14.771: Private and Social Returns to Education

Esther Duflo

MIT

Estimating the returns to human capital investment

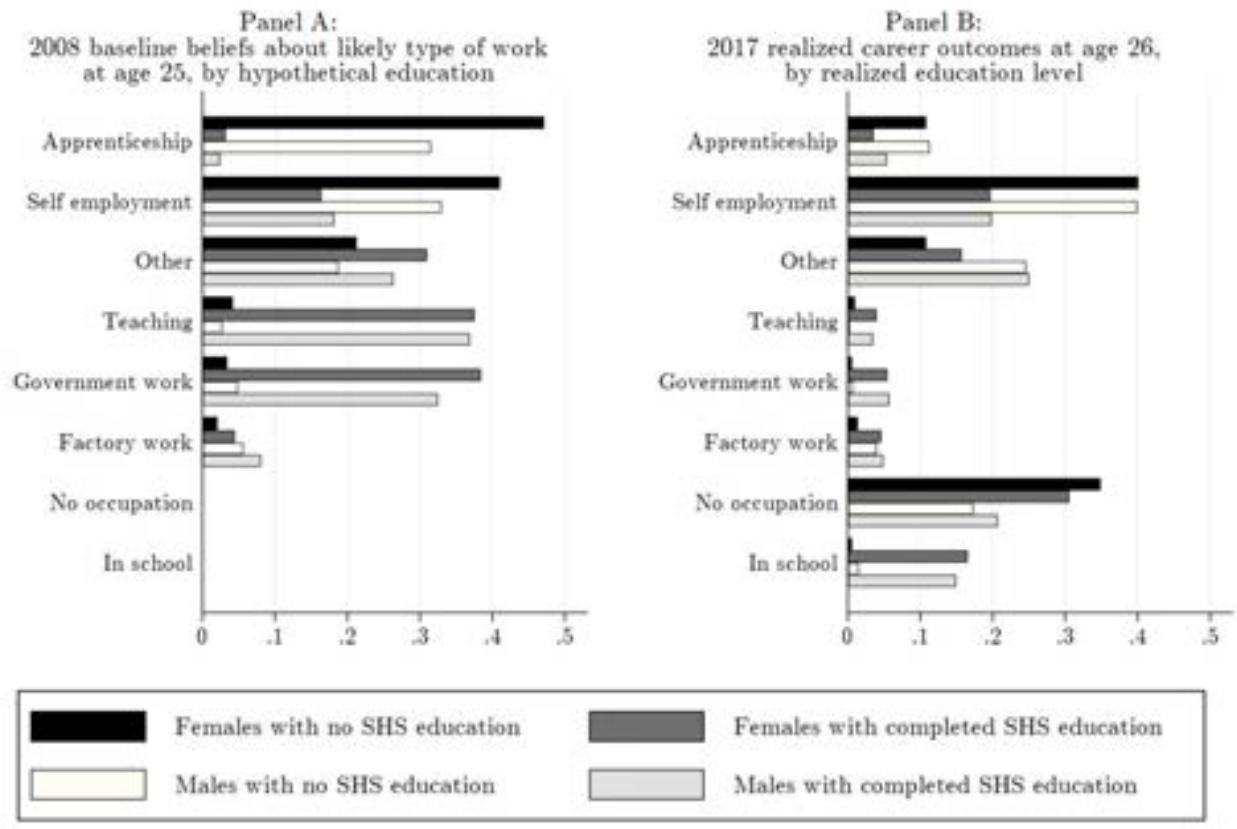
- Plenty of evidence of correlation of human capital and earnings.
- ▸ Education , ▸ health
- And other benefits of both health and education which could be represented (with a stretch) as forms of consumption (political attitudes, etc)
- There is also evidence of interactions between different form of human capital
  - Educated people (and their children) are healthier
  - Healthier children miss fewer days of school, and earn more as adults (deworming).
- What could be the problem with the problem with interpreting these correlations as the causal effect of human capital –which is what parents should care about when investing?
- What do authors mention specifically in the Ghana and Indonesia paper?

Private vs Social returns

- Private returns: for any individual, how much more money to they make with or without education
- Social returns: What is the value for society of an individual being more educated.
- Private and social returns are likely to differ for all sorts of reasons
  - Equilibrium effects (competition on the labor markets, for other educated or for uneducated workers).
  - An extreme version of this is rent seeking (see Ghana paper): education is a ticket to rationed jobs, but if there are no more rationed jobs, there is little you gain
  - Positive impacts on other workers
  - Impacts in other spheres (health, politics, etc.)

Parent's perception of returns and their sources

Figure 1: Type of work, by education level: Baseline Expectations vs. Realizations



Notes: Data from 2008 in-person baseline survey of participants (Panel A) and 2017 phone survey (Panel B). SHS stands for Senior High School. In Panel A, respondents (aged 17 on average at the time) were asked in 2008: "If you never go to SHS or continue any other higher education in the future, what types of work do you think you would do when you are 25 years old?" and "Imagine that you complete Senior High School in the future, what types of work do you think you would do when you are 25 years old?" In Panel B, data from the 2017 phone survey on the realized career outcomes of students who did and did not complete SHS is shown. We plot answers separately by respondent gender, pooling treatment and control groups.

Ghana: An example of random assignment

- What is randomly assigned in Ghana?
- In what sample?
- What can we confidently estimate?

	(1) Total years of education to date (2019)	(2) Total years of SHS to date (2019)	(3) Completed SHS (2019)	(4) Completed TVI (2019)	(5) Completed tertiary (2019)	(6) Ever enrolled in tertiary program (2019)	(7) Currently enrolled in tertiary program (2019)	(8) Ever enrolled in tertiary program (2020)
Panel A: All								
Treatment	1.241	1.251	0.272	-0.020	0.035	0.044	0.010	0.047
	(0.104)	(0.079)	(0.022)	(0.006)	(0.015)	(0.018)	(0.011)	(0.020)
P-value	0.000	0.000	0.000	0.001	0.019	0.016	0.346	0.019
Comparison mean	11.387	1.842	0.436	0.029	0.087	0.154	0.049	0.179
N	1924	1925	1952	1952	1952	1951	1951	1740
Panel B: Female								
Treatment	1.313	1.208	0.258	-0.005	0.040	0.077	0.029	0.088
	(0.155)	(0.119)	(0.032)	(0.008)	(0.020)	(0.025)	(0.015)	(0.028)
P-value	0.000	0.000	0.000	0.573	0.048	0.002	0.059	0.002
Comparison mean	11.030	1.627	0.389	0.017	0.078	0.126	0.035	0.146
N	968	968	986	986	986	986	986	856
Panel C: Male								
Treatment	1.141	1.273	0.281	-0.035	0.030	0.010	-0.009	0.007
	(0.138)	(0.104)	(0.031)	(0.009)	(0.022)	(0.026)	(0.016)	(0.029)
P-value	0.000	0.000	0.000	0.000	0.162	0.709	0.588	0.804
Comparison mean	11.758	2.065	0.485	0.041	0.096	0.184	0.065	0.210
N	956	957	966	966	966	965	965	884

	(1) Total years of education (2013)	(2) Total cognitive score (2013)	(3) Political knowledge score (2013)	(4) Media engagement (radio, newspaper, TV, internet) (2013)	(5) Knows how to use internet (2013)	(6) Has a bank account (2013)	(7) ICT/Social media adoption index (2016)	(8) Uses fertilizer (if in farming) (2017)	(9) Used internet in the past month (2019)
Panel A: All									
Treatment	1.191	0.157	0.095	0.060	0.086	0.058	0.062	-0.024	0.059
	(0.077)	(0.046)	(0.046)	(0.025)	(0.047)	(0.023)	(0.037)	(0.037)	(0.024)
P-value	0.000	0.001	0.040	0.018	0.069	0.011	0.095	0.527	0.013
Comparison mean	10.787	-0.000	0.000	-0.020	0.000	0.314	-0.133	0.471	0.493
N	2064	1983	1981	1981	1983	1984	1995	769	1950
Panel B: Female									
Treatment	1.186	0.194	0.075	0.074	0.050	0.098	0.090	0.020	0.076
	(0.114)	(0.069)	(0.058)	(0.032)	(0.058)	(0.031)	(0.054)	(0.057)	(0.033)
P-value	0.000	0.005	0.192	0.023	0.386	0.001	0.095	0.720	0.023
Comparison mean	10.575	-0.175	-0.381	-0.165	-0.333	0.236	-0.416	0.410	0.402
N	1036	1002	1001	1001	1001	1002	1007	337	985
Panel C: Male									
Treatment	1.183	0.113	0.094	0.035	0.101	0.016	0.016	-0.059	0.033
	(0.101)	(0.059)	(0.061)	(0.037)	(0.067)	(0.033)	(0.045)	(0.050)	(0.033)
P-value	0.000	0.054	0.126	0.347	0.133	0.630	0.724	0.238	0.310
Comparison mean	11.006	0.183	0.397	0.131	0.346	0.396	0.162	0.522	0.590
N	1028	981	980	980	982	982	988	432	965

