Instrumental Variable: Applications

Econ 140, Section 8

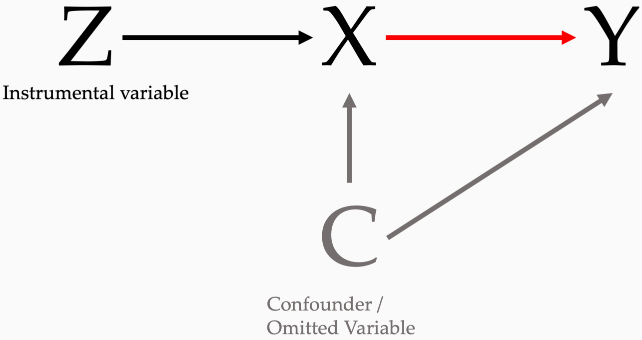
Jonathan Old

Roadmap

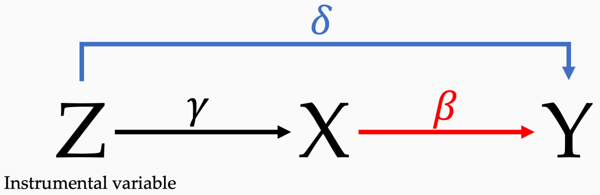
1. Recap: IV
2. Group work
3. Colonial Origins of Development [SA8-Q1]
4. LATE [SA8-Q2]

Recap: IV

Recap: IV



Recap: IV ”rescales” the effect



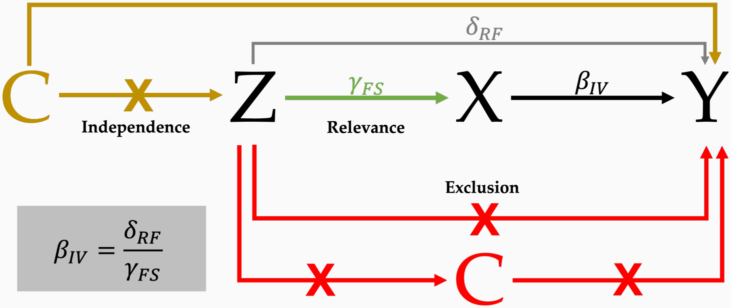
A simple example:

- We want to know the effect of chocolate (X) on happiness (Y), using a randomized voucher as instrument (Z).
- We find: people with voucher were 3 points more happy ($\delta = 3$), and ate 0.5 more chocolates ($\gamma = 0.5$).
- Then, the effect of eating one more chocolate is $\beta = \delta/\gamma = 3/0.5 = 6$.

Recap: IV summary

We need the following three assumptions for IV to work:

1. **Relevance**: Z must truly affect X
2. **Independence**: Z is as good as randomly assigned
3. **Exclusion Restriction**: The **only** way that Z affects Y is via X.



$$\beta_{IV} = \frac{\delta_{RF}}{\gamma_{FS}}$$

Group work

Get into groups of 4 …

… Your job today: Ask as many ”silly” questions as possible!

Group work

Group 1: We are interested in the effect of being in the army on crime. We instrument being in the army with a lottery (paper)

Group 2: We are interested in the effect of income on conflict. We instrument income with rainfall (paper)

Group 3: We are interested in the effect of air pollution on mortality. We instrument local air pollution with wind direction (paper)

1. Relevance: Z must truly affect X
2. Independence: Z is as good as randomly assigned
3. Exclusion restriction: The **only** way that Z affects Y is via X

Your job: Discuss whether these assumptions hold!

Any questions?

... Remember: The more questions, the better!

Colonial Origins of Development [SA8-Q1]

Introduction

Settler Mortality: OLS [SA8-Q1b]

The authors are interested in the effect of good institutions (C_i) on economic development (log GDP per capita, $\ln y_i$). They also have a series of control variables X_i, and run the following regression:

$$\ln y_i = \alpha + \beta C_i + \mathbf{X_i}\gamma + u_i$$

- Interpret β.
- Suppose that institutional quality is measured from 0 (worst institutions) to 1 (best institutions). What sign do you expect β to have?
- You find that $\hat{\beta} = 0.8$. What do you conclude about the causal effect of good institutions on economic development?

