Human Capital and Skill Differentials

UniTo, Labor Economics Part II
Christoph Albert

Lecture 1

Agenda

Course Logistics and Requirements

Human Capital

- General and specific human capital
- HC over the life-cycle and the Ben-Porath model
- The Mincer regression and returns to schooling

Empirical Applications

- Bhuller, Mogstad and Salvanes (2017) on lifecycle and OVB
- Nybom (2017) on selection on gains and MTEs

Extensions

- Income processes beyond Mincer
- Multiplicity, dynamic complementarity and social returns

Course Logistics and Requirements

- We mainly focus on **very recent** papers in empirical labor economics
- Useful auxiliary readings are
  "Labor Economics" by Cahuc, Carcillo and Zylberberg (CCZ),
  "Lectures in Labor Economics" by Acemoglu and Autor (AA).
- christoph.albert@carloalberto.org
  (office hours: before class or by appointment)

Weekly Presentations

One presentation by a student at the beginning of each class. You can choose a paper from the reading list (in moodle) based on first-come first-served basis.

The presentation should answer the following questions

1. What is the research question?
2. What are the methods used to answer this questions
3. What are the **strengths** and **weaknesses** of the approach to answering that question?
4. How does this work advance knowledge on the question, i.e, what's the contribution?
5. What would be one or two valuable next steps to advance the question?

Research Projects/Proposals

Proposal or project of no more than 3500 words

- Pose a research question and explain why it is important.
- Summarize the state of knowledge on this question.
- Explain your idea for advancing knowledge on this question.
- Present the research plan: empirical design, experimental design, simple model, etc.
- Discuss your implementation strategy: data requirements, experimental setting, etc.

Projects vs Proposals

For proposals:

- Discuss the next steps you plan to take to implement this strategy as well as the roadblocks you may face
- We do not expect to see any preliminary results

For projects:

- Describe your data sources and implement your empirical strategy
- A full replication package including all codes and data used to generate your results as well as a Readme file will be part of the assessment.

A note on grading of research proposals vs projects:

Requirements for a good grade in terms of the craft, clarity, and specificity will be higher for a proposal.

Agenda

Course Logistics and Requirements

Human Capital

- General and specific human capital
- HC over the life-cycle and the Ben-Porath model
- The Mincer regression and returns to schooling

Empirical Applications

- Bhuller, Mogstad and Salvanes (2017) on lifecycle and OVB
- Nybom (2017) on selection on gains and MTEs

Extensions

- Income processes beyond Mincer
- Multiplicity, dynamic complementarity and social returns

Human capital: Introduction

Human capital:

-
-
-

Human capital investment:

-
-

Human capital: Introduction

Alternative views and aspects of human capital (see AA's book):

Becker (1964)

- Investment into human capital 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)

- Formal education "signals" inherent ability

Agenda

Course Logistics and Requirements

Human Capital

- General and specific human capital
- HC over the life-cycle and the Ben-Porath model
- The Mincer regression and returns to schooling

Empirical Applications

- Bhuller, Mogstad and Salvanes (2017) on lifecycle and OVB
- Nybom (2017) on selection on gains and MTEs

Extensions

- Income processes beyond Mincer
- Multiplicity, dynamic complementarity and social returns

General and specific human capital

Human capital theory 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 (brief reminder)
2. Optimal investments into 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:
- Specific HC:

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 $y_2'' < 0$

Free market entry implies zero-profit condition for firm:

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

General human capital

Assume on-the-job training generates general HC

- Trained worker produces $y_2(H)$ at *all* firms in the 2nd period. **Under competitive markets, what will other firms offer?**
-

  What will the firm that provided the training offer?
-

Implications:

-
-
- Other contract types by firms to avoid losses through training?
- **Key difference to physical capital?**

Specific human capital

Assume training generates specific HC

- 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:

-
-
-

Agenda

Course Logistics and Requirements

Human Capital

- General and specific human capital
- HC over the life-cycle and the Ben-Porath model
- The Mincer regression and returns to schooling

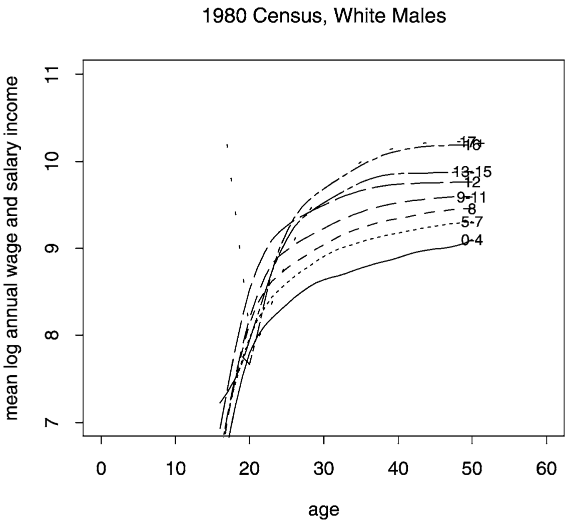
Empirical Applications

- Bhuller, Mogstad and Salvanes (2017) on lifecycle and OVB
- Nybom (2017) on selection on gains and MTEs

Extensions

- Income processes beyond Mincer
- Multiplicity, dynamic complementarity and social returns

Figure: Age-earnings profiles



Source: Heckman, Lochner and Todd (2006): 1980 Census, white males

Human capital over the life cycle

Age-earnings profiles follow a particular pattern:

-
-
-

These pattern can be rationalized with a life-cycle model of human capital accumulation.

Human capital over the life cycle

A life-cycle model of human capital accumulation, based on Ben-Porath (1967). Assumptions (see Section 2.2.1 in CCZ or Chapter 1 in AA's book):

- Human capital is a 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 efficiency of HC accumulation ("learning speed")
- Wage = f(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)

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

where r is the interest rate.