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Worked for pay in past 6 months (2019)	Has wage contract with employer (2019)	Job with benefits (2019)	Public sector employee (2019)	Lives in urban area (2019)	Self-employed (2019)	Total earnings in past 6 months (2019)	Could not cope with 200 GHX emergency (2019)
Panel A: All								
Treatment	0.011	0.039	0.030	0.019	-0.015	-0.029	37.123	-0.027
	(0.021)	(0.015)	(0.015)	(0.013)	(0.015)	(0.020)	(93.450)	(0.017)
P-value	0.589	0.008	0.052	0.157	0.330	0.153	0.691	0.117
Comparison mean	0.730	0.084	0.099	0.077	0.123	0.245	1456.217	0.161
N	1952	1951	1951	1952	1921	1952	1915	1951
Panel B: Female								
Treatment	0.033	0.041	0.020	0.041	-0.029	-0.012	35.794	-0.044
	(0.033)	(0.019)	(0.019)	(0.019)	(0.021)	(0.031)	(108.464)	(0.024)
P-value	0.314	0.032	0.283	0.031	0.152	0.683	0.741	0.070
Comparison mean	0.602	0.063	0.075	0.063	0.119	0.287	951.456	0.176
N	986	986	986	986	973	986	972	986
Panel C: Male								
Treatment	-0.020	0.035	0.037	-0.003	-0.001	-0.042	-12.740	-0.009
	(0.024)	(0.023)	(0.024)	(0.019)	(0.023)	(0.026)	(145.790)	(0.024)
P-value	0.405	0.119	0.126	0.874	0.959	0.106	0.930	0.718
Comparison mean	0.864	0.106	0.125	0.092	0.128	0.201	1993.862	0.146
N	966	965	965	966	948	966	943	965

Table 7: Labor Market Outcomes during COVID Crisis

	(1) Worked for pay in past 6 months (2020)	(2) Has wage contract with employer (2020)	(3) Job with benefits (2020)	(4) Public sector employee (2020)	(5) Lives in urban area (2020)	(6) Self-employed (2020)	(7) Total earnings in past 6 months (2020)	(8) Total earnings April (2020)	(9) Coeff. of variation of monthly earnings (if > 0) (GHX) (2020)
Panel A: All									
Treatment	0.045	0.047	0.005	0.015	-0.019	-0.066	62.606	30.743	-0.776
	(0.020)	(0.016)	(0.015)	(0.014)	(0.016)	(0.023)	(125.938)	(25.424)	(3.910)
P-value	0.027	0.003	0.732	0.303	0.238	0.004	0.619	0.227	0.843
Comparison mean	0.760	0.081	0.102	0.082	0.124	0.342	1808.426	252.299	76.538
N	1737	1730	1730	1735	1714	1733	1672	1713	1251
Panel B: Female									
Treatment	0.057	0.067	0.031	0.039	-0.039	-0.068	262.979	67.620	-10.473
	(0.034)	(0.021)	(0.019)	(0.020)	(0.022)	(0.034)	(149.376)	(24.744)	(6.201)
P-value	0.092	0.001	0.107	0.046	0.074	0.048	0.079	0.006	0.092
Comparison mean	0.632	0.050	0.058	0.056	0.125	0.386	1008.077	115.131	89.819
N	856	853	853	855	846	853	826	843	513
Panel C: Male									
Treatment	0.028	0.027	-0.022	-0.009	0.001	-0.063	-167.123	-10.776	5.911
	(0.021)	(0.024)	(0.024)	(0.021)	(0.024)	(0.031)	(191.467)	(42.674)	(5.024)
P-value	0.179	0.252	0.359	0.660	0.964	0.042	0.383	0.801	0.240
Comparison mean	0.887	0.112	0.146	0.108	0.124	0.298	2607.316	388.009	67.376
N	881	877	877	880	868	880	846	870	738
P-val male=fem	0.463	0.234	0.097	0.097	0.256	0.849	0.089	0.128	0.037

Notes: See Table 2 notes. 2020 survey was administered over the phone (no in-person tracking) between May 19 and September 25 2020.

Discussion

- What is pretty cool about this experiment?
- But is it really what we want here?
- Not quite!! This is the effect of a scholarship. Not the effect of education.

Randomized evaluation as an instrumental Variable

- The question: How much does education improve earnings (or test scores or health...)?
- Notation, assume earnings can be written as:

$$Y_i = \alpha + \beta S_i + \epsilon_i$$

  where S_i is the years of schooling for individual i, and Y_i is earnings
- Note that this formulation assumes that the effect of education is the same for all people, which is not an assumption we will continue to make below: we also have some results on how to estimate a relationship where we don't make this assumption, but we will not cover them now)

Randomized Scholarship

- We have a potential instrument in Ghana. Scholarship were randomly assigned to students who qualified for secondary school on a basis of a competitive test scores but had not yet joined.
- Let Z_i be a dummy variable equal to 1 if one is assigned to the treatment group (and were therefore offered the scholarship), 0 otherwise.
- Getting scholarship increases the probability to ever enroll in high school by about 25 pp
  - most kids but not all enroll with scholarship
  - some kids don't enroll even without the scholarship
  - non compliance both ways

Combining the two: an instrumental variable estimate of the effect of going to school on later outcomes

Effect of treatment on participation can be measured by :

$$E[S_i|Z_i = 1] - E[S_i|Z_i = 0] \tag{1}$$

Effect of treatment on outcome down the road could be measured by:

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] \tag{2}$$

Using our expression for Y_i, we have:

$$E[Y_i|Z_i = 1] = \alpha + \beta E[S_i|Z_i = 1] + E[\epsilon_i|Z_i = 1]$$

and:

$$E[Y_i|Z_i = 0] = \alpha + \beta E[S_i|Z_i = 0] + E[\epsilon_i|Z_i = 0]$$

Therefore

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] =$$

$$\beta(E[S_i|Z_i = 1] - E[S_i|Z_i = 0]) +$$

$$E[\epsilon_i|Z_i = 1] - E[\epsilon_i|Z_i = 0]$$

- What can we assume about $E[\epsilon_i|Z_i = 1] - E[\epsilon_i|Z_i = 0]$?
- What underlies this assumption, and is this justified?

Putting everything together:

$$\hat{\beta} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[S_i|Z_i = 1] - E[S_i|Z_i = 0]} \tag{3}$$

RCT as IV

$$\hat{\beta} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[S_i|Z_i = 1] - E[S_i|Z_i = 0]}$$

- Careful: never forget to check *both* conditions when thinking about using an instrument. The second condition is often not verified even when the first is.
- If assumptions are verified: We obtain the effect of health on knowledge/earnings/anything else by dividing the effect of the program on cognitive scores by the effect of the program on education.

$$\hat{\beta} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[S_i|Z_i = 1] - E[S_i|Z_i = 0]}$$

Equation 1 is the *first stage* relationship (the numerator). Equation 2 is the *reduced form* relationship (the denominator). $\hat{\beta}$ given by equation 15 is the *Wald estimate* of the effect of SHS participation. It is the simplest form of the instrumental variable estimator (Z_i is our instrument).

