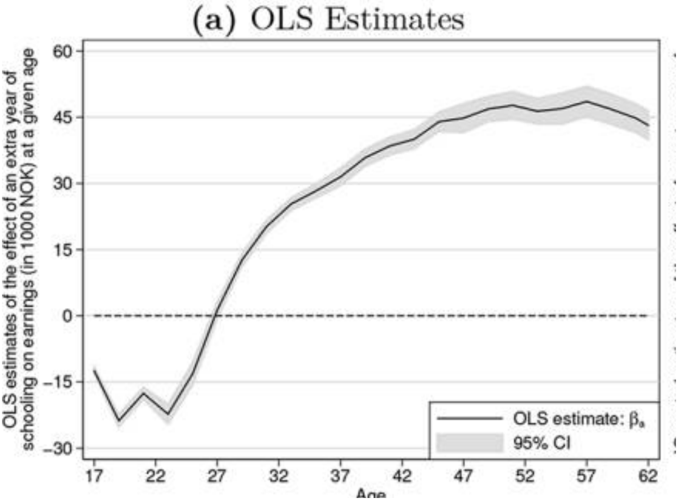
Bhuller, Mogstad and Salvanes (2017)

Bhuller, Mogstad and Salvanes (2017) apply all three approaches, using great data:

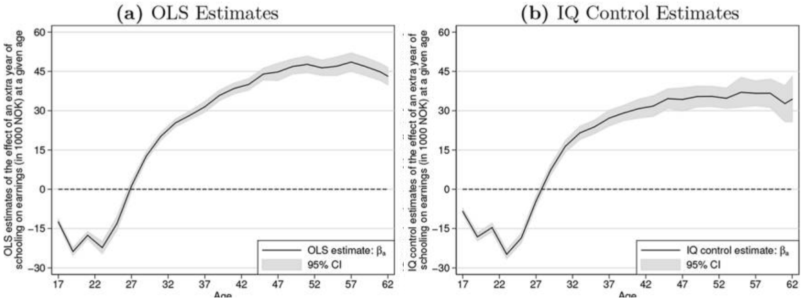
- Administrative panel data for Norway
- Full population, near career-long earning histories (1967-2014)
- Pre-tax labor income and benefits / post-tax / benefits
- Ability/IQ measures from military enlistment tests
- IV: staggered implementation of compulsory school reform

Separate regression at each age:

Bhuller, Mogstad and Salvanes (2017): OLS



Bhuller, Mogstad and Salvanes (2017): Control approach



- Estimated return decreases when controlling for IQ
- Suggestive of upward ability bias in OLS estimates

Bhuller, Mogstad and Salvanes (2017): IV approach

IV approach:

- exploits gradual roll-out of compulsory school reform across counties
- increases schooling duration (“first stage”)
- reform has biggest effect on those with low schooling
- implemented at different times, so can control for county and time fixed effects (similar as in difference-in-differences setting)

Bhuller, Mogstad and Salvanes (2017): IV first stage

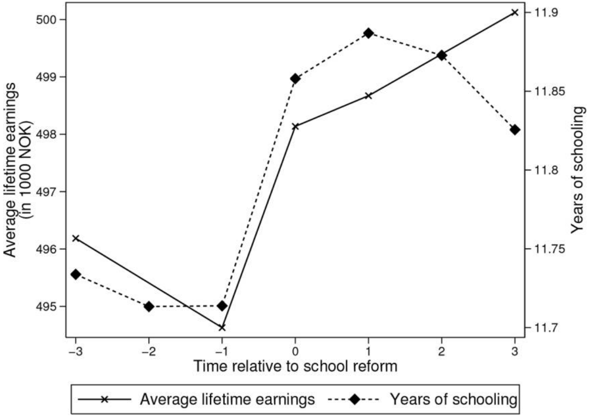
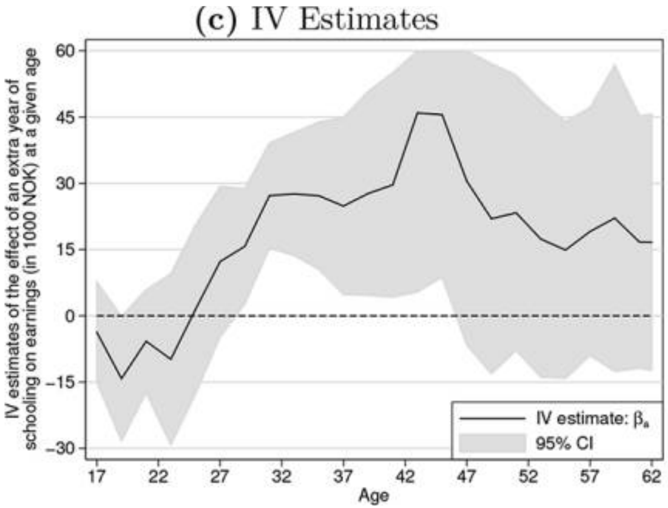


