Lecture 8: Returns to Education and Fixed Effects

Hammad Shaikh

Pecuniary returns to education

- Education investment is an important decision made by parents and teenagers

- Educational investments are associated with costs

K-12 funding + subsidized college

- Do the benefits exceed the costs?

Benefits ↑tax, ↓Crime, ↓Welfare

- Why is it important to know the causal effect of years of schooling on earnings from a policy perspective?

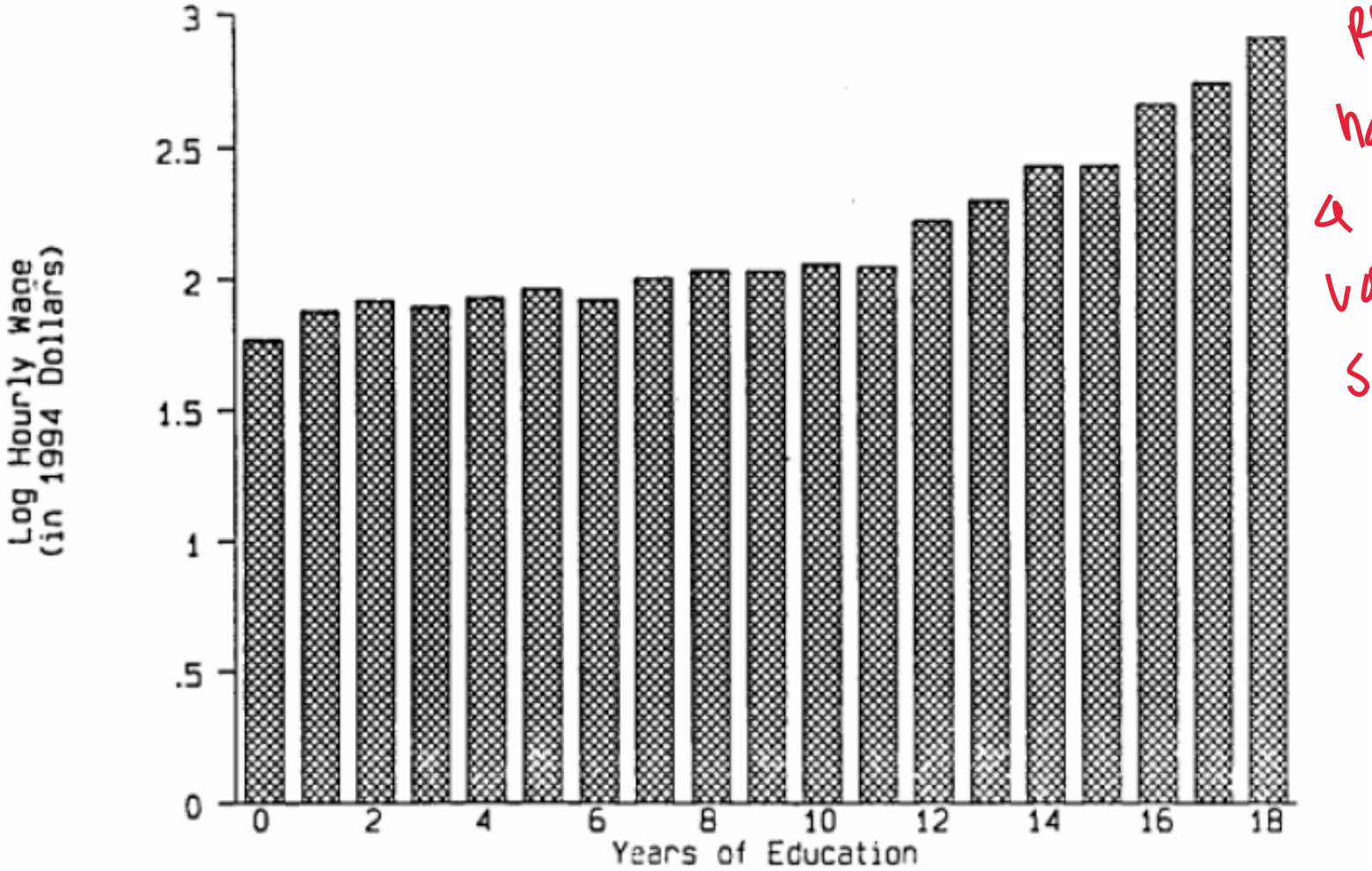
Conduct cost-benefit analysis

Education and wages

1) Signaling says more educated people signal higher ability

2) Human capital says more educated people have a more valuable skills

Figure 1. Average Log Hourly Wage by Years of Education



Causality review for returns to education

- Why isn't it sufficient to regress income on education?

$$Wage_i = \beta_0 + \beta_1 Yrs Educ_i + \varepsilon_i$$

- Shouldn't the coefficient on years of education tell us the marginal returns to schooling?
  - Education is likely to be correlated with ability, which will also be independently correlated with income

$$Cov(\varepsilon_1, YrsEduc_i) \neq 0 \text{ is problem}$$

  - Education is likely to be correlated with family socioeconomic status, which is also likely to be correlated with future income for other reasons

Ideal experiment for measuring education returns

- If we had infinite resources and ethics weren't an issue, what would be the best way to estimate the returns to education?