Scholarship and participation in Senior High School

	(1) Total years of education to date (2019)	(2) Total years of SHS to date (2019)	(3) Completed SHS (2019)	(4) Completed TVI (2019)	(5) Completed tertiary (2019)	(6) Ever enrolled in tertiary program (2019)	(7) Currently enrolled in tertiary program (2019)
Panel A: All							
Treatment	1.241	1.251	0.272	-0.020	0.035	0.044	0.010
	(0.104)	(0.079)	(0.022)	(0.006)	(0.015)	(0.018)	(0.011)
P-value	0.000	0.000	0.000	0.001	0.019	0.016	0.346
Comparison mean	11.387	1.842	0.436	0.029	0.087	0.154	0.049
N	1924	1925	1952	1952	1952	1951	1951
Panel B: Female							
Treatment	1.313	1.208	0.258	-0.005	0.040	0.077	0.029
	(0.155)	(0.119)	(0.032)	(0.008)	(0.020)	(0.025)	(0.015)
P-value	0.000	0.000	0.000	0.573	0.048	0.002	0.059
Comparison mean	11.030	1.627	0.389	0.017	0.078	0.126	0.035
N	968	968	986	986	986	986	986

Scholarship and cognitive test scores

	(1) Total years of education (2013)	(2) Total cognitive score (2013)	(3) Political knowledge score (2013)	(4) Media engagement (radio, newspaper, TV, internet) (2013)	(5) Knows how to use internet (2013)	(6) Has a bank account (2013)	(7) ICT/Social media adoption index (2016)	(8) Uses fertilizer (if in farming) (2017)	(9) Used internet in the past month (2019)
Panel A: All									
Treatment	1.191	0.157	0.095	0.060	0.086	0.058	0.062	-0.024	0.059
	(0.077)	(0.046)	(0.046)	(0.025)	(0.047)	(0.023)	(0.037)	(0.037)	(0.024)
P-value	0.000	0.001	0.040	0.018	0.069	0.011	0.095	0.527	0.013
Comparison mean	10.787	-0.000	0.000	-0.020	0.000	0.314	-0.133	0.471	0.493
N	2064	1983	1981	1981	1983	1984	1995	769	1950
Panel B: Female									
Treatment	1.186	0.194	0.075	0.074	0.050	0.098	0.090	0.020	0.076
	(0.114)	(0.069)	(0.058)	(0.032)	(0.058)	(0.031)	(0.054)	(0.057)	(0.033)
P-value	0.000	0.005	0.192	0.023	0.386	0.001	0.095	0.720	0.023
Comparison mean	10.575	-0.175	-0.381	-0.165	-0.333	0.236	-0.416	0.410	0.402
N	1036	1002	1001	1001	1001	1002	1007	337	985
Panel C: Male									
Treatment	1.183	0.113	0.094	0.035	0.101	0.016	0.016	-0.059	0.033
	(0.101)	(0.059)	(0.061)	(0.037)	(0.067)	(0.033)	(0.045)	(0.050)	(0.033)
P-value	0.000	0.054	0.126	0.347	0.133	0.630	0.724	0.238	0.310
Comparison mean	11.006	0.183	0.397	0.131	0.346	0.396	0.162	0.522	0.590
N	1028	981	980	980	982	982	988	432	965

Scholarship and cognitive test scores

	(1) Worked for pay in past 6 months (2019)	(2) Has wage contract with employer (2019)	(3) Job with benefits (2019)	(4) Public sector employee (2019)	(5) Lives in urban area (2019)	(6) Self-employed (2019)	(7) Total earnings in past 6 months (2019)	(8) Could not cope with 200 GHX emergency (2019)
Panel A: All								
Treatment	0.011	0.039	0.030	0.019	-0.015	-0.029	37.123	-0.027
	(0.021)	(0.015)	(0.015)	(0.013)	(0.015)	(0.020)	(93.450)	(0.017)
P-value	0.589	0.008	0.052	0.157	0.330	0.153	0.691	0.117
Comparison mean	0.730	0.084	0.099	0.077	0.123	0.245	1456.217	0.161
N	1952	1951	1951	1952	1921	1952	1915	1951
Panel B: Female								
Treatment	0.033	0.041	0.020	0.041	-0.029	-0.012	35.794	-0.044
	(0.033)	(0.019)	(0.019)	(0.019)	(0.021)	(0.031)	(108.464)	(0.024)
P-value	0.314	0.032	0.283	0.031	0.152	0.683	0.741	0.070
Comparison mean	0.602	0.063	0.075	0.063	0.119	0.287	951.456	0.176
N	986	986	986	986	973	986	972	986
Panel C: Male								
Treatment	-0.020	0.035	0.037	-0.003	-0.001	-0.042	-12.740	-0.009
	(0.024)	(0.023)	(0.024)	(0.019)	(0.023)	(0.026)	(145.790)	(0.024)
P-value	0.405	0.119	0.126	0.874	0.959	0.106	0.930	0.718
Comparison mean	0.864	0.106	0.125	0.092	0.128	0.201	1993.862	0.146
N	966	965	965	966	948	966	943	965

Wald and IV

- Let us calculate the Wald estimator ourselves, for cognitive scores or earnings .
- Compare to the IV .
- Compare to the OLS

	OLS (no cont)	OLS	IV	DML	DML (LATE)
Sexual Behavior Index (2013-2017)					
Effect of year of education	-0.123	-0.121	-0.074	-0.084	-0.087
Standard Error	(0.007)*	(0.008)*	(0.025)*	(0.007)*	(0.008)*
Social Media Adoption Index (2017)					
Effect of year of education	0.132	0.128	0.062	0.107	0.104
Standard Error	(0.01)*	(0.01)*	(0.032)	(0.01)*	(0.011)*
Technology Adoption Index (2013)					
Effect of year of education	0.15	0.142	0.049	0.1	0.102
Standard Error	(0.01)*	(0.01)*	(0.031)	(0.012)*	(0.12)*
Labor Index (2017)					
Effect of year of education	0.124	0.022	0.084	0.015	0.018
Standard Error	(0.01)*	(0.01)*	(0.028)*	(0.01)	(0.011)

Let's discuss potential violation of exclusion restriction

- You can see that even a "small" violation of either of the conditions for the validity of the instrument can result in very large bias. Any bias in the reduced form will be "blown up" when I divide by the first stage difference.
-
-
-
-

Estimates for whom?

- Some kids would have gone to school anyways
- Some kids did not go to school even without the scholarship
- Some kids were moved to the scholarship to go to school
- How might the returns line up if they are not homogenous?

Wald estimate with heterogeneity in treatment effect

- Let Z_i be an *instrument*, which affects the probability that an individual is treated
- Let $W_i(1)$ be the treatment status for individual i if $Z = 1$, and $W_i(0)$ the treatment status of the same individual if $Z_i = 0$.
- The observed treatment is : $W_i = Z_i W_i(1) + (1 - Z_i) W_i(0)$
- As before, $Y_i(1)$ is potential outcome of treated (if $W_i = 1$) and $Y_i(0)$ is potential outcome if non-treated.
- Identification assumptions (Imbens and Angrist):
  1. *All* Potential outcomes are independent of the Instrument

$$(Y_i(1), Y_i(0), W_i(1), W_i(0)) \perp Z_i$$

  2. What does this imply?
     - Treatment assignment is randomly assigned (or can be treated as such)
     - Treatment has no direct impact on the outcome (that is not implied by randomization of the instrument and has to be argued on a case by case basis!)
  3. Monotonicity: $W_i(1) \geq W_i(0)$ for everyone

More on Monotonicity

Three groups of people :

1. The Compliers: $W_i(1) = 1$ and $W_i(0) = 0$.
2. The Never-Takers: $W_i(1) = 0$ and $W_i(0) = 0$
3. The Always-Takers: $W_i(1) = 1$ and $W_i(0) = 1$
4. The Defiers: $W_i(1) = 0$ and $W_i(0) = 1$

The monotonicity assumption means that there are no defiers. This is not a testable assumption, and needs to be assessed on a case by case basis.

Heterogenous treatment Effect

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]$$

$$= E[W_i(1)Y_i(1) + (1 - W_i(1))Y_i(0)|Z_i = 1]$$
$$-E[W_i(0)Y_i(1) + (1 - W_i(0))Y_i(0)|Z_i = 0]$$

$$= E[(W_i(1) - W_i(0))(Y_i(1) - Y_i(0))] + E[Y_i(0)|Z_i = 1] - E[Y_i(0)|Z_i = 0]$$

$= E[(W_i(1) - W_i(0))(Y_i(1) - Y_i(0))]$ (by independence)
$= E[-(Y_i(1) - Y_i(0))|W_i(1) - W_i(0) = -1]P(W_i(1) - W_i(0) = -1)$
$+E[0 * (Y_i(1) - Y_i(0))|W_i(1) - W_i(0) = 0]P(W_i(1) - W_i(0) = 0)$
$+E[(Y_i(1) - Y_i(0))|W_i(1) - W_i(0) = 1]P(W_i(1) - W_i(0) = 1)$
$= E[Y_i(1) - Y_i(0)|W_i(1) - W_i(0) = 1] * P(W_i(1) - W_i(0) = 1)$
(by monotonicity)
$= E[Y_i(1) - Y_i(0)|W_i(1) - W_i(0) = 1] * (E[W_i(1)] - E[W_i(0)]$

Wald Estimate is treatment effect on the compliers

$$\hat{\beta}_{IV} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[W_i|Z_i = 1] - E[W_i|Z_i = 0]}$$

$$= E[Y_i(1) - Y_i(0)|W_i(1) - W_i(0) = 1]$$

Who are the compliers?