FIG. 2.—Graphical illustration of the instrumental variable (IV) approach. For each municipality, we recenter the data such that time zero is the year in which the reform was implemented. Variables are residuals from a regression on birth cohort and municipality fixed effects (adding in a common intercept). For each individ-

Bhuller, Mogstad and Salvanes (2017): IV approach



- IV estimates similar as OLS estimates, but much noisier

Bhuller, Mogstad and Salvanes (2017): Twin approach

Twin approach

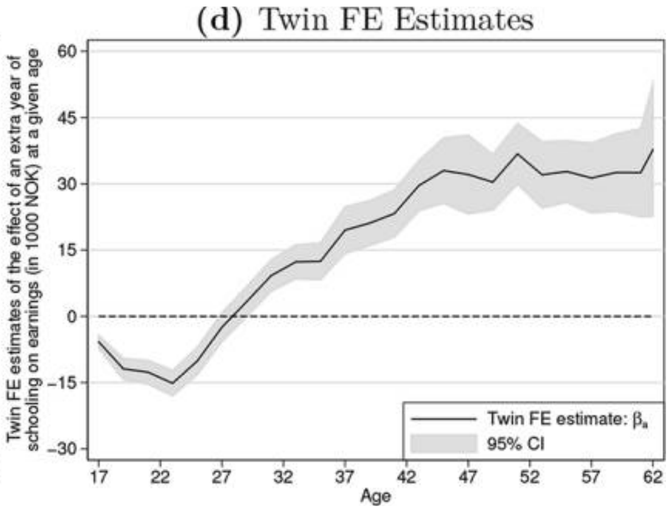
- 6,434 twins (monozygotic and dizygotic)
- Compare difference in schooling of twins with their difference in earnings

Intuition:

- Twins have similar abilities (e.g. monozygotic twins share both genetic endowments and family environment)
- If they have the *same* ability then differencing would eliminate ability from the Mincer regression (see Griliches 1979)
- In practice, questionable if differencing eliminates more of the omitted variable (ability) or the regressor of interest (schooling)

Bhuller, Mogstad and Salvanes show that military test scores are similar for twins

Bhuller, Mogstad and Salvanes (2017): Twin approach



Bhuller, Mogstad and Salvanes (2017): IRR

IRR defined here as the discount rate ρ that equates ...

$$\sum_{age=17}^{62} \tilde{r} \frac{\beta_{age}}{(1+\rho)^{age-16}} = 0$$

	Full Sample, OLS Estimate (1)	IQ Sample, IQ Control Estimate (2)	IV Sample, IV Estimate (3)	Twins Sample, Twin FE Estimate (4)
Pretax earnings	.093***	.083***	.112**	.089***
	(.002)	(.003)	(.048)	(.008)
After-tax income	.069***	.068***	.091**	.072***
	(.002)	(.003)	(.041)	(.007)
After-tax income + pension income	.069***	.069***	.091**	.072***
	(.002)	(.003)	(.038)	(.007)
N	601,290	325,417	577,098	6,434

NOTE.—For each identification strategy, we report estimates of internal rates of return in pretax earnings, after-tax income, and the sum of after-tax income and pension entitlements. All regressions include fixed effects for childhood municipality and birth cohort. Standard errors (in parentheses) are computed by nonparametric bootstrap with 250 replications. FE = fixed effects.

** $p < .05$.

*** $p < .01$.

Still a puzzle

Even after addressing selection, returns to schooling appear much larger than the market interest rate.

In fact, IV estimates of β_1 in

$$logw_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 x_i^2 + \varepsilon_i$$

typically *exceed* the OLS estimates. Potential explanations:

- Measurement error
- Invalid instruments (e.g. quarter of birth)
- Publication bias: IV estimates often have larger SEs than OLS $\rightarrow$ IV estimates are statistically significant and get published only if they are large (Ashenfelter, 1999)
- IV estimates based on a particular type of transition and effect of education may be non-linear
- IV identifies only a *local average treatment effect* (LATE)