- Randomly assign some students to get more education than others
  - Outcome: measure their wages 20 years later

- Why does this solve the problem of estimating the causal effect of education on income?

Since YrsEduc randomly assigned
$\Rightarrow \quad \varepsilon \perp$ YrsEduc

Natural experiments for measuring education returns

- Short of a randomized experiment, what can we do?
- Option 1: We can try to “control" for the differences in ability and socio-economic status
- Option 2: We can look for “natural experiments" that appear to have randomly induced some people to get more education than others

↳ Use instrumental variable, regression discont. and diff. in diff.

Ashenfelter and Krueger (1994) Introduction

- Estimate returns to education using option 1 (control for differences across more and less educated people)

  Control for family background

- Collect wage and education data on about 300 identical twins

  Twins festival

- Strategy: compare wages of identical twins with different levels of education

  Use only "within family" variation

Ashenfelter and Krueger (1994) Model

- y_{ji} is log wage for twin $j \in \{1, 2\}$ in family i

Family charac. SES, Educ, Occupation, etc

- $y_{ji} = \alpha X_i + \beta YrsEduc_{ji} + \eta_{ji}$ for twin j

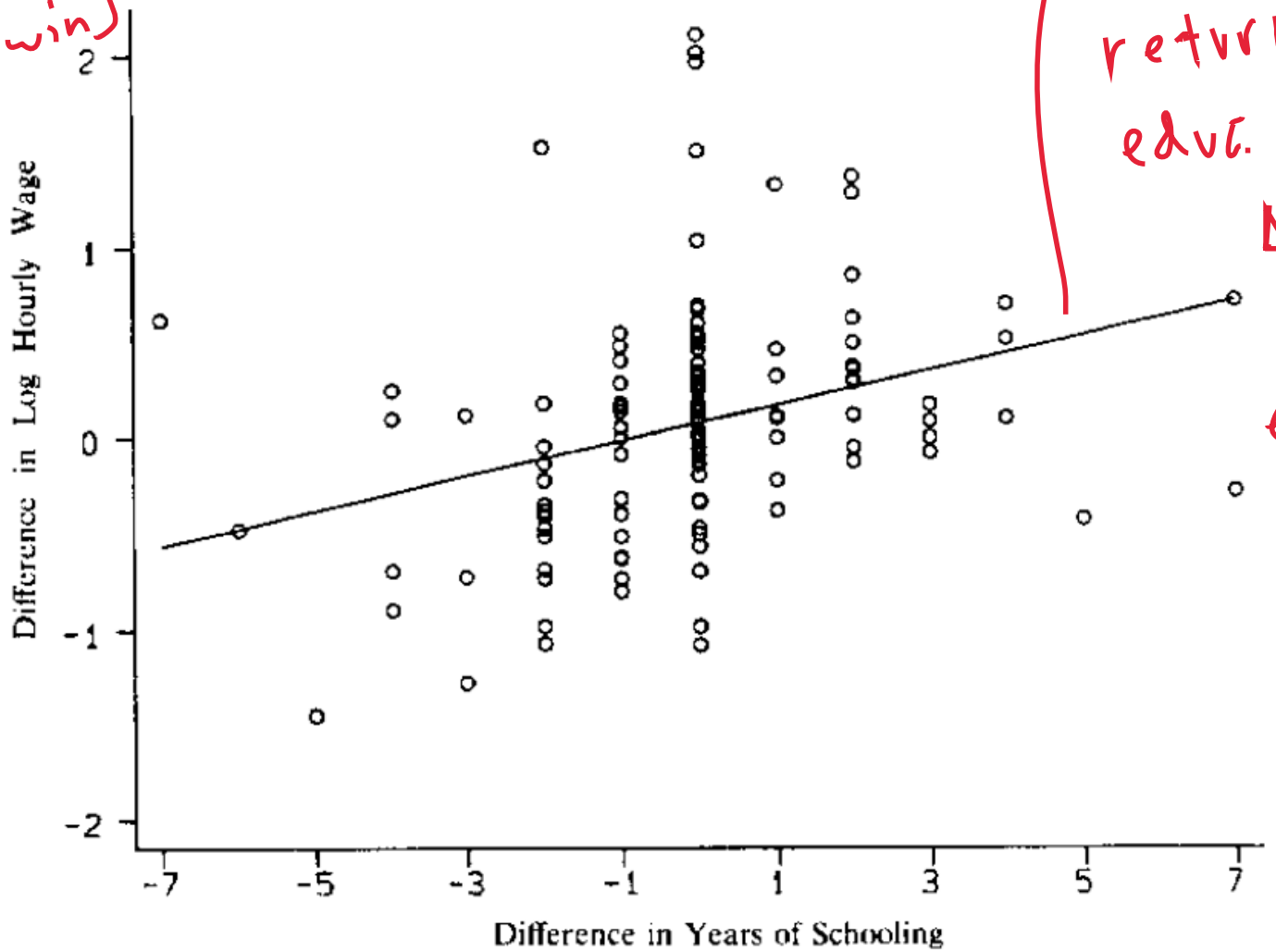
η_{ji} = unobs. family + unobs. indiv. charac.