Trade-off between opportunity costs (lost earnings) and benefits (higher future earnings) implies an optimal schooling duration s^*,

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

Human capital over the life cycle

The optimal duration of schooling

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

- increases with life expectancy T
  - individuals or populations with longer expected life-span will acquire more human capital.
  - disease will reduce human capital investments (e.g. Manuelli and Yurdagul, 2020)
- increases with efficiency θ
  - most efficient learner spends longest time in education
- decreases with interest rate r

Extended model with three phases of HC accumulation

For a more realistic hump-shaped profile for wages, consider the extended 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 $g'' < 0$)

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

- Full-time education while opportunity costs (foregone earnings) are small and returns large (many working years left)

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

- Some education, some work (e.g., on-the-job training)

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

- No education, HC and earnings decline due to depreciation

Human capital over the life cycle

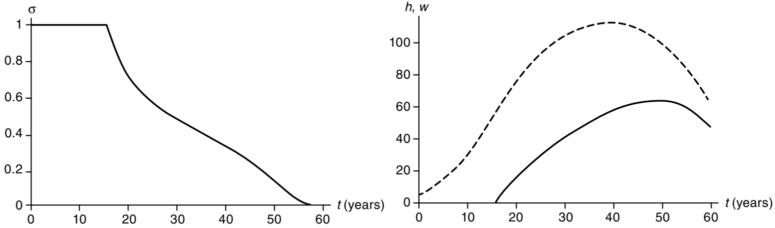


Figure 4.9
The law of motion of time dedicated to education (graph on the left), stock of human capital (dotted line in the graph on the right), and wage gains (solid line in the graph on the right) in the human capital model for an efficiency coefficient $\theta = 0.5$.

Human capital over the life cycle

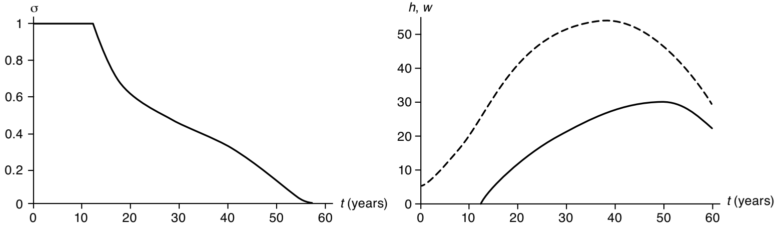


FIGURE 4.10
The law of motion of time dedicated to education (graph on the left), stock of human capital (dotted line in the graph on the right), and wage gains (solid line in the graph on the right) in the human capital model for an efficiency coefficient $\theta = 0.4$.

Other determinants of life cycle profiles

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

Alternative models for HC accumulation:

- "Learning-by-doing" and experience effects

Alternative mechanisms:

- "Search capital" or "Job ladder", as older individuals had more time to search for a better job ($\rightarrow$ Search and monopsony models, such as Burdett and Mortensen 1998)

Agenda

Course Logistics and Requirements

Human Capital

- General and specific human capital
- HC over the life-cycle and the Ben-Porath model
- The Mincer regression and returns to schooling

Empirical Applications

- Bhuller, Mogstad and Salvanes (2017) on lifecycle and OVB
- Nybom (2017) on selection on gains and MTEs

Extensions

- Income processes beyond Mincer
- Multiplicity, dynamic complementarity and social returns

The Mincer regression

As noted by Mincer (1958, 1974), our simple life-cycle model of human capital accumulation implies an earnings function of the form (CCZ 4.1.1)

$$\log w(s_i) = \log w(0) + \rho s_i$$

Derivation from Compensating Differences Model:

Let the present value of the income stream be

$$V(s) = Y(s)\int_s^T e^{-rt}dt = \frac{Y(s)}{r}(e^{-rs} - e^{-rT})$$

Equilibrium across schooling levels requires:

$$\ln Y(s) = \ln Y(0) + rs + \ln\left((1-e^{-rT})/(1-e^{-r(T-s)})\right)$$

The Mincer regression

The extended model with on-the-job learning implies
the so-called **Mincer regression** (CCZ 4.1.2)

$$\log w_i = \beta_0 + \beta_1 s_i + \beta_2 x_i + \beta_3 x_i^2 + \varepsilon_i$$

where s_i is formal years of schooling and x_i is labor market experience (interpreted as on-the-job learning), and $\beta_1 > 0$, $\beta_2 > 0$ and $\beta_3 < 0$ are functions of the efficiency of HC accumulation.

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

Figure: Log earnings are linear in schooling? US data

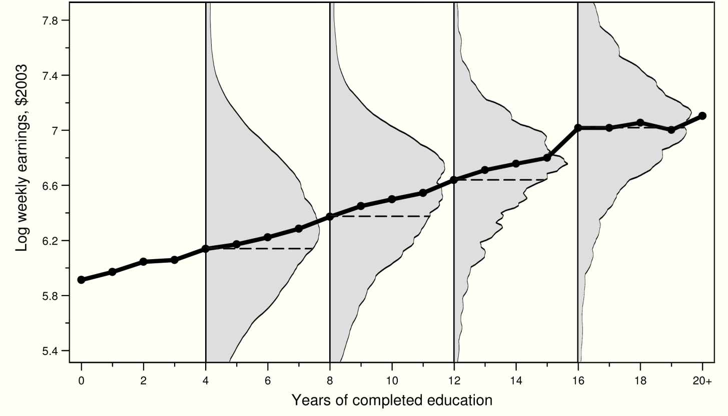
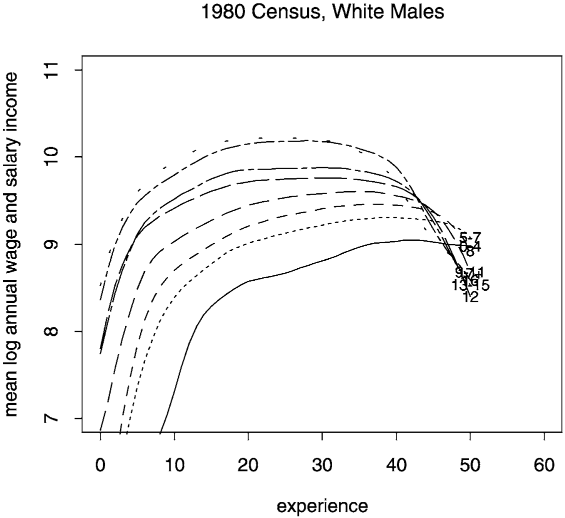


Figure 3.1.1: Raw data and the CEF of average log weekly wages given schooling. The sample includes white men aged 40-49 in the 1980 IPUMS 5 percent file.