OLS and IV weights

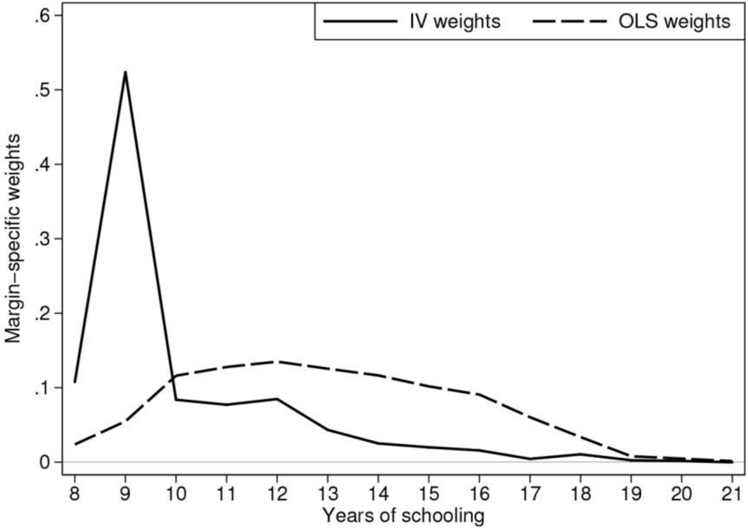


FIG. C1.—Ordinary least squares (OLS) and instrumental variable (IV) weights for every grade-specific effect.

Bhuller, Mogstad and Salvanes (2017)

Bhuller, Mogstad and Salvanes (2017) main findings:

1. Estimated internal rate of return ($\approx 11\%$) much larger than market interest rate, even when addressing selection bias
2. Mincer regressions strongly understate returns to schooling because some of its assumptions [parallelism over experience across schooling levels, stationary environment, no earnings while in school, exogenous employment status] are violated

IV estimates, LATE, and MTE

IV estimates of

$$logw_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 x_i^2 + \varepsilon_i$$

typically *exceed* the OLS estimates.

Potential explanation:

- Returns to schooling vary across individuals (Carneiro, Heckman and Vytlacil, 2011)
- Individuals select schooling based on their idiosyncratic gains. In particular, returns to schooling might be high for those who acquire schooling, but lower for those who do not
- Standard IV identifies only the gains for those individuals whose schooling level has changed because of the instrument $\rightarrow$ local average treatment effect (LATE)

Reminder: Potential outcome model

The *"Potential Outcome Model"*:

The **treatment** (e.g. college attendance yes/no)

$$D_i = \begin{cases} 1 & \text{individual } i \text{ receives treatment} \\ 0 & \text{individual } i \text{ does not receive treatment} \end{cases}$$

The **observed outcome** is a function of ***potential outcomes***

$$Y_i = \begin{cases} Y_{1i} & \text{if } D_i = 1 \text{ (treated outcome)} \\ Y_{0i} & \text{if } D_i = 0 \text{ (untreated outcome)} \end{cases}$$

The observed outcome can be written as $Y_i = Y_{0i} + (Y_{1i} - Y_{0i})\, D_i$, where $\tau_i = Y_{1i} - Y_{0i}$ is the individual treatment effect

Definition of treatment effects

Average treatment effect (ATT):

$$\tau_{ATE} = E[\tau_i] = E[Y_{1i} - Y_{0i}]$$

Average treatment effect for the treated (ATT):

$$\tau_{ATT} = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]$$

Average treatment effect for the untreated (ATU):

$$\tau_{ATU} = E[Y_{1i}|D_i = 0] - E[Y_{0i}|D_i = 0]$$

Local average treatment effect (LATE):

$$\tau_{LATE} = E[Y_{1i} - Y_{0i}|Compliers]$$

where compliers are those individuals whose treatment status has switched because of the instrument

IV estimates, LATE, and MTE

Standard IV identifies only local average treatment effect (LATE):

$$\tau_{LATE} = E[Y_{1i} - Y_{0i} | Compliers]$$

- e.g. IV estimate in Bhuller, Mogstad and Salvanes (2017) identifies returns to schooling for individuals whose schooling decisions changed because of compulsory schooling law
- Very particular group → very "local" effect

Potential solution: Marginal treatment effect (MTE) approach (Björklund and Moffitt 1987; Heckman and Vytlacil 1999, 2005)

- Intuition: Estimate many LATEs for many different groups of compliers

Application: Nybom (2017)

Nybom (2017) estimates lifetime earnings returns to college in Sweden, and how those returns vary with

- Observed characteristics (cognitive skills, non-cognitive skills, parental income)
- Unobserved characteristics ($\rightarrow$ MTE)

Heavy data requirements to estimate the MTE semi-parametrically. Local IV approach requires continuous instrument

Instruments used in Nybom (2017):

1. Distance to closest university
2. Short-run fluctuations in local labor market conditions in final year of high school

Exclusion restrictions may not be perfectly satisfied

Nybom (2017): OLS estimates

Table 2
Ordinary Least Squares (OLS) Estimates of the Return to a Year of College