- Special case: Treatment on the Treated:
  - When $W_i(0) = 0$ (e.g. randomized evaluation: all the control stays control)
- General case: Those are compelled by the instrument to get the treatement: external validity?
- While we cannot know who the compliers are, we can describe their characteristics

Interpretation of the IV in the Ghana case?

- Who are the people who get scholarship and do not join SHS? (never takers)
- Who are the people who do not get scholarship but join SHS? (always takers)
- Who are the people who are swayed by the scholarship (complier)?
- is this an interesting group of people?

Comparing IV with other strategies

- DML=double machine learning (Chernozukhov et al.)
- DML-Late: weighted DML.
- "Lalonde" exercise: comparison of DML strategy to IV.

Comparing IV with other strategies

	OLS (no cont)	OLS	IV	DML	DML (LATE)
Sexual Behavior Index (2013-2017)					
Effect of year of education	-0.123	-0.121	-0.074	-0.084	-0.087
Standard Error	(0.007)*	(0.008)*	(0.025)*	(0.007)*	(0.008)*
Social Media Adoption Index (2017)					
Effect of year of education	0.132	0.128	0.062	0.107	0.104
Standard Error	(0.01)*	(0.01)*	(0.032)	(0.01)*	(0.011)*
Technology Adoption Index (2013)					
Effect of year of education	0.15	0.142	0.049	0.1	0.102
Standard Error	(0.01)*	(0.01)*	(0.031)	(0.012)*	(0.12)*
Labor Index (2017)					
Effect of year of education	0.124	0.022	0.084	0.015	0.018
Standard Error	(0.01)*	(0.01)*	(0.028)*	(0.01)	(0.011)

Let's discuss the substantive findings

- Do we find positive effects of education?
- Do we find financial returns to education ?
- Are these returns private or social?

Constructing an instrumental variable from observational data

- Some time you may not have a randomized instrument at your disposal
- But policy variations may create variations in human capital
- Exploiting these requires an extra step: building a solid empirical strategy for the first stage and the reduced form.
- Usually it takes additional assumptions, which you will defend with institutional knowledge.
- Duflo (2001) builds the IV on the DD strategy

"Fuzzy" DD

- First strategy: ratio of the DD (Wald DiD)
- Chaisemartin and Hautefeuille (2017) Show that this identifies the LATE under rather restrictive assumptions (only if the effect of the treatment is stable over time, and if the effect of the treatment is the same in the treatment and in the control group)
- they propose an alternative IV based on a control group where the exposure to the treatment does not change over time.

Instrumental variable

- What can we use as instruments?
  - If we wanted to use just one instrument
  - If we wanted to use many instruments?
- What are the identification assumptions? Do we believe in them?
- ▸ Results Did the IV make a big difference?
- What is the interpretation of the estimate? What are the years of education we are estimating the returns for?
- Interpretation of IV when the treatment takes more than one value: weighted average of marginal effects (going from 0 to 1, 1 to 2, etc..), where the weights are the fraction of people who are moved from one value of the instrument to another.
- See ▸ impact of programs by year of education

Non Experimental approach: Duflo, 2004

- Strategy of the "affected un-treated"
- Use the INPRES program
- We want to estimate the "social returns to education"
- Do we expect externalities to be positive or negative? (why?)
- We are looking to estimate:

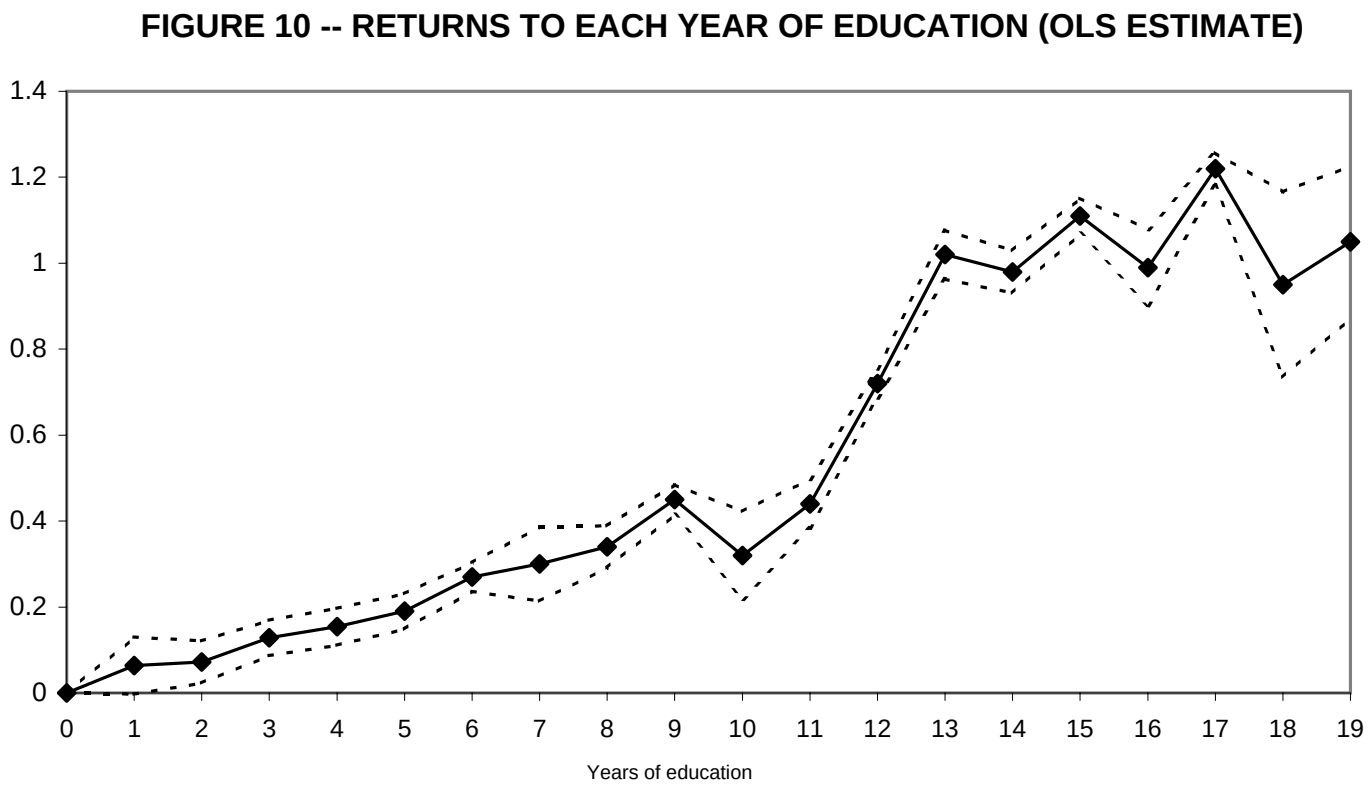
$$y_i = \alpha + \beta S_i + \beta \overline{S_i} + \epsilon_i$$

- Two estimation problem: we need an instrument for S_i and an instrument for $\overline{S_i}$ (Acemoglu and Angrist).
- Consider a cohort who was 12 or older in 1973, and is thus not exposed by the program
- Until 1979, no-one in the labor market is educated in the new schools.
- Starting in 1979, slow influx of the graduate of the new schools ▸ Graph