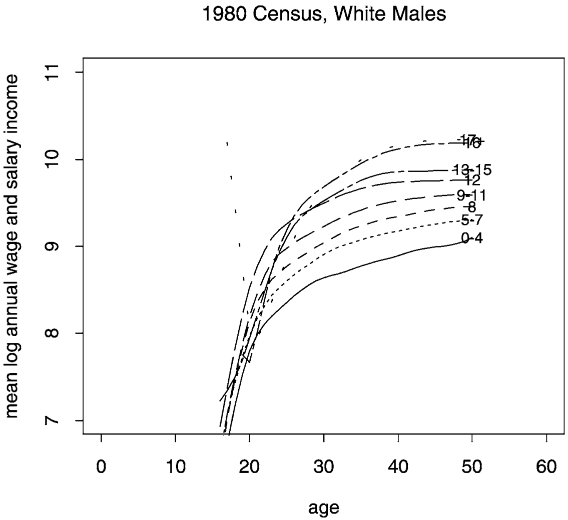
Source: Angrist and Pischke (2008), cross-sectional data.

Figure: Experience-earnings (log mean) profile



Source: Heckman, Lochner and Todd (2006)

Figure: Age-earnings (log mean) profile



Source: Heckman, Lochner and Todd (2006)

Returns to schooling

The *Mincer equation*,

$$\log w_i = \beta_0 + \beta_1 s_i + \beta_2 x_i + \beta_3 x_i^2 + \varepsilon_i \tag{4}$$

where s_i is schooling, x_i is labor market experience.

A puzzling finding:

- OLS estimates of Mincer equation imply that in the US, the returns to an additional year of schooling are ten percent or more (e.g., $\hat{\beta}_1 \approx 0.12$ in Heckman et al. 2006)
- Few other investments have such high returns. Why are not more people attending university?

Two possible answers:

1. OLS estimates of β_1 are biased (omitted variable bias)
2. Heterogeneity in returns/costs to education (sorting on gains)

Internal rate of return to schooling

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, β_1 from the Mincer regression is the IRR to schooling (see Heckman, Lochner and Todd 2006)

Agenda

Course Logistics and Requirements

Human Capital

- General and specific human capital
- HC over the life-cycle and the Ben-Porath model
- The Mincer regression and returns to schooling

Empirical Applications

- Bhuller, Mogstad and Salvanes (2017) on lifecycle and OVB
- Nybom (2017) on selection on gains and MTEs

Extensions

- Income processes beyond Mincer
- Multiplicity, dynamic complementarity and social returns

Empirical applications

Empirical applications:

1. Bhuller, Mogstad and Salvanes (2017) on returns to education, lifecycle bias and omitted variable bias (and LATE)
2. Nybom (2017) on returns to education and selection on gains (and MTE)

Agenda

Course Logistics and Requirements

Human Capital

- General and specific human capital
- HC over the life-cycle and the Ben-Porath model
- The Mincer regression and returns to schooling

Empirical Applications

- Bhuller, Mogstad and Salvanes (2017) on lifecycle and OVB
- Nybom (2017) on selection on gains and MTEs

Extensions

- Income processes beyond Mincer
- Multiplicity, dynamic complementarity and social returns

Omitted variable bias

Abstract from experience in eq. (4) and assume that ε_i partially reflects unobserved "ability" a_i, such that $\varepsilon_i = \gamma a_i + \tilde{\varepsilon}_i$ and

$$\log w_i = \beta_0 + \beta_1 s_i + \gamma a_i + \tilde{\varepsilon}_i \qquad (5)$$

where $\tilde{\varepsilon}_i$ is uncorrelated with s_i or a_i.

The probability limit of the OLS estimator of $\log w_i$ on s_i is then

$$plim\, \hat{\beta}_1 = \frac{Cov(\log w_i, s_i)}{Var(s_i)} = \beta_1 + \gamma \frac{Cov(a_i, s_i)}{Var(s_i)}$$

How is this most likely to be biased?

Omitted variable bias

How to deal with omitted variable bias? Common strategies:

(1) **Selection-on-observables approach**

- Try to control for "ability" in the Mincer regression, e.g. by including IQ scores (Grilliches 1977)

(2) **Twin approach**

- Try to difference out ability from Mincer regression

(3) **Instrumental variable approach**

- A valid IV predicts schooling (rank condition) and affects wages only via schooling (exclusion restriction).
- Examples: Date of birth interacted with compulsory schooling age; compulsory school reforms; distance to college

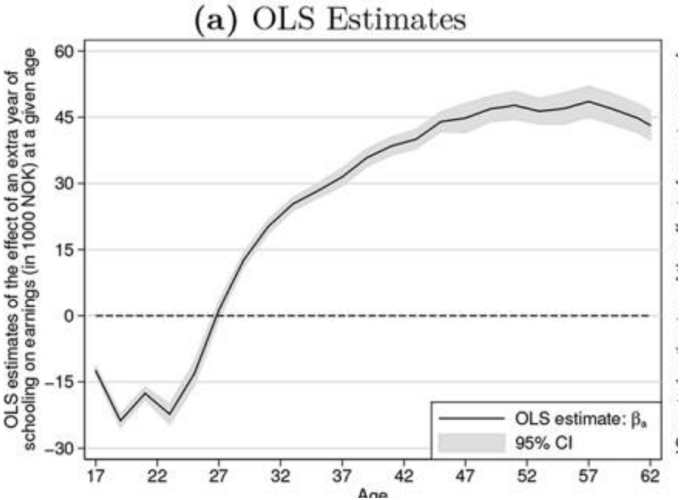
Application 1: Bhuller, Mogstad and Salvanes (2017)