Settler Mortality: OLS [SA8-Q1b]

- β measures the association between institutional quality and GDP per capita in the data. An increase in institutional quality by one unit is associated with an increase in GDP per capita by $\beta 100\%$, all else equal (the last because we have control variables included).

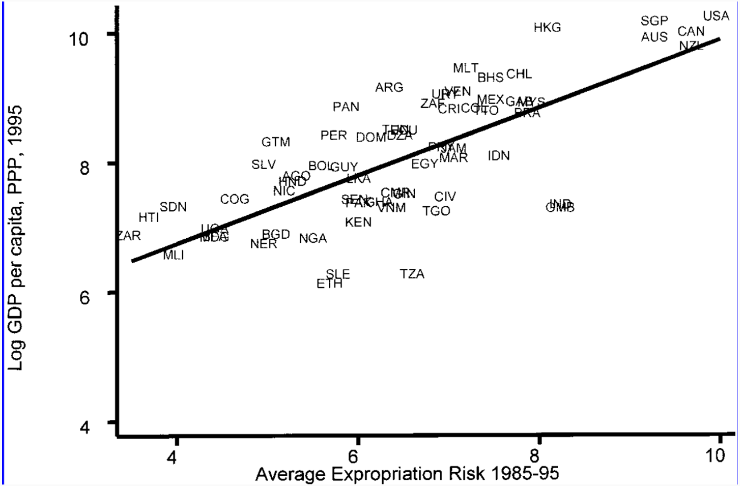
Settler Mortality: OLS [SA8-Q1b]

- β measures the association between institutional quality and GDP per capita in the data. An increase in institutional quality by one unit is associated with an increase in GDP per capita by $\beta 100\%$, all else equal (the last because we have control variables included).
- In the world, we typically see that richer countries have better institutions. Therefore, we expect β to be positive.

Settler Mortality: OLS [SA8-Q1b]

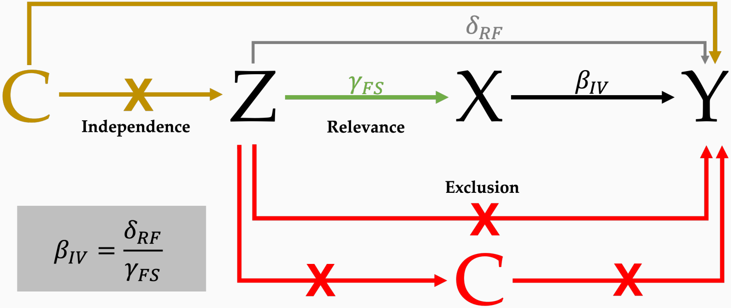
- β measures the association between institutional quality and GDP per capita in the data. An increase in institutional quality by one unit is associated with an increase in GDP per capita by $\beta 100\%$, all else equal (the last because we have control variables included).
- In the world, we typically see that richer countries have better institutions. Therefore, we expect β to be positive.
- The coefficient means that, in the data, we observe that the countries with the best institutions have around 80% higher GDP per capita than the countries with the worst institutions. However, since there are omitted variables, this does not tell us anything about the causal effect of institutions.