Empirical Strategy

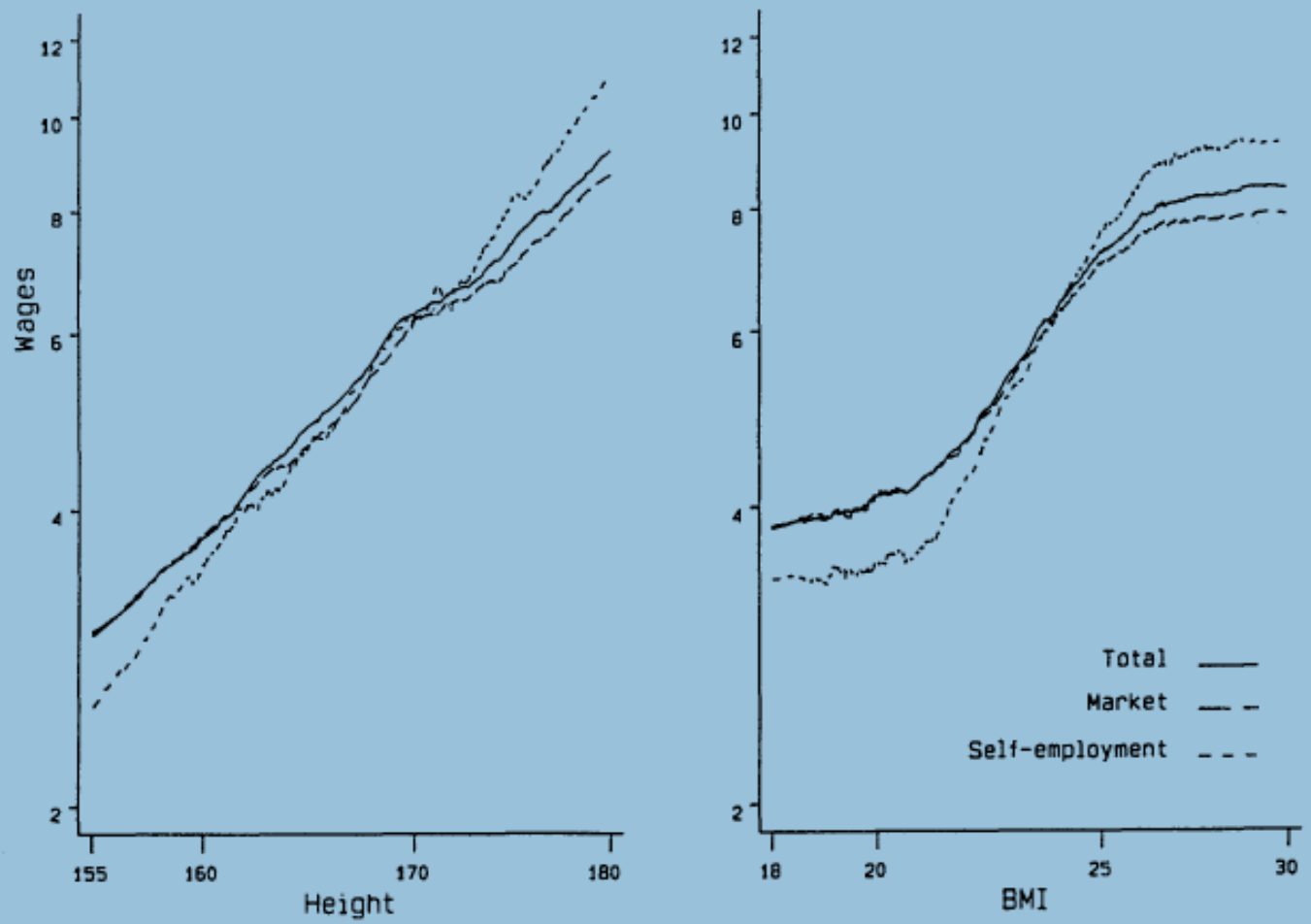
- Fix the cohort, let the years vary.
- Survey Year*Region are instrument for $\overline{S_i}$. Are they correlated with S_i?
- Results (▸ Graph, ▸ Table): Mushy, but if anything, equilibrium effects are negative: consistent with no "A" externality and negative pecuniary externalities

Log(wages) and years of education in Indonesia



Source: Duflo (2001) "Schooling and labor market consequences of school construction in Indonesia"

Figure 1: Non parametric estimates of wage -- health functions
Males



Duflo (2001)

TABLE 3—MEANS OF EDUCATION AND LOG(WAGE) BY COHORT AND LEVEL OF PROGRAM CELLS

	Years of education			Log(wages)		
	Level of program in region of birth			Level of program in region of birth		
	High (1)	Low (2)	Difference (3)	High (4)	Low (5)	Difference (6)
Panel A: Experiment of Interest						
Aged 2 to 6 in 1974	8.49	9.76	−1.27	6.61	6.73	−0.12
	(0.043)	(0.037)	(0.057)	(0.0078)	(0.0064)	(0.010)
Aged 12 to 17 in 1974	8.02	9.40	−1.39	6.87	7.02	−0.15
	(0.053)	(0.042)	(0.067)	(0.0085)	(0.0069)	(0.011)
Difference	0.47	0.36	0.12	−0.26	−0.29	0.026
	(0.070)	(0.038)	(0.089)	(0.011)	(0.0096)	(0.015)
Panel B: Control Experiment						
Aged 12 to 17 in 1974	8.02	9.40	−1.39	6.87	7.02	−0.15
	(0.053)	(0.042)	(0.067)	(0.0085)	(0.0069)	(0.011)
Aged 18 to 24 in 1974	7.70	9.12	−1.42	6.92	7.08	−0.16
	(0.059)	(0.044)	(0.072)	(0.0097)	(0.0076)	(0.012)
Difference	0.32	0.28	0.034	0.056	0.063	0.0070
	(0.080)	(0.061)	(0.098)	(0.013)	(0.010)	(0.016)

Notes: The sample is made of the individuals who earn a wage. Standard errors are in parentheses.

Duflo (2001)

TABLE 4—EFFECT OF THE PROGRAM ON EDUCATION AND WAGES: COEFFICIENTS OF THE INTERACTIONS BETWEEN COHORT DUMMIES AND THE NUMBER OF SCHOOLS CONSTRUCTED PER 1,000 CHILDREN IN THE REGION OF BIRTH

		Dependent variable					
		Years of education			Log(hourly wage)		
	Observations	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Experiment of Interest: Individuals Aged 2 to 6 or 12 to 17 in 1974							
(Youngest cohort: Individuals ages 2 to 6 in 1974)							
Whole sample	78,470	0.124	0.15	0.188			
		(0.0250)	(0.0260)	(0.0289)			
Sample of wage earners	31,061	0.196	0.199	0.259	0.0147	0.0172	0.0270
		(0.0424)	(0.0429)	(0.0499)	(0.00729)	(0.00737)	(0.00850)
Panel B: Control Experiment: Individuals Aged 12 to 24 in 1974							
(Youngest cohort: Individuals ages 12 to 17 in 1974)							
Whole sample	78,488	0.0093	0.0176	0.0075			
		(0.0260)	(0.0271)	(0.0297)			
Sample of wage earners	30,225	0.012	0.024	0.079	0.0031	0.00399	0.0144
		(0.0474)	(0.0481)	(0.0555)	(0.00798)	(0.00809)	(0.00915)
Control variables:							
Year of birth*enrollment rate in 1971		No	Yes	Yes	No	Yes	Yes
Year of birth*water and sanitation program		No	No	Yes	No	No	Yes

Notes: All specifications include region of birth dummies, year of birth dummies, and interactions between the year of birth dummies and the number of children in the region of birth (in 1971). The number of observations listed applies to the specification in columns (1) and (4). Standard errors are in parentheses.

Duflo (2001)

TABLE 5—EFFECT OF THE PROGRAM ON EDUCATION AND WAGES: COEFFICIENTS OF THE INTERACTIONS BETWEEN DUMMIES INDICATING AGE IN 1974 AND THE NUMBER OF SCHOOLS CONSTRUCTED PER 1,000 CHILDREN IN REGION OF BIRTH

	Dependent variable: years of education						Dependent variable: log(hourly wage)		
	Whole sample			Sample of wage earners					
Age in 1974	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
12	−0.035	−0.025	0.002	−0.040	−0.010	0.009	0.016	0.019	0.027
	(0.047)	(0.048)	(0.054)	(0.077)	(0.078)	(0.091)	(0.013)	(0.013)	(0.015)
11	0.011	0.025	0.018	0.008	0.014	−0.003	−0.014	−0.013	−0.009
	(0.046)	(0.047)	(0.051)	(0.073)	(0.074)	(0.083)	(0.012)	(0.013)	(0.014)
10	0.059	0.049	0.078	0.10	0.092	0.13	0.0036	0.0042	0.0059
	(0.047)	(0.049)	(0.054)	(0.075)	(0.076)	(0.090)	(0.013)	(0.013)	(0.015)
9	0.14	0.14	0.15	0.067	0.063	0.17	0.0095	0.010	0.018
	(0.039)	(0.041)	(0.044)	(0.065)	(0.066)	(0.077)	(0.011)	(0.011)	(0.013)
8	0.088	0.11	0.11	0.19	0.20	0.28	0.019	0.021	0.027
	(0.049)	(0.050)	(0.054)	(0.078)	(0.079)	(0.089)	(0.013)	(0.013)	(0.015)
7	0.12	0.14	0.16	0.11	0.13	0.16	−0.0095	−0.0049	0.0066
	(0.044)	(0.046)	(0.051)	(0.072)	(0.073)	(0.084)	(0.012)	(0.012)	(0.014)
6	0.14	0.17	0.26	0.23	0.23	0.32	0.011	0.013	0.018
	(0.042)	(0.044)	(0.049)	(0.070)	(0.070)	(0.084)	(0.012)	(0.012)	(0.014)
5	0.10	0.13	0.13	0.14	0.16	0.27	0.021	0.023	0.052
	(0.043)	(0.045)	(0.050)	(0.075)	(0.075)	(0.088)	(0.013)	(0.013)	(0.015)
4	0.11	0.12	0.18	0.19	0.19	0.29	0.019	0.020	0.038
	(0.039)	(0.041)	(0.046)	(0.069)	(0.069)	(0.082)	(0.012)	(0.012)	(0.014)
3	0.11	0.14	0.20	0.15	0.17	0.30	0.0079	0.013	0.027
	(0.044)	(0.046)	(0.053)	(0.079)	(0.080)	(0.097)	(0.013)	(0.014)	(0.016)
2	0.14	0.19	0.19	0.20	0.22	0.25	0.016	0.023	0.040
	(0.041)	(0.043)	(0.049)	(0.073)	(0.074)	(0.088)	(0.012)	(0.013)	(0.015)
Control variables:[a]									
Year of birth*enrollment rate in 1971	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Year of birth*water and sanitation program	No	No	Yes	No	No	Yes	No	No	Yes
F-statistic[b]	4.03	5.18	6.15	2.70	2.74	4.38	1.13	1.29	2.05
R^2	0.19	0.19	0.17	0.14	0.14	0.13	0.14	0.15	0.13
Number of observations	152,989	152,495	143,107	60,633	60,466	55,144	60,633	60,466	55,144