- Problem: $\eta_{ji} = \mu_i + \epsilon_{ji}$ and $Corr(\mu_i, YrsEduc_{ji}) \neq 0$

- Solution: $y_{2i} - y_{1i} = \beta(YrsEduc_{2i} - YrsEduc_{1i}) + (\epsilon_{2i} - \epsilon_{1i})$

Assume Δ YrsEduc $\perp \Delta\epsilon \Rightarrow$ unbiased estimate for β

Returns to schooling for identical twins



$\Delta \log(\text{wage})$ across twins

slope $\hat{b}_1$ is the causal returns to educ. if Δ Yrs Educ is exogenous

Δ Yrs Educ. across twins

Measurement error problem

$\varepsilon \sim N(0,1)$

- $\widetilde{X}$ is measured with error if $\widetilde{X} = X + \epsilon$

$\widetilde{X}$ = survey yrs educ, X = Actual yrs educ

only observe $\widetilde{X}$

- Measurement error is a type of endogenity

Biases slope estimate towards 0

- $Y = \beta_0 + \beta_1 \widetilde{X} + \epsilon$ then $E(\widehat{\beta}_1) = \lambda\beta_1$ where $\lambda < 1$

$\beta_1 \neq 0 \Rightarrow E[\hat{\beta}_1] < \beta_1 \Rightarrow \hat{\beta}_1$ downwards bias

- Solution: use instrumental variable Z for $\widetilde{X}$

ASK use Z = j^{th} Educ. reported by twin i

Ashenfelter and Krueger (1994) Results

Comparing twins:

Extra yr of edu ↑ wages by 9.2%, on avg.

TABLE 3—ORDINARY LEAST-SQUARES (OLS), GENERALIZED LEAST-SQUARES (GLS), INSTRUMENTAL-VARIABLES (IV), AND FIXED-EFFECTS ESTIMATES OF LOG WAGE EQUATIONS FOR IDENTICAL TWINS[a]

Variable	OLS (i)	GLS (ii)	GLS (iii)	IV[a] (iv)	First difference (v)	First difference by IV (vi)
Own education	0.084 (0.014)	0.087 (0.015)	0.088 (0.015)	0.116 (0.030)	0.092 (0.024)	0.167 (0.043)
Sibling's education	—	—	−0.007 (0.015)	−0.037 (0.029)	—	—
Age	0.088 (0.019)	0.090 (0.023)	0.090 (0.023)	0.088 (0.019)	—	—
Age squared (÷100)	−0.087 (0.023)	−0.089 (0.028)	−0.090 (0.029)	−0.087 (0.024)	—	—
Male	0.204 (0.063)	0.204 (0.077)	0.206 (0.077)	0.206 (0.064)	—	—
White	−0.410 (0.127)	−0.417 (0.143)	−0.424 (0.144)	−0.428 (0.128)	—	—
Sample size:	298	298	298	298	149	149
R^2:	0.260	0.219	0.219	—	0.092	—

estimate after correcting for measurement error

Fixed effects approach

"family fixed effects"

- Comparisons like this is called "fixed effects" because we are making comparisons within a fixed group

Comparing returns to educ. for twins within family

- Fixed effect regression gives causal estimate if education is not correlated with ability or socioeconomic status within the fixed group

$\Delta\varepsilon \perp \Delta Y$ vs educ within families

Ashenfelter and Krueger (1994) Summary

- An extra year of school has at least 9% increase in wages on average

- Potential short comings of paper:
  - Assumption: Differences in educational attainment across twins is random
    ↳ Violated if higher ability twin recieves more educ. & parental investment
  - Small sample size
    ↳ $n = 298$
  - Selection into "Twins Festival" where sample is collected