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

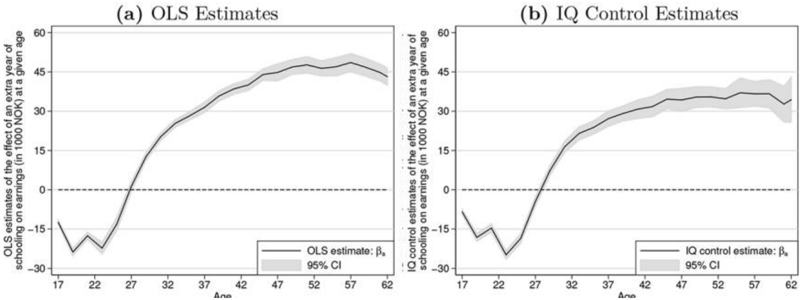
- Administrative panel data for Norway
- Full population, near career-long earning histories (1967-2014)
- Pre-tax labor income and benefits / post-tax / benefits
- Ability/IQ measures from military enlistment tests
- IV: staggered implementation of compulsory school reform

They estimate a separate Mincer regression at each age (with municipality and cohort FEs):

Bhuller, Mogstad and Salvanes (2017): OLS



Bhuller, Mogstad and Salvanes (2017): SoO approach



- Estimated returns decrease when controlling for IQ
- Suggestive of upward ability bias in OLS estimates, but the estimates do not drop by much

Bhuller, Mogstad and Salvanes (2017): IV approach

Their IV approach exploits the gradual roll-out of compulsory school reform across counties (similar reforms have been studied in the US, UK, Sweden and other countries)

- the reform increases schooling duration ("first stage")
- reform has biggest effect on those with low schooling
- municipalities are treated at different times ("event study", but that label was not yet fashionable in 2017)

Bhuller, Mogstad and Salvanes (2017): IV approach

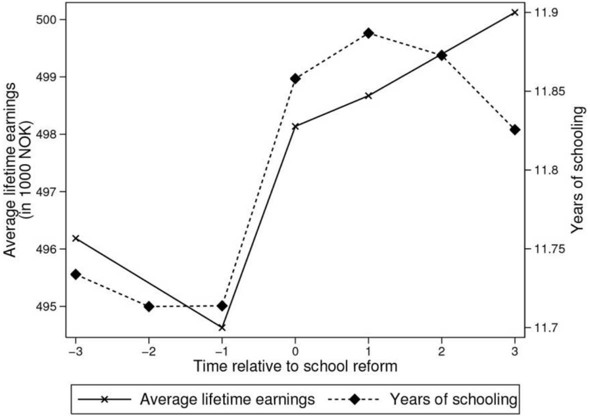
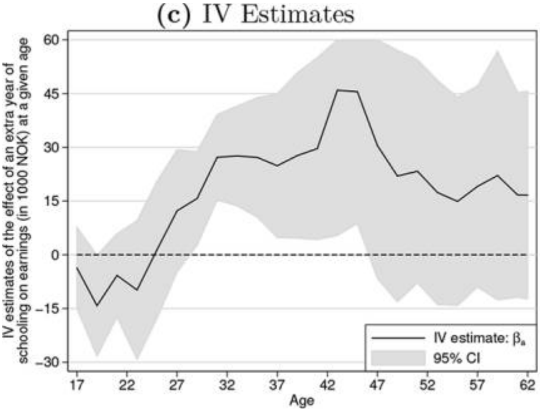


FIG. 2.—Graphical illustration of the instrumental variable (IV) approach. For each municipality, we recenter the data such that time zero is the year in which the reform was implemented. Variables are residuals from a regression on birth cohort and municipality fixed effects (adding in a common intercept).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

Bhuller, Mogstad and Salvanes (2017): IV approach



- IV estimates similar as OLS estimates, but much noisier

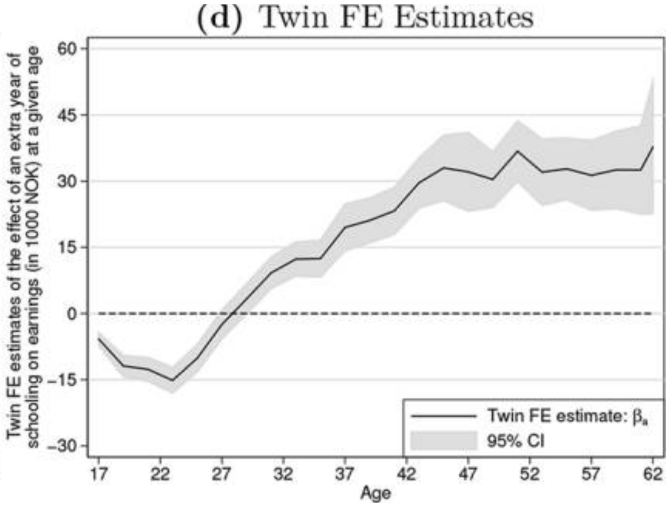
Bhuller, Mogstad and Salvanes (2017): Twin approach

- Compare difference in schooling of twins with their difference in earnings (6,434 monozygotic and dizygotic twins)
- Twins have more similar abilities than other siblings (e.g. monozygotic twins share genes).
- If they have the *same* ability then differencing a Mincer regression such as eq. (5) for twins A and B,

  $$log(w_i^A) - log(w_i^B) = \beta_1(s_i^A - s_i^B) + \gamma(a_i^A - a_i^B) + \tilde{\varepsilon}_i^A - \tilde{\varepsilon}_i^B$$

  would eliminate the ability bias (Griliches 1979)
- But does differencing eliminate more variation in the omitted variable (ability) or the regressor of interest (schooling)? In BP model schooling depends on ability ("learning efficiency")!
- Mixed evidence: Bhuller et al find no systematic differences, but Sandewall, Cesarini and Johannesson (2014) find that IQ and educational differences do correlate within twin pairs.

Bhuller, Mogstad and Salvanes (2017): Twin approach



Bhuller, Mogstad and Salvanes (2017): IRR

IRR defined 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 non-parametric bootstrap with 250 replications. FE = fixed effects.

** $p < .05$.

*** $p < .01$.

BMS further show that direct estimates from a Mincer regression are much smaller (→ Section 4).