Settler Mortality: OLS [SA8-Q1b]



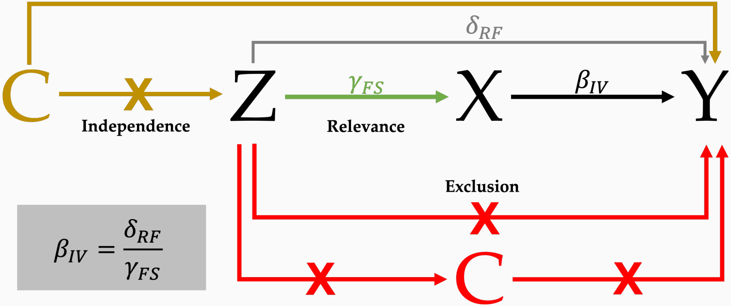
Settler Mortality: IV Approach [SA8-Q1c]

- The authors propose to use an instrumental variables approach. They use settler mortality (S_i) as an instrument for the quality of institutions (C_i).
- (In Groups:) Describe in words the three IV assumptions that settler mortality has to fulfil here to be valid.



Settler Mortality: IV Approach [SA8-Q1d]

- (In Groups:) For each of the three IV conditions, give one argument why they hold, and one example how they could be violated



$$\beta_{IV} = \frac{\delta_{RF}}{\gamma_{FS}}$$

Q1c,d: Sketch of solutions

1. Relevance: historical settler mortality and current institutions must be strongly related (an empirical question we can verify. Maybe, conditional on control variables, they are not!)

Q1c,d: Sketch of solutions

1. Relevance: historical settler mortality and current institutions must be strongly related (an empirical question we can verify. Maybe, conditional on control variables, they are not!)
2. Independence (Exogeneity): historical settler mortality cannot be correlated with anything (besides current institutions and our control variables) that impact GDP (what if climate or temperature affects both settler mortality and GDP today – e.g., via malaria)

Q1c,d: Sketch of solutions

1. Relevance: historical settler mortality and current institutions must be strongly related (an empirical question we can verify. Maybe, conditional on control variables, they are not!)
2. Independence (Exogeneity): historical settler mortality cannot be correlated with anything (besides current institutions and our control variables) that impact GDP (what if climate or temperature affects both settler mortality and GDP today – e.g., via malaria)
3. Exclusion: historical settler mortality can only impact GDP per capita through its impact on current institutions (What if settler mortality also led to different probabilities of being colonized?)

Theory

- (potential) settler mortality
- settlements
- early institutions
- current institutions
- current performance

Settler Mortality: IV Relevance [SA8-Q1e]

- How would you write out the regression equation to estimate whether the first stage is strong?

Settler Mortality: IV Relevance [SA8-Q1e]

- How would you write out the regression equation to estimate whether the first stage is strong?

$$C_i = \delta + \eta S_i + \mathsf{X_i}\theta + v_i$$

Settler Mortality: IV Relevance [SA8-Q1e]

- How would you write out the regression equation to estimate whether the first stage is strong?

$$C_i = \delta + \eta S_i + \mathbf{X_i}\theta + v_i$$

- What coefficient should we look at to evaluate the strength of the first stage?

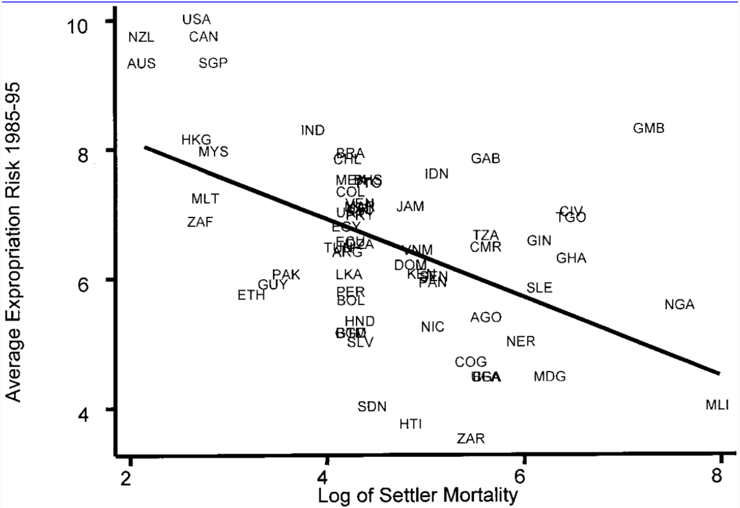
Settler Mortality: IV Relevance [SA8-Q1e]

- How would you write out the regression equation to estimate whether the first stage is strong?

$$C_i = \delta + \eta S_i + \mathbf{X_i}\theta + v_i$$

- What coefficient should we look at to evaluate the strength of the first stage? We look at $\hat{\eta}$: This should be very significant. The relevant F-statistic (with one instrument, this is the **square** of the t-statistic) should be at least 10.