Duflo (2001)

TABLE 7—EFFECT OF EDUCATION ON LABOR MARKET OUTCOMES: OLS AND 2SLS ESTIMATES

Method	Instrument	(1)	(2)	(3)	(4
Panel A: Sample of Wage Earners					
Panel A1: Dependent variable: log(hourly wage)					
OLS		0.0776	0.0777	0.0767	
		(0.000620)	(0.000621)	(0.000646)	
2SLS	Year of birth dummies*program intensity in region of birth	0.0675	0.0809	0.106	0.0
		(0.0280)	(0.0272)	(0.0222)	(0.0
		[0.96]	[0.9]	[0.93]	[0.9
2SLS	(Aged 2–6 in 1974)*program intensity in region of birth	0.0752	0.0862	0.104	
		(0.0338)	(0.0336)	(0.0304)	
		(0.0338)	(0.0336)	(0.0304)	
Panel A2: Dependent variable: log(monthly earnings)					
OLS		0.0698	0.0698	0.0689	
		(0.000601)	(0.000602)	(0.000628)	
2SLS	Year of birth dummies*program intensity in region of birth	0.0756	0.0925	0.0913	0.1
		(0.0280)	(0.0278)	(0.0219)	(0.0
		[0.73]	[0.63]	[0.58]	[0.7
Panel B: Whole Sample					
Panel B1: Dependent variable: participation in the wage sector					
OLS		0.0328	0.0327	0.0337	
		(0.00311)	(0.000311)	(0.000319)	
2SLS	Year of birth dummies*program intensity in region of birth	0.101	0.118	0.0892	
		(0.0210)	(0.0197)	(0.0162)	
		[0.66]	[0.93]	[1.12]	
Panel B2: Dependent variable: log(monthly earnings), imputed for self-employed individuals					
OLS		0.0539	0.0539	0.0539	
		(0.000354)	(0.000354)	(0.000355)	
2SLS	Year of birth dummies*program intensity in region of birth	0.0509	0.0745	0.0346	
		(0.0157)	(0.0136)	(0.0138)	
		[0.68]	[0.58]	[1.16]	
Control variables:					

Duflo (2001)

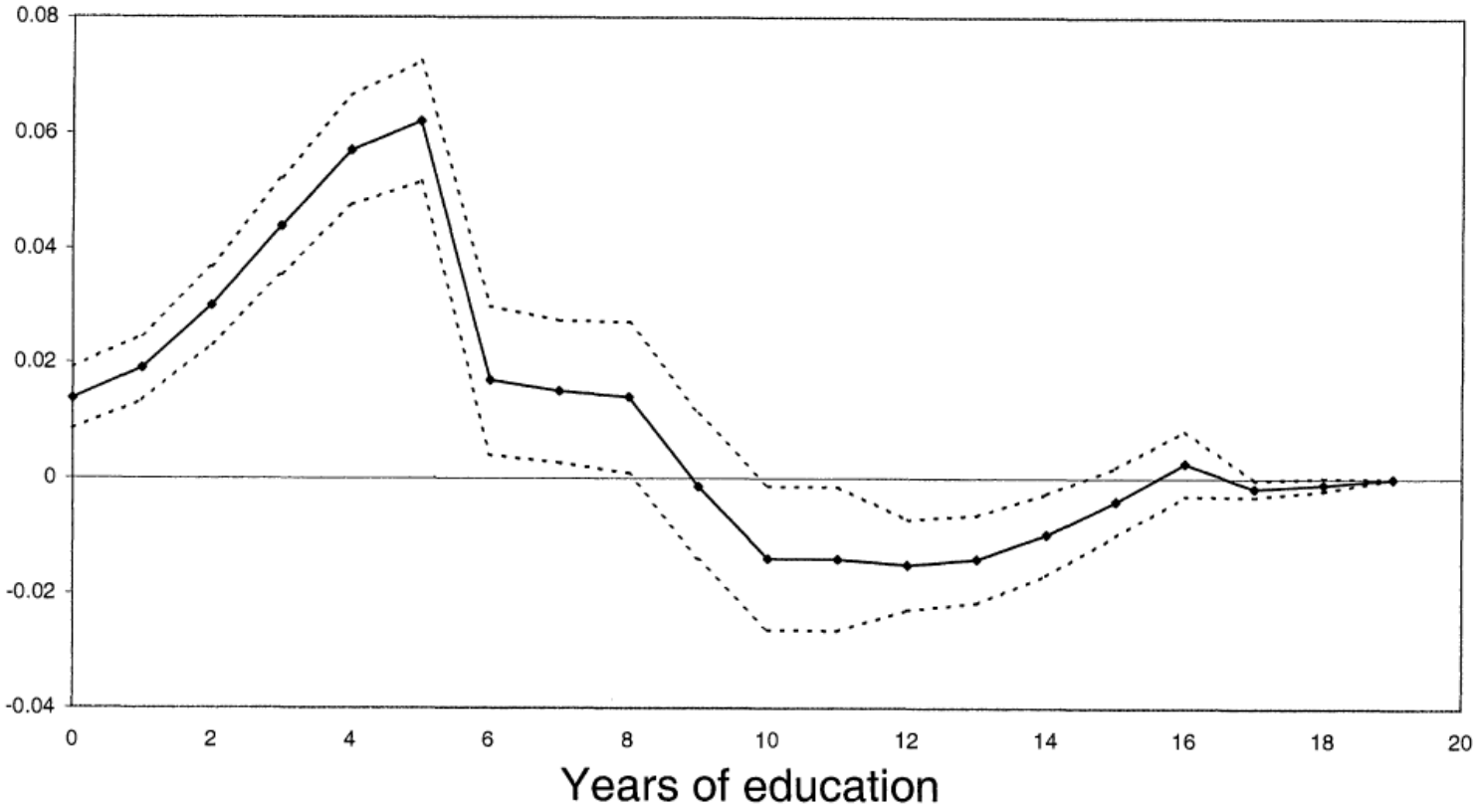


FIGURE 2. DIFFERENCE IN DIFFERENCES IN CDF (ESTIMATED FROM LINEAR PROBABILITY MODEL) WITH 95-PERCENT CONFIDENCE INTERVAL

Duflo (2001)

FIGURE 3. COEFFICIENTS OF THE INTERACTIONS AGE IN 1974* PROGRAM INTENSITY IN THE REGION OF BIRTH IN THE WAGE AND EDUCATION EQUATIONS

Duflo (2004)