	OLS Coefficients				
	(1)	(2)	(3)	(4)	(5)
College dummy (S)	.0572	.0571	.0452	.0388	.0391
	(.0008)	(.0008)	(.0008)	(.0009)	(.0009)
$S \times \mathbf{A}$ (cognitive)				.0067	.0064
				(.0006)	(.0006)
$S \times \mathbf{A}$ (noncognitive)				.0056	.0050
				(.0004)	(.0004)
$S \times \mathbf{A}$ (father's earnings)				−.0002	−.0005
				(.0012)	(.0014)
Conditional on **A**			X	X	X
Interactions $S \times \mathbf{A}$				X	X
Interactions $S \times \mathbf{X}$		X			X

NOTE.—This table reports OLS estimates of the return to college. All specifications control for **X**, which includes region and cohort dummies, linear and quadratic terms of father's and mother's years of schooling, number of siblings, and local long-run earnings at age 20. Specifications 3–5 include linear and quadratic terms of cognitive and noncognitive ability and log of father's earnings (i.e., **A**), specifications 2 and 5 include interactions between S and all components of **X**, and specifications 4 and 5 include interactions between S and all components of **A**. The interaction terms ($S \times \mathbf{A}$) are reported as average derivatives. I obtain annualized returns by dividing all estimates by 4.3, which is the average difference in years of schooling for those with $S = 1$ and those with $S = 0$. Standard errors (from 1,000 bootstrap replications) are in parentheses.

Potential outcomes (observed/unobserved heterogeneity)

Let S be binary choice indicator

$$S_i = 0 \text{ no college (untreated)}$$
$$S_i = 1 \text{ college (treated)}$$

Assume potential outcomes are

$$Y_{0i} = \mu_0(X_i) + U_{0i}$$
$$Y_{1i} = \mu_1(X_i) + U_{1i}$$

where X_i are observed regressors, such as cognitive and non-cognitive skills

Idiosyncratic gains from treatment are

$$Y_{1i} - Y_{0i} = \mu_1(X_i) - \mu_0(X_i) + U_{1i} - U_{0i}$$

Generalized Roy model

The propensity score

$$P_i(x,z) = Pr(S_i = 1 | X_i = x, Z_i = z)$$

denote the conditional probability to attend college for people with characteristics $X_i = x$ and $Z_i = z$.

- Z_i assumed to be valid instrument: affects the college decision (rank condition) but not potential outcomes (exclusion restriction)

However, there is unobserved heterogeneity. Let U_{Si} represent (the quantiles of) idiosyncratic and latent **resistance** to college. Then

$P_i(x,z) > U_{Si}$ individual attends college ($S_i = 1$)
$P_i(x,z) = U_{Si}$ individual is indifferent
$P_i(x,z) < U_{Si}$ individual does not attend college ($S_i = 0$)

MTE: Definition

MTE is defined as

$$MTE(X_i = x, U_{Si} = u_S) = E[Y_{1i} - Y_{0i} | X_i = x, U_{Si} = u_s] \qquad (5)$$

where U_{Si} is the individual unobserved resistance to treatment

MTE may vary with X_i or U_{Si}

- If MTE sloped with respect to U_s: "unobserved heterogeneity"
- If MTE is sloped with respect to X: "observed heterogeneity"

MTE: Estimation

MTE can be estimated parametrically or semi-parametrically.

Nybom (2017) implements both approaches, but focuses on semi-parametric **local IV approach**.

Local IV approach:

- Estimate $E[Y_i|X_i = x, P(Z_i) = p]$ semi-parametrically for all values of $x, p = u_S$
- Compute derivative with respect to p, because

$$MTE(X_i = x, U_{Si} = u_S) = \left. \frac{\partial E[Y_i|X_i = x, P(Z_i) = p]}{\partial p} \right|_{p=u_s} \tag{6}$$

This MTE can then be aggregated for certain subgroups, e.g. to compute the ATE, ATT or ATU

MTE: Continuous instrument

Local IV approach identifies the MTE under minimal assumption. However, requires a *continuous* instrument to generate marginal expansions in college attendance

To understand the intuition, assume (from Cornelissen, Dustmann, Raute and Schönberg, 2016):

- those living right next to college always attend college
- those extremely far away from college never attend college
- Gradually decreasing distance will then gradually all “types” (U_S) into college

See figure (next slide)