Still a puzzle

Even after addressing omitted variable bias, returns to schooling appear larger than the market interest rate.

In fact, IV estimates of β_1 in

$$\log w_i = \beta_0 + \beta_1 s_i + \beta_2 x_i + \beta_3 x_i^2 + \varepsilon_i$$

typically *exceed* the OLS estimates (see previous slide). Potential explanations:

- Measurement error
- Invalid instruments
- Publication bias: IV estimates tend to be noisier than OLS $\rightarrow$ get published only if they are large (Ashenfelter, Harmon and Oosterbeek, 1999)
- IV identifies only a local average treatment effect (LATE)

OLS and IV weights

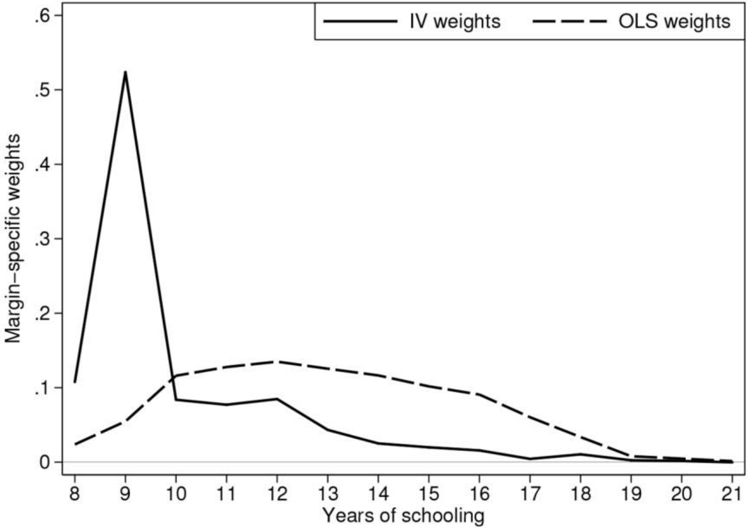


FIG. C1.—Ordinary least squares (OLS) and instrumental variable (IV) weights for every grade-specific effect.

From local average to marginal treatment effects

- Returns to schooling may vary across individuals and selection into schooling may depend on idiosyncratic gains (Carneiro, Heckman and Vytlacil, 2011)
- In particular, returns might be high for those who acquire schooling, but lower for those who do not ("selection on gains")
- IV estimator only identifies gains for "compliers" whose schooling has changed because of the instrument $\rightarrow$ LATE
- However, if we have "many instruments" we could estimate many LATEs for different groups of compliers $\rightarrow$ marginal treatment effects (Björklund and Moffitt 1987, Heckman and Vytlacil 1999, 2005)

Agenda

Course Logistics and Requirements

Human Capital

General and specific human capital

HC over the life-cycle and the Ben-Porath model

The Mincer regression and returns to schooling

Empirical Applications

Bhuller, Mogstad and Salvanes (2017) on lifecycle and OVB

Nybom (2017) on selection on gains and MTEs

Extensions

Income processes beyond Mincer

Multiplicity, dynamic complementarity and social returns

Application 2: 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)

Continuous instrument, heavy data requirements to estimate the MTE semi-parametrically. Instruments used in Nybom (2017):

1. Distance to closest college
2. Short-run fluctuations in local labor market conditions

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 $\mathbf{A}$			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 a binary choice indicator

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

Assume potential outcomes Reminder PO model 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

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

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

Instrument Z_I is assumed to be valid: affects college decision (rank condition) but not potential outcomes (exclusion restriction).

However, there is unobserved heterogeneity. Let U_{Si} represent (the quantiles of) an idiosyncratic latent "resistance" (or "distaste") to college, such that

$P_i(x,z) > U_{Si}$ 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

MTEs are defined as

$$MTE(X_i = x, U_{Si} = u_S) = E[Y_{1i} - Y_{0i} | X_i = x, U_{Si} = u_s] \qquad (6)$$

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

MTE may vary with X_i or U_{Si} (i.e. $U_1 - U_0$ may correlate with U_S)

- Slope of MTE with respect to X: "observed heterogeneity"
- Slope of MTE with respect to U_s: "unobserved heterogeneity"

MTE: Estimation

MTE can be estimated parametrically or semi-parametrically.

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

1. Estimate $E[Y_i|X_i = x, P(Z_i) = p]$ semi-parametrically for all values of $x, p = u_S$
2. Compute derivative with respect to p, as (Carneiro et al. 2011)

$$MTE(X_i = x, U_{Si} = u_S) = \left.\frac{\partial E[Y_i|X_i = x, P(Z_i) = p]}{\partial p}\right|_{p=u_s} \quad (7)$$

This MTE can then be aggregated for certain subgroups to compute the ATE, ATT or other treatment effects of interest.

MTE: Continuous instrument

Local IV approach identifies the MTE under minimal assumption. However, we require 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 push gradually "all types" (U_S) into college

See figure (next slide)

MTE: Continuous instrument

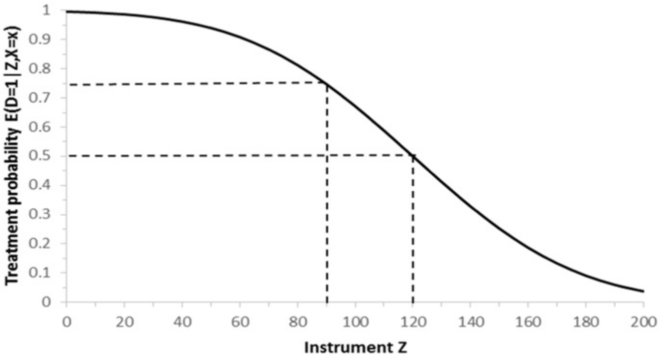


Fig. 1. Treatment probability as a function of a continuous instrument. *Notes:* Based on hypothetical data, the figure shows the effect of a continuous instrument Z on the probability of treatment in a sample with fixed covariates ($E[D = 1, Z, X = x]$). For example, the horizontal axis could represent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