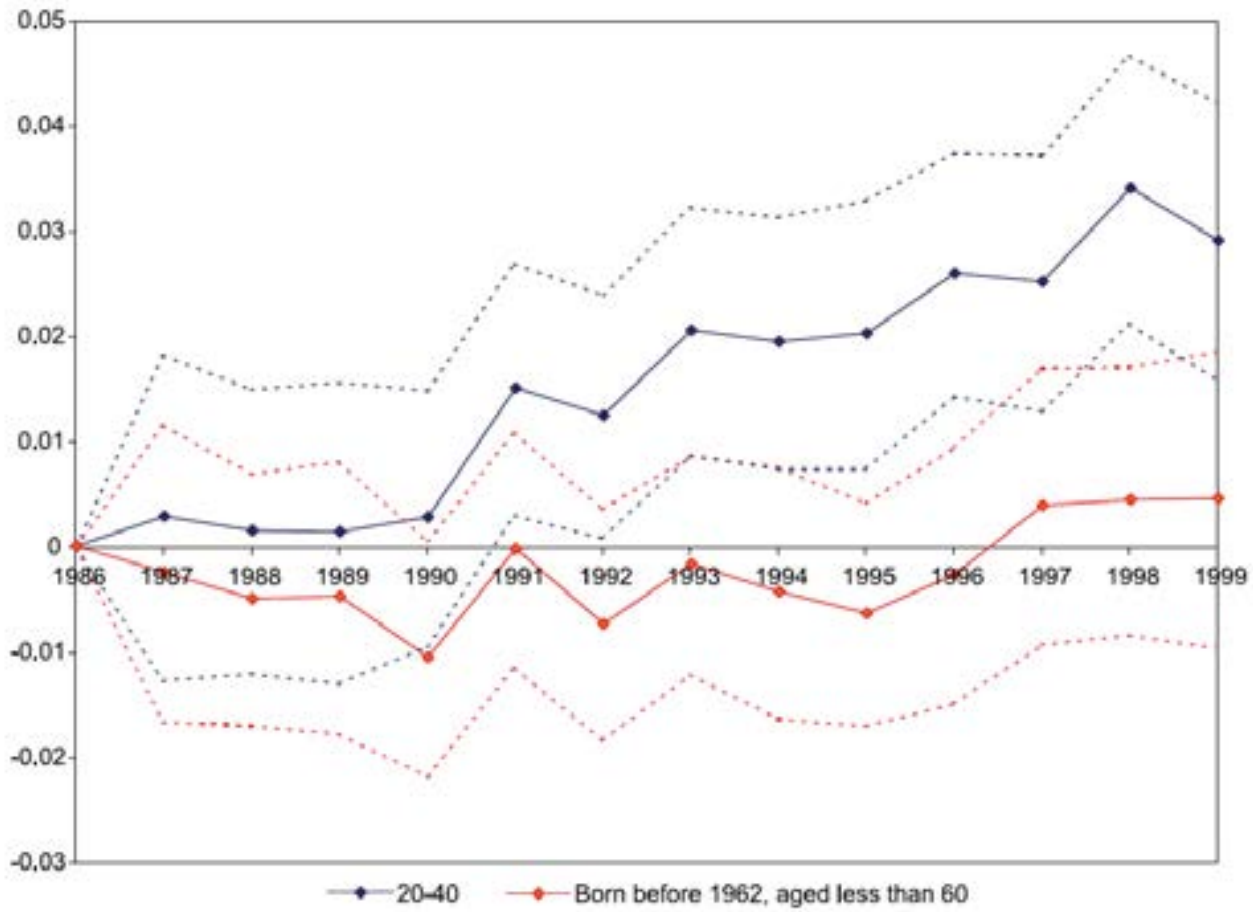


Fig. 2. Coefficients of the interactions of program intensity and survey year dummies. Dependent variable: % of primary school graduates.

Duflo (2004)

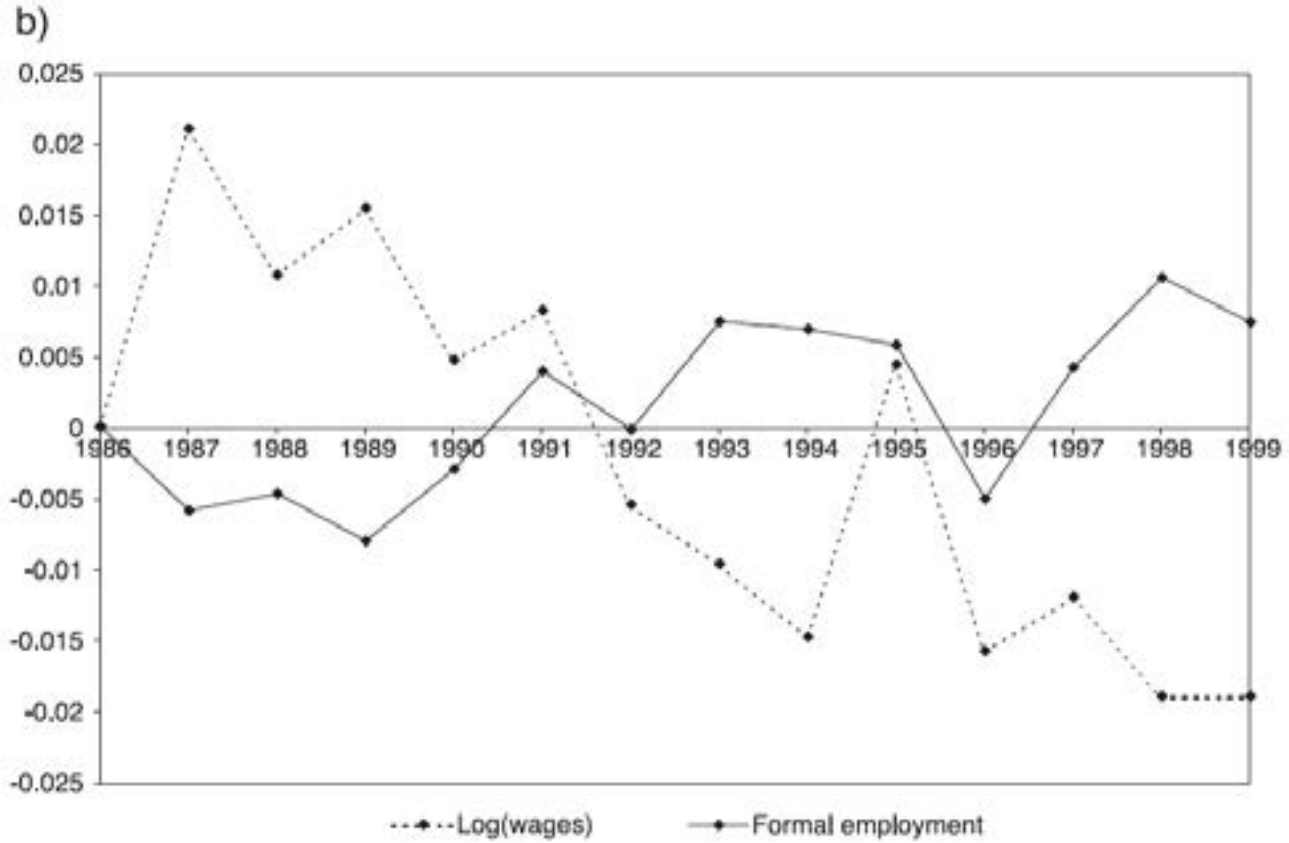


Fig. 4. (a) Coefficients of the interactions of program intensity and survey year dummies. Dependent variables: log(wage) and formal sector employment (individuals born before 1962 and aged less than 60). Sample: urban and rural regions. (b) Coefficients of the interactions of program intensity and survey year dummies. Dependent variables: average log(wage) and average formal sector employment among individuals born before 1962 and aged less than 60. Sample: rural regions.

Duflo (2004)

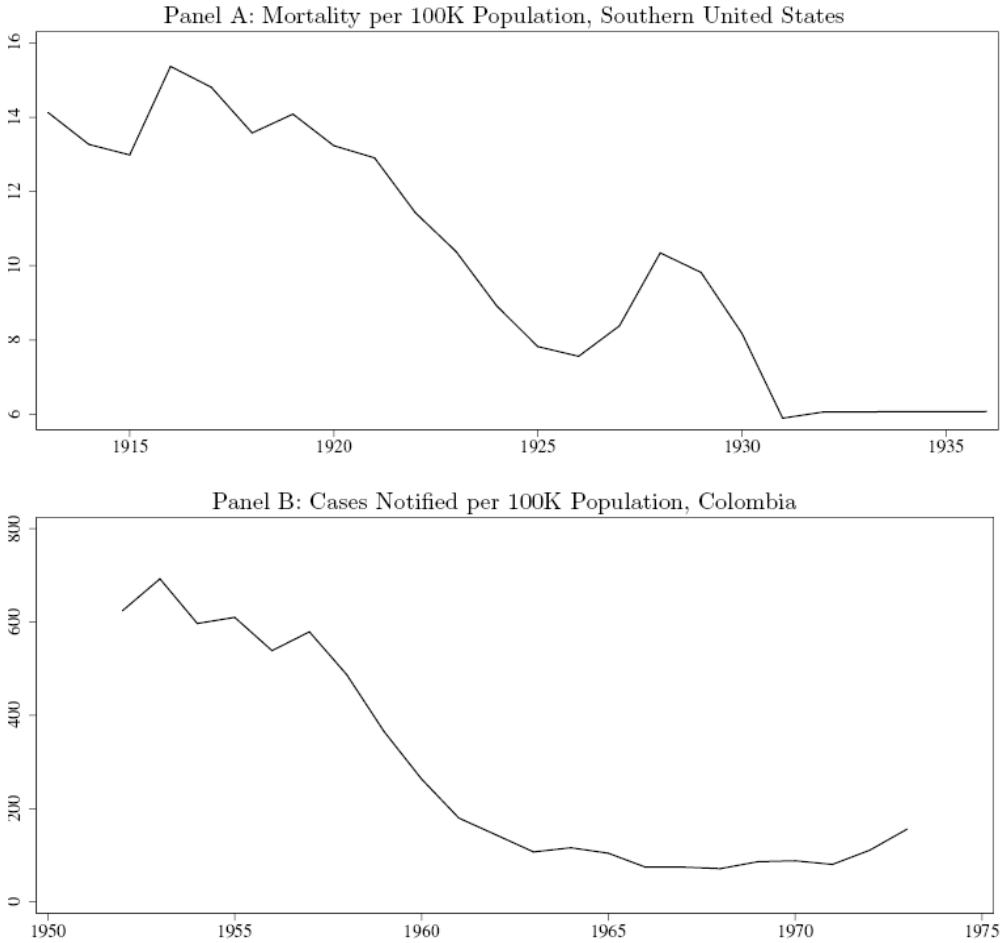
Table 6
2SLS estimates of the impact of average education on individual wages