MTE: Continuous instrument

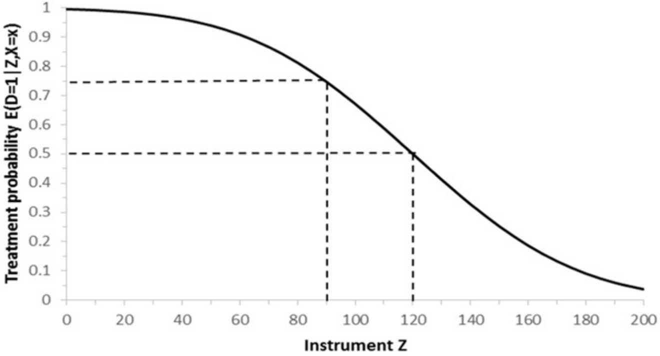


Fig. 1. Treatment probability as a function of a continuous instrument. *Notes:* Based on hypothetical data, the figure shows the effect of a continuous instrument Z on the probability of treatment in a sample with fixed covariates ($E[D = 1,Z,X = x]$). For example, the horizontal axis could represent distance to college and the vertical axis could represent the probability to attend college. *Data source:* Simulated hypothetical data.

Source: Cornelissen, Dustmann, Raute and Schönberg (2016)

MTE: Continuous instrument

In the ideal case, in which the instrument varies strongly conditional on X, estimation of MTE requires assumptions that are no stronger than the assumptions for conventional IV estimation

- Need instrument with very large support

In practice, we rarely have that much support in the instrument and require auxiliary assumptions

- For example, Nybom (2017) assumes that X_i is independent of (U_{0i}, U_{1i}, U_{Si}), i.e. imposes separability on the conditional expectation in equation (6)
- Intuition: Assume that treatment effect varies with the unobserved characteristic in same way irrespective of other, observed characteristic

Nybom (2017): Common Support

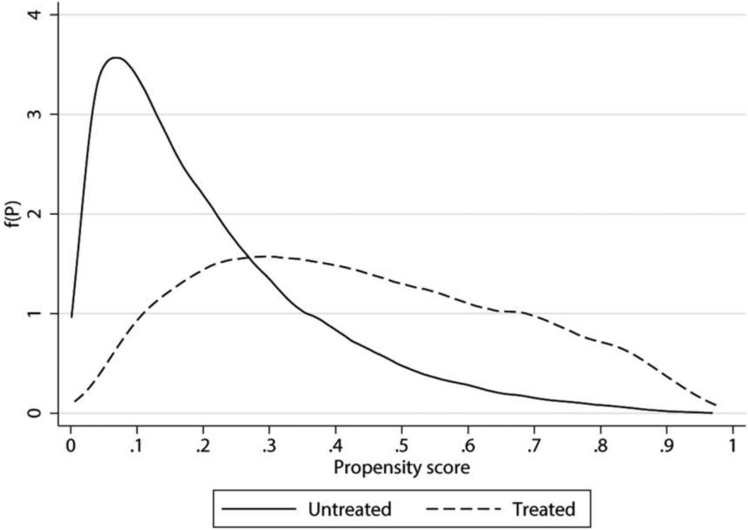


FIG. 2.—Support of $P(\mathbf{Z})$ for untreated ($S = 0$) and treated ($S = 1$). This figure shows the support of $P(\mathbf{Z})$ for the treated and the untreated. The parameter $P(\mathbf{Z})$ is the probability of going to college estimated in a probit regression (see table 5).

Nybom (2017): Semi-parametric estimates of MTE

Figure: Returns to a Year of College

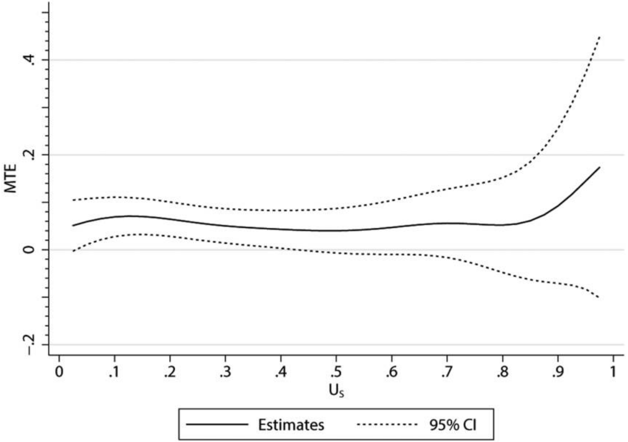


FIG. 3.—Marginal treatment effect (MTE) by U_S estimated by semiparametric local instrumental variable analysis. This figure shows point estimates and 95% confidence bands of the MTE from the semiparametric model in equation (5). The model is estimated using the local quadratic regression procedure described in Section II. All estimates are conditioned on mean values of **X** and **A**. Standard errors are boot-

Nybom (2017): Semi-parametric estimates of MTE

Findings:

- MTE flat $\rightarrow$ not much heterogeneity in returns to college with respect to unobserved determinants of college decision
- Difference between lowest and highest MTEs only about 3 percentage points
- MTE decreasing at lower values of U_S (selection on gains)
- MTE increasing at higher values of U_S (*negative* selection on gains)