Nybom (2017): Semi-parametric estimates of MTE

Figure: Returns to a Year of College

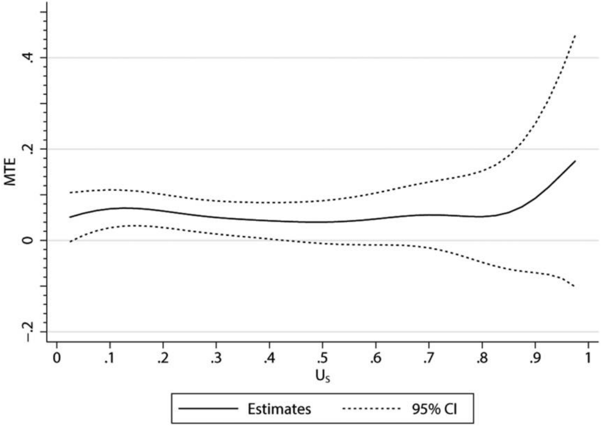


FIG. 3.—Marginal treatment effect (MTE) by U_S estimated by semiparametric local instrumental variable analysis. This figure shows point estimates and 95% confidence bands of the MTE from the semiparametric model in equation (5). The model is estimated using the local quadratic regression procedure described in Section II. All estimates are conditioned on mean values of **X** and **A**. Standard errors are bootstrapped (1,000 replications). CI = confidence interval.

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 decreases in U_S at lower values of U_S (selection on gains)
- MTE increasing in U_S at higher values of U_S (*negative* selection on gains)

Heterogeneity in Returns to Observed Characteristics

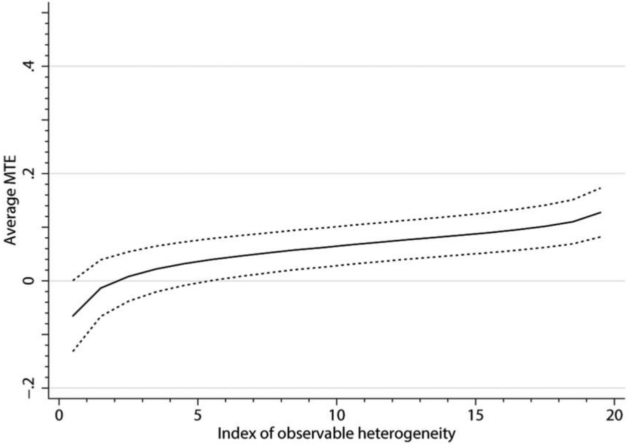


FIG. 4.—Average marginal treatment effect (MTE) by total observed heterogeneity. This figure shows the average MTE with 95% confidence bands across the index of observed heterogeneity. The index is computed by estimating $\mathbf{X}(\delta_1 - \delta_0) + \mathbf{A}(\gamma_1 - \gamma_0)$ for each individual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

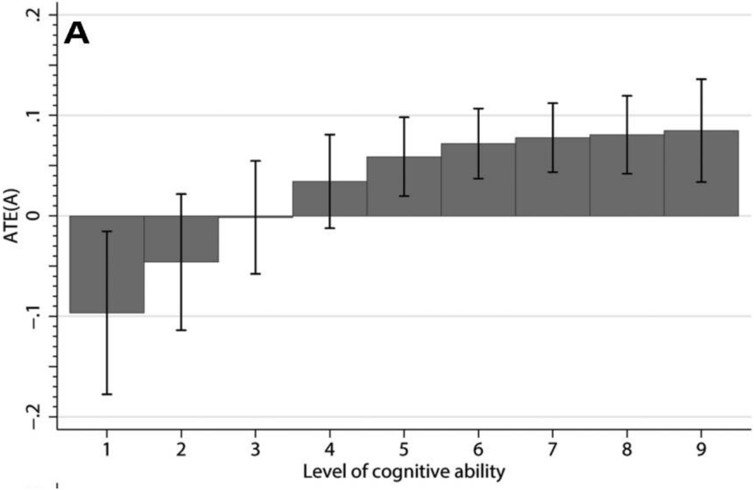
Heterogeneity in Returns to Observed Characteristics

Findings:

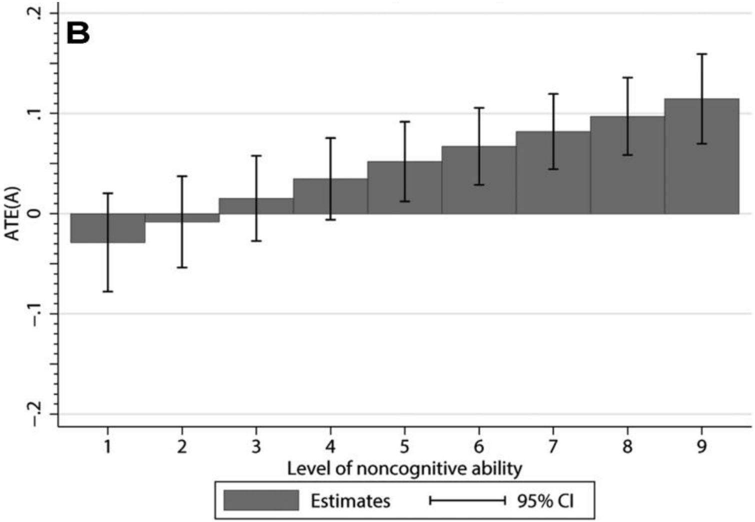
- Substantial variation in lifetime earnings returns with respect to observable characteristics
- Lifetime returns are 20 percentage points higher for those with high cognitive and non-cognitive skills compared to those with the least favorable combination of characteristics
- Observed heterogeneity much more important than unobserved heterogeneity

Quality of observables might partly explain difference in results compared to Carneiro, Heckman and Vytlacil (2011).