Settler Mortality: IV Relevance [SA8-Q1e]



Settler Mortality: IV Reduced Form [SA8-Q1f]

- The authors document a strong ”reduced-form” relationship. Write out the regression equation used to estimate the reduced form

Settler Mortality: IV Reduced Form [SA8-Q1f]

- The authors document a strong ”reduced-form” relationship. Write out the regression equation used to estimate the reduced form

$$\ln y_i = \mu + \lambda S_i + \mathsf{X_i}\kappa + e_i$$

Settler Mortality: IV Reduced Form [SA8-Q1f]

- The authors document a strong ”reduced-form” relationship. Write out the regression equation used to estimate the reduced form

$$\ln y_i = \mu + \lambda S_i + \mathsf{X_i}\kappa + e_i$$

- Why don’t we just estimate the reduced form?

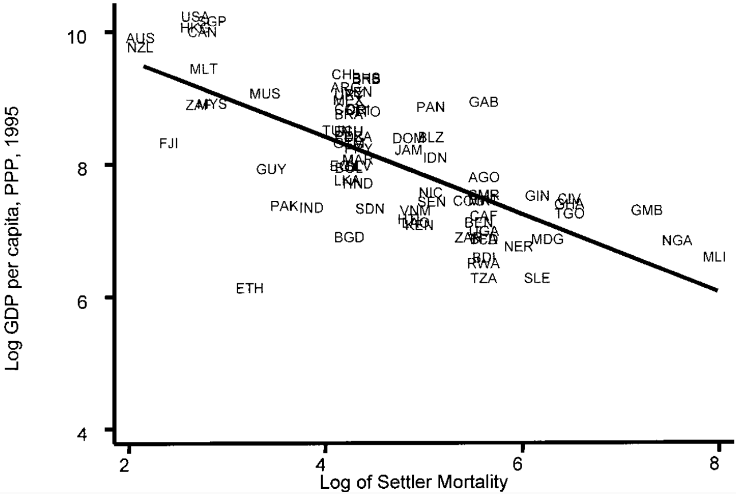
Settler Mortality: IV Reduced Form [SA8-Q1f]

- The authors document a strong ”reduced-form” relationship. Write out the regression equation used to estimate the reduced form

$$\ln y_i = \mu + \lambda S_i + \mathsf{X_i} \kappa + e_i$$

- Why don’t we just estimate the reduced form? We are not primarily interested in the effect of settler mortality, but we use settler mortality as a way to find out the causal effect of institutions on development outcomes.

Settler Mortality: IV Reduced Form [SA8-Q1f]



Settler Mortality: Results [SA8-Q1g]

	OLS			2SLS		
Average protection against expropriation risk (s_i)	0.52 (0.06)	0.47 (0.06)	0.41 (0.06)	0.94 (0.16)	1.00 (0.22)	1.10 (0.46)
Latitude		0.89 (0.49)	0.92 (0.63)		-0.65 (1.34)	- 1.20 (1.8)
Asia dummy			-0.60 (0.23)			-1.10 (0.52)
Africa dummy			-0.90 (0.17)			-0.44 (0.42)
"Other" continent dummy			-0.04 (0.32)			-0.99 (1.00)

- Interpret the results of the OLS regressions
- Interpret the results of the 2SLS (IV) regressions
- How do the OLS and 2SLS estimates differ? What does this imply about the OLS regression from part b)?