Nybom (2017): Implied ATE, ATT and ATU

Table 4
Returns to a Year of College

	Normal Model (1)	Semiparametric Model (2)
ATE	.0269	.0573
	(.0043)	(.0201)
ATT	.0380	.0751
	(.0048)	(.0179)
ATU	.0227	.0506
	(.0042)	(.0210)
ATT − ATU	.0152	.0245
	(.0012)	(.0049)
ATT − ATE	.0111	.0177
	(.0009)	(.0035)
ATE − ATU	.0042	.0067
	(.0003)	(.0013)

NOTE.—This table reports estimates of the average treatment effect (ATE), treatment effect on the treated (ATT), and treatment effect on the untreated (ATU) from the normal selection model in equation (9) and the semiparametric model. The latter is thus for $\widetilde{ATE}$, $\widetilde{ATT}$, and $\widetilde{ATU}$, indicating that they are sample-specific parameters with weights integrating to one over the empirical support of U_S. Rows 4–6 show the estimated differences between the treatment effect parameters. All estimates are annualized, reflecting the average difference in years of schooling for those with $S = 1$ and those with $S = 0$. Standard errors are obtained using the bootstrap (1,000 replications).

- ATE > ATU, but lifetime earnings returns for those who do not go to college would still be positive

Nybom (2017): Heterogeneity in Returns to Observed Characteristics

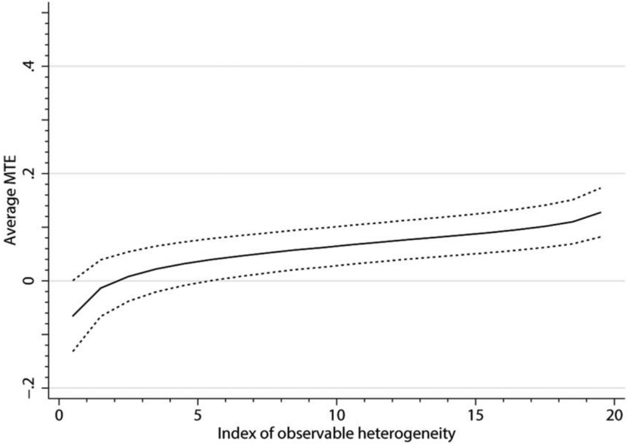


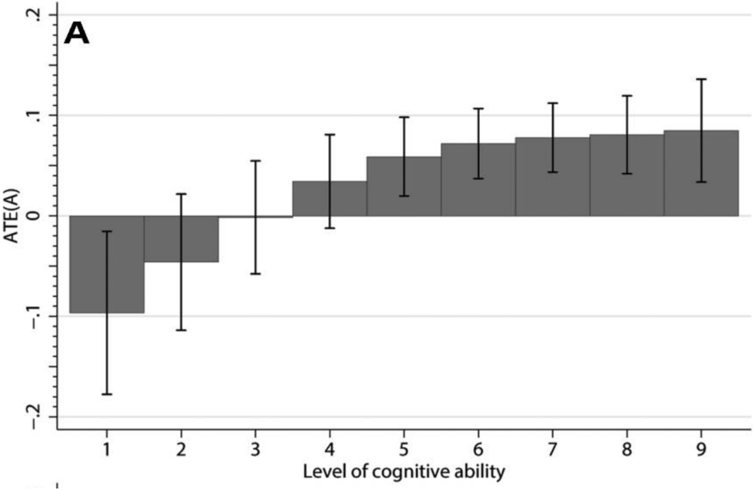
FIG. 4.—Average marginal treatment effect (MTE) by total observed heterogeneity. This figure shows the average MTE with 95% confidence bands across the index of observed heterogeneity. The index is computed by estimating $\mathbf{X}(\delta_1 - \delta_0) + \mathbf{A}(\gamma_1 - \gamma_0)$ for each individual and splitting the sample into 20 uniformly distributed groups. Standard errors are bootstrapped (1,000 replications).