Heterogeneity in Returns to Observed Characteristics



Heterogeneity in Returns to Observed Characteristics



Nybom (2017): Summary

Main findings of Nybom (2017):

- College attendance is based on idiosyncratic gains (ATT > ATU), but the differences are small and ATU remains positive
- Returns to college increase steeply with cognitive and non-cognitive abilities
  - 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
- Points to complementarities between education and abilities

Agenda

Course Logistics and Requirements

Human Capital

- General and specific human capital
- HC over the life-cycle and the Ben-Porath model
- The Mincer regression and returns to schooling

Empirical Applications

- Bhuller, Mogstad and Salvanes (2017) on lifecycle and OVB
- Nybom (2017) on selection on gains and MTEs

Extensions

- Income processes beyond Mincer
- Multiplicity, dynamic complementarity and social returns

Agenda

Course Logistics and Requirements

Human Capital

- General and specific human capital
- HC over the life-cycle and the Ben-Porath model
- The Mincer regression and returns to schooling

Empirical Applications

- Bhuller, Mogstad and Salvanes (2017) on lifecycle and OVB
- Nybom (2017) on selection on gains and MTEs

Extensions

- Income processes beyond Mincer
- Multiplicity, dynamic complementarity and social returns

Beyond Mincer

Beyond Mincer and life-cycle bias:

- Estimates of the returns to education are sensitive to age at measurement, in line with the BP model and Mincer equation.
- A similar "life-cycle bias" may occur in many other contexts ($\rightarrow$ extreme example: intergenerational studies)
- But what about income dynamics *within* education groups or conditional on other individual characteristics?

Such questions lead us from the Mincer equation to the broader literature on income processes.

Income process literature

The income process literature models the lifecycle profile of income and related variables.

Interesting debate on whether the shape of income profiles is better described by HIP or RIP process:

- Restricted income profiles (RIP): individuals are subject to *permanent* income shocks ($\rightarrow$ random walk)
- Heterogeneous income profiles (HIP): individuals face individual-specific (and predictable?) income profiles

The distinction matters. In particular, are "permanent income shocks" as identified in the RIP process really unexpected "shocks"?

Mello, Nybom and Stuhler (2022)

Mello, Nybom and Stuhler (2022) study the shape of age-earning profiles in Swedish data. Two parts:

1. Study lifecycle dynamics of child income by parental income
2. Develop a "life-cycle estimator" for intergenerational mobility

We focus on the first step here:

- Key finding: Steeper earnings growth for children from high-income families, even conditional on their own education
- Consistent with HIP but not RIP; "permanent shocks" in the RIP model may not be "shocks" at all.
- Recall that in the Ben-Porath model, earnings growth depends on "learning speed". Abilities affect earning levels *and* slopes?

Table: Heterogeneity in Income Growth by Parental Income (MNS 2022)

	(1)	(2)	(3)	(4)	(5)	(6)
Log (Father's Income)/100						
x Age 25-30	13.386***	3.675***	3.493***	5.831***	1.785***	1.872***
	(0.252)	(0.227)	(0.263)	(0.237)	(0.225)	(0.262)
x Age 30-35	6.813***	1.503***	0.823***	3.524***	1.164***	0.633**
	(0.202)	(0.204)	(0.231)	(0.205)	(0.204)	(0.233)
x Age 35-40	3.188***	0.139	0.065	1.184***	-0.117	-0.098
	(0.193)	(0.198)	(0.226)	(0.199)	(0.200)	(0.230)
x Age 40-45	0.738***	-0.476*	-0.272	0.353	-0.267	-0.133
	(0.181)	(0.188)	(0.216)	(0.188)	(0.191)	(0.220)
x Age 45-50	-0.543**	-0.123	-0.277	-0.075	0.035	-0.224
	(0.177)	(0.183)	(0.211)	(0.185)	(0.187)	(0.216)
x Age 50-55	-2.463***	-0.873***	-0.670**	-1.308***	-0.539**	-0.369
	(0.174)	(0.179)	(0.206)	(0.183)	(0.184)	(0.212)
Education x Age		X	X		X	X
Occupation x Age				X	X	X
Skill scores x Age			X			X
Demographics x Age						X
N	950263	950125	744286	919473	919346	720085
R-sq	0.053	0.117	0.122	0.095	0.132	0.137

Agenda

Course Logistics and Requirements

Human Capital

- General and specific human capital
- HC over the life-cycle and the Ben-Porath model
- The Mincer regression and returns to schooling

Empirical Applications

- Bhuller, Mogstad and Salvanes (2017) on lifecycle and OVB
- Nybom (2017) on selection on gains and MTEs

Extensions

- Income processes beyond Mincer
- Multiplicity, dynamic complementarity and social returns

Related topics

Other important topics that for time reasons we cover only briefly:

1. Multiplicity of skills, cognitive vs. non-cognitive skills
2. Early-childhood investments and dynamic complementarity
3. Private vs. social returns to human capital

Deming (2017)

Human capital may consist of multiple distinct skills. One useful distinction is between cognitive and non-cognitive skills.

Motivational fact: no increase in returns to to cognitive skills since 2000 despite skill-biased technical change since 1990s.

$\rightarrow$ Slowdown in technical progress?
$\rightarrow$ Technical progress substituting cognitive skills?

Evidence by Deming (2017):

- Labor market increasingly rewards *social* skills
- In 1980-2012, share of U.S. jobs 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
- To explain these findings, Deming proposes a model in which there is "trade" in tasks and social skills reduce trade costs

Deming (2017)

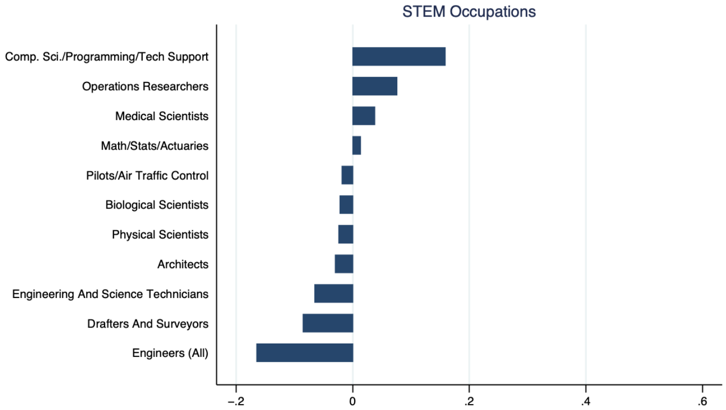


FIGURE I

Change in Relative Employment for Cognitive Occupations, 2000-2012

Deming (2017)