	Independent variable: % of primary school graduates in the 20–40 sample		Independent variable: % of primary school graduates in the 20–60 sample	
	Sample: rural and urban areas	Sample: rural areas only	Sample: rural and urban areas	Sample: rural areas only
	(1)	(2)	(3)	(4)
Panel A: years 1986–1999				
Log (wage)	−0.204 (0.443)	−0.834 (0.701)	−0.208 (0.615)	−0.871 (0.837)
Log (wage) residual	−0.292 (0.355)	−0.633 (0.431)	−0.379 (0.512)	−0.994 (0.556)
Skill premium	−0.434 (0.916)	−0.982 (1.408)	−0.596 (1.197)	−0.636 (1.645)
Formal employment	0.441 (0.159)	0.454 (0.203)	0.661 (0.238)	0.745 (0.352)
Formal employment among educated workers	0.432 (0.197)	0.501 (0.259)	0.543 (0.264)	0.713 (0.406)
Formal employment among uneducated workers	0.379 (0.203)	0.409 (0.232)	0.510 (0.354)	0.318 (0.318)
Panel B: years 1986–1997				
Log (wage)	−0.358 (0.493)	−0.710 (0.821)	−0.451 (0.716)	−0.480 (0.801)
Log (wage) residual	−0.330 (0.412)	−0.588 (0.529)	−0.437 (0.618)	−0.902 (0.602)
Skill premium	−0.225 (1.033)	−0.635 (1.461)	−0.291 (1.488)	0.536 (1.576)
Formal employment	0.463 (0.183)	0.442 (0.233)	0.716 (0.282)	0.694 (0.379)
Formal employment among educated workers	0.428 (0.229)	0.473 (0.301)	0.530 (0.317)	0.622 (0.479)
Formal employment among uneducated workers	0.478 (0.249)	0.449 (0.277)	0.624 (0.415)	0.263 (0.319)

Men aged 20–60 and born before 1962.
1. Survey year dummies, region dummies, interactions between survey year dummies and the enrollment rate in 1971, and interactions between survey year dummies and the number of children are included in the regressions. 2. Regression run using kabupaten-year averages, weighted by the number of observations in each kabupaten-year cell. 3. The instruments are interactions between survey year dummies and the program intensity. 4. The standard errors are corrected for auto-correlation within kabupaten.

Bleakley (2006)

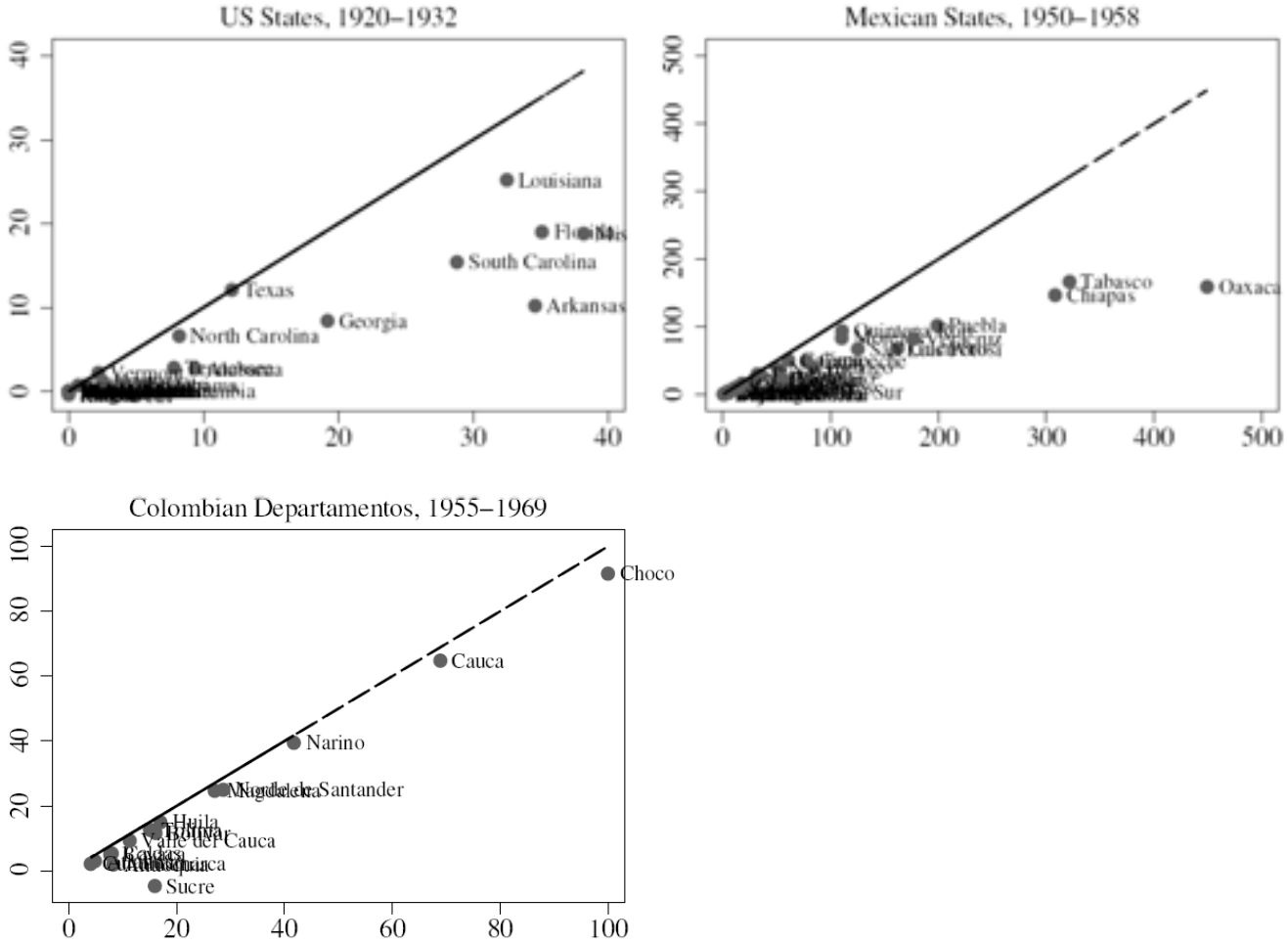
Figure 1: Malaria Incidence Before and After the Eradication Campaigns



Notes: Panel A plots the estimated malaria mortality per capita for the Southern region and bordering states. Because the death registration system was being phased in over the period, a regression model with state fixed effects is used to control for sample changes, and the time series is constructed from the year dummies in the regression, normalized to match the end-of-period data when all states were represented. (Census Bureau *Vital Statistics*, various years, and author's calculations.) Panel B reports data on notified cases of malaria for Colombia (SEM, 1979).

Bleakley (2006)

Figure 2: Highly Infected Areas Saw Greater Declines in Malaria



Notes: The y axis displays the estimated decrease in malaria mortality post-intervention. The x axis is the pre-campaign malaria mortality rate. The 45-degree line represents complete eradication. Both variables are expressed per 100,000 population. United States data are reported in Maxcy (1923) and *Vital Statistics* (Census, 1933). Mexican data are drawn from Pesqueira (1957) and from the Mexican *Anuario Estadístico* (Dirección General de Estadística, 1960). SEM (1957) and the Colombian *Anuario de Salubridad* (DANE, 1968-70) are the sources for the Colombian data.

14.771: Development Economics

Fall 2021