Nybom (2017): Heterogeneity in Returns to Observed Characteristics

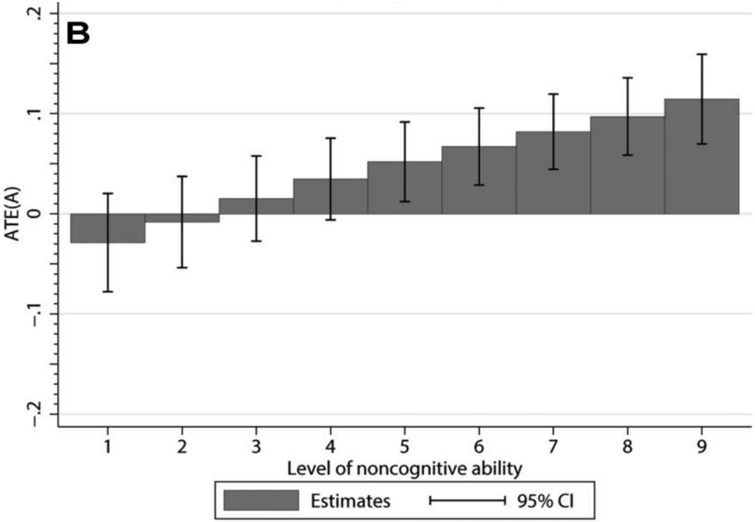
Findings:

- Substantial variation in lifetime earnings returns with respect to observable characteristics
- Those with most complementary characteristics (e.g. those with high cognitive and non-cognitive skills) have a lifetime return to college that is 20 percentage points higher than those with the least favorable combination of characteristics
- Observed heterogeneity much more important than unobserved heterogeneity

Nybom (2017): Heterogeneity in Returns to Observed Characteristics



Nybom (2017): Heterogeneity in Returns to Observed Characteristics



Nybom (2017)

Main findings:

- Individuals select into college based on idiosyncratic gains (ATT > ATU)
- But difference is small and ATU remains positive
- Substantial heterogeneity in returns with respect to cognitive and non-cognitive ability
  - Individuals at bottom of ability distribution have negative returns to college
  - Individuals at top earn returns that are twice as high as the average return
- Suggests school-skill complementarities in the labor market

Deming (2015)

Cognitive and non-cognitive skills are important determinants of earnings.

Recent findings: Deming (2017)

- Labor market increasingly rewards non-cognitive and *social* skills
- Between 1980 and 2012, share of jobs in U.S. requiring high levels of social interaction grew by nearly 12 percent
- Employment and wage growth particularly strong for jobs requiring high levels of both math and social skills

Proposes a model in which social skills reduce coordination costs

Deming (2015)

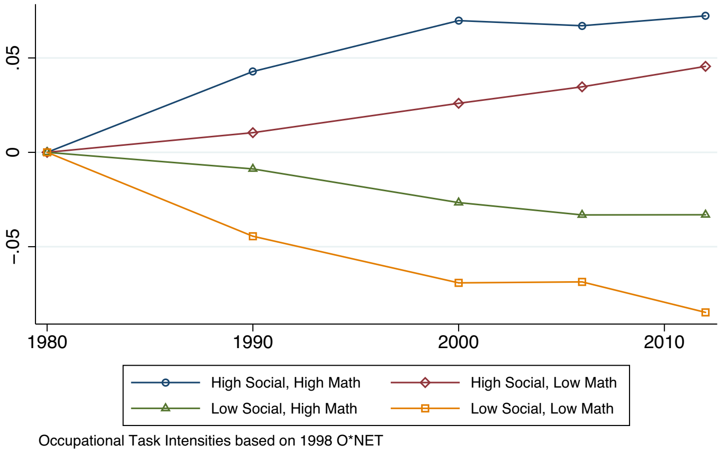


FIGURE IV

Cumulative Changes in Employment Share by Occupation Task Intensity, 1980–2012

Deming (2015)

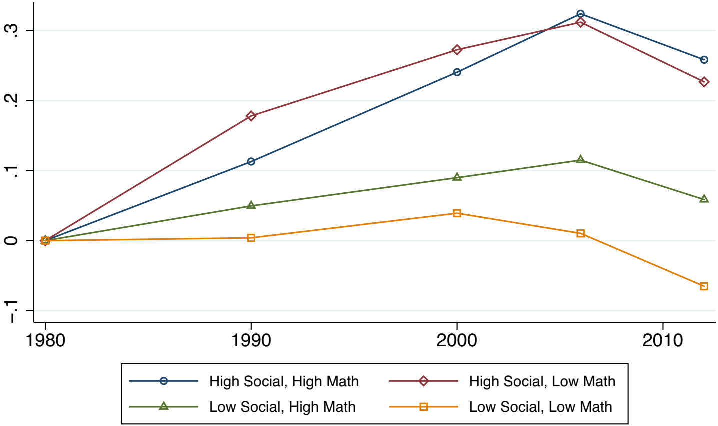


FIGURE V

Cumulative Changes in Real Hourly Wages by Occupation Task Intensity, 1980–2012

Dynamic complementarity

At what age are HC investments most effective? Important policy question

The empirical evidence suggests that interventions in early life (e.g. preschool programs, formal childcare) tend to be more effective