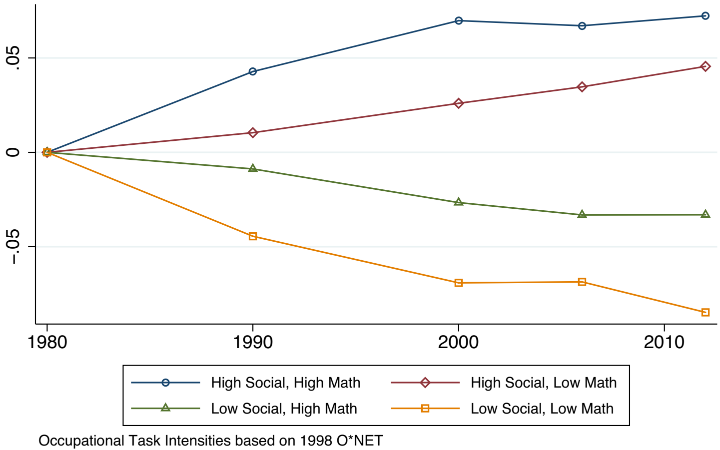


FIGURE IV

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

Deming (2017)

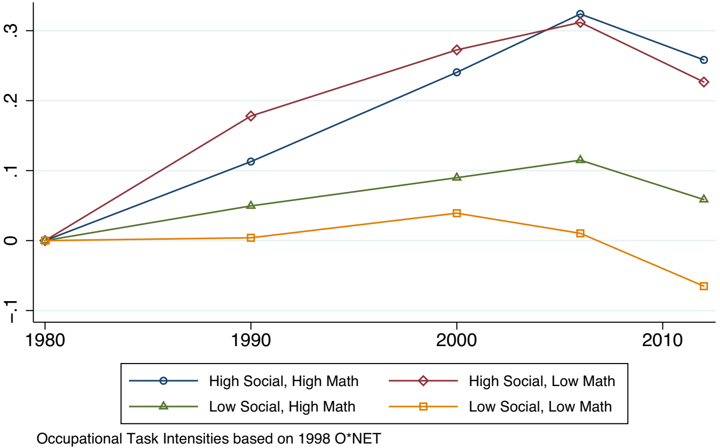


FIGURE V

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

Deming (2017)

Main estimating equation:

$$\log wage_{ijt} = \beta_1 COG_i + \beta_2 SS_i + \beta_3 COG_i \times SS_i + \gamma X_{ijt} + \delta_j + \zeta_t + \varepsilon_{ijt}$$

- Controls: race, gender, region, age (j) and year (t)
- Returns to both types of skills are positive
- β_3 is posiitve $\Rightarrow$ complementarity between skills

Dynamic complementarity

At what age are HC investments most effective?

- Some evidence 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 later stages are complementary to earlier investments + self-productivity = dynamic complementarity
- Policy implication: Investments in disadvantaged young children can be both fair and efficient, while investments in disadvantaged adolescents might be fair but less efficient.

Returns to HC investment over age

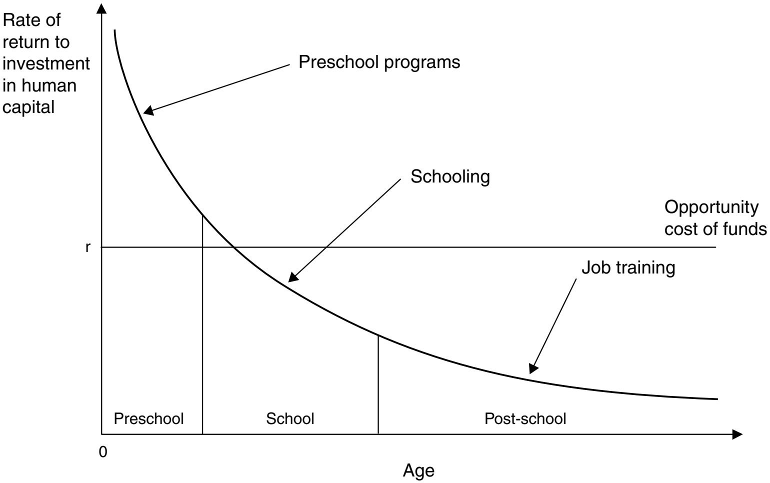


FIGURE 4.20
Rate of return to investment in human capital by age.

Social returns to education

Education might generate large social returns (returns to society as a whole, minus private returns):

- knowledge externalities
- education appears to have a negative effect on crime
- ...

But in principle, social returns could also be negative:

- Example: signaling model by Spence (1973)

Social returns are difficult to estimate:

- Example: Moretti (2004) estimates spillovers from share of college graduates on workers in a city
- Finds large positive externality, in particular on less educated workers

Readings

Readings:

- Bhuller, Mogstad and Salvanes (2017): understand main research designs and source of "life-cycle bias"
- Heckman, Lochner and Todd (2006): Skim through, in particular Sections 1, 3, 4, 6, 7, 11
- Nybom (2017): Definition and intuition of MTE

Appendix

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)

$$\Omega = \int_0^T e^{-rt}\left[Ah(t)(1-\sigma(t))\right]dt \tag{8}$$

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

$$\begin{aligned}
\frac{\partial \Omega}{\partial \sigma(t)} &= \underbrace{-e^{-rt}Ah(t)}_{\text{opportunity costs}} + \int_0^T e^{-rz}A\left[1-\sigma(z)\right]\frac{\partial h(z)}{\partial \sigma(t)}dz \\
&= \underbrace{-e^{-rt}Ah(t)}_{\text{opportunity costs}} + \int_t^T e^{-rz}A\left[1-\sigma(z)\right]\theta h(z)dz
\end{aligned} \tag{9}$$

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

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

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 (9)

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

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 (9)

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

Back

Reminder: Potential outcome model

The *"Potential outcome or "Rubin causal" 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 (ATE):

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

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):

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

for compliers whose treatment status has switched because of the instrument → a "local" effect. Back