This finding can be motivated with an investment model with **dynamic complementarity** (Heckman and Cunha 2007, 2009):

- HC investments in different stages of the life cycle are complementary; self-productivity
- HC investment in early childhood makes later investments more efficient

Described in Heckman and Mosso (2014)

Returns to HC investment over age

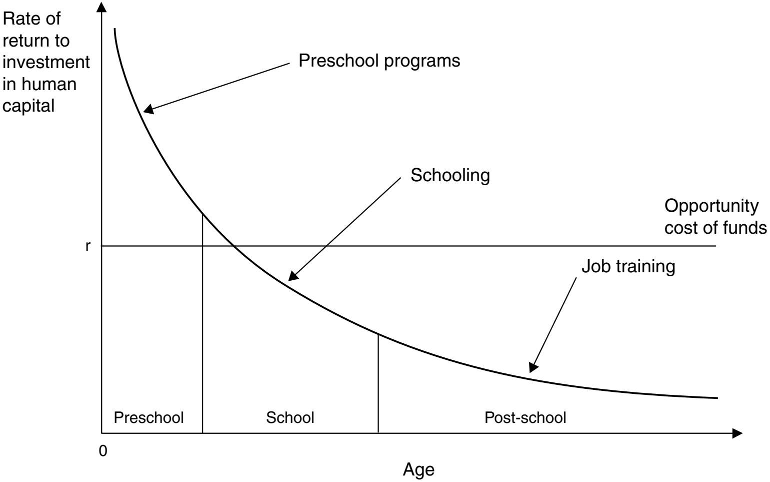


FIGURE 4.20
Rate of return to investment in human capital by age.

Social returns

Social returns: returns to society as a whole (minus private returns)

Many reasons why education could have positive social returns

- Knowledge externalities
- Education appears to have a strong, negative effect on crime
- Geographical mobility increases in education, so local labor markets may adjust more quickly to demand shocks

But in principle, social returns could also be negative

- e.g. see signaling model by Spence

However, social returns are difficult to estimate

Social returns: Moretti (2004)

Moretti (2004) estimates spillovers from share of college graduates on workers in a city.

Controls for ...

Individual-level heterogeneity

- Individual fixed effects

City-level heterogeneity over time

- Control for city-specific demand shocks that can be predicted by industry composition in city (similar as *Bartik* instrument)
- IV strategy: Differences in demographic structure across cities leads to differential trends in college share

Results: College share has large positive externality, in particular on low skilled workers. See Tables 8 and 9 in paper

Human capital: Aggregate implications

Human capital has important *aggregate* implications:

Key determinant of economic growth

- Human capital is complementary to physical capital
- Investment in education may increase society's productive capacity (Adam Smith's "The Wealth of Nations", 1776)
- Cross-country evidence: rapid economic growth only in countries with mass primary schooling (Easterlin, 1981)

... and inequality?

- Technology may be *skill-biased*
- Inequalities widen if HC accumulation does not keep pace with technological change (Tinbergen's "Race between education and technology")
- Investments into human capital could reduce inequalities, but HC accumulation is slowing

Human capital: Aggregate implications

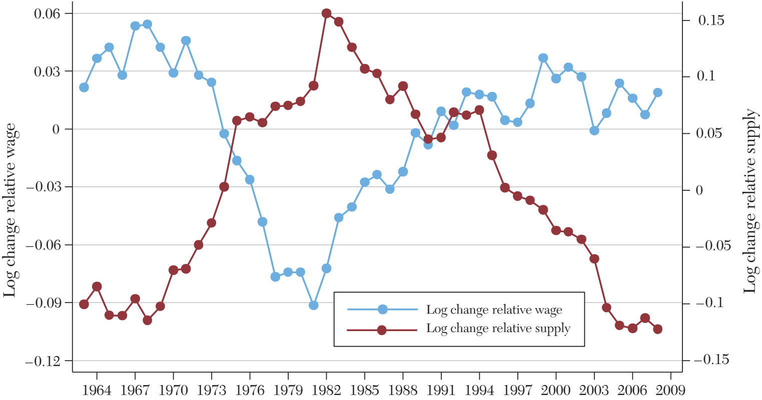


Figure 2. Detrended Changes in College–High School Relative Supply and Relative Wages

Source: March CPS data for earnings years 1963–2008. See notes to figure 1. The detrended supply and wage series are the residuals from separate OLS regressions of the relative supply and relative wage measures on a constant and a linear time trend.

Source: Acemoglu and Autor (2012)

Readings

Required Readings:

- Heckman, Lochner and Todd (2006): Sections 1, 3, 4, 6, 7, 11
- Heckman and Mosso (2014): Sections 1-3 + skim through rest
- Nybom (2017): definition and intuition of MTE