Settler Mortality: Results [SA8-Q1g]

- OLS: An increase in the index of protection against expropriation by one unit is associated with around 40-50% higher GDP per capita (keeping latitude and continent fixed)

Settler Mortality: Results [SA8-Q1g]

- OLS: An increase in the index of protection against expropriation by one unit is associated with around 40-50% higher GDP per capita (keeping latitude and continent fixed)
- IV: An increase in the index of protection against expropriation by one unit leads to around 100% higher GDP per capita (keeping latitude and continent fixed)

Settler Mortality: Results [SA8-Q1g]

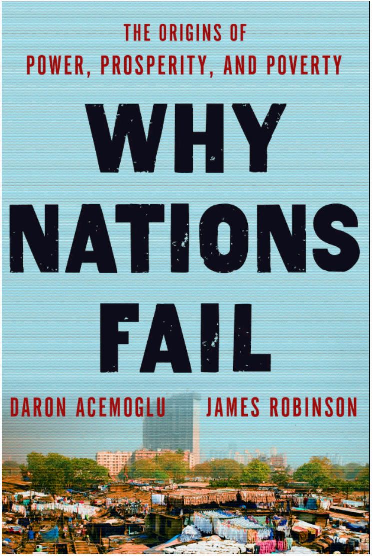
- OLS: An increase in the index of protection against expropriation by one unit is associated with around 40-50% higher GDP per capita (keeping latitude and continent fixed)
- IV: An increase in the index of protection against expropriation by one unit leads to around 100% higher GDP per capita (keeping latitude and continent fixed)
- The IV estimates are larger than the OLS estimates. Negative OVB or reverse causality should make the IV estimate smaller than the OLS estimate, so they may not be important here. The OLS estimate could be biased towards zero because of measurement error. In addition, the countries for which settler mortality had a big effect on today's institutions may be the countries for which institutions had the biggest effect on GDP (LATE).

Settler Mortality: Results [SA8-Q1g]

All the ways IV estimates can be different from OLS estimates

- **OVB:** If OLS had omitted variable bias, and our instrument is valid, then the IV estimate should be different. We can check whether this is plausible with the OVB formula
- **Measurement error:** If we have random measurement error in the independent variable (X), then we can use IV to overcome this. In that case, the IV coefficient will be larger (in absolute value) than the OLS coefficient
- **LATE:** OLS gives us ATT + Selection Bias, while IV gives us the treatment effect on the compliers (LATE). The ATT may be different from LATE, even without selection bias.
- **Invalid IV:** Hard to determine what exactly is going on
- **Sampling variation**: This can just happen by chance

Want to know more?



Any questions?

… Remember – Every question is useful!

LATE [SA8-Q2]

LATE: IV gives us the treatment effect for the compliers

Potential outcomes! (unobserved)		***Does not get voucher (Z=0)***	
Gets voucher (Z=1)		*Eats chocolate (D=1)*	*Does not eat chocolate (D=0)*
	Eats chocolate (D=1)	Always-takers: E(D|Z=1)=E(D|Z=0)=1 ➔ E(Y|Z=1)=E(Y|Z=0)	Compliers
	Does not eat chocolate (D=0)	Defiers	Never-takers: E(D|Z=1)=E(D|Z=0)=0 ➔ E(Y|Z=1)=E(Y|Z=0)

Health Insurance and LATE [SA8-Q2a]

Suppose individuals are eligible for a public health insurance program and are randomly assigned a priority number Z, which influences how likely they are to be enrolled in the program. Let β_{1i} represent the causal effect for individual i of enrolling in public health insurance on health outcomes, and let π_{1i} represent the effect of Z on the likelihood that individual i enrolls.

Half the individuals are aware that their health outcomes would improve from the program and thus may decide to enroll if given the option. For this population, would β_{1i} be positive, zero, or negative? What about π_{1i} ? Call these β_{1i}^{A} and π_{1i}^{A}.

Never takers? [SA8-Q2b]

The other half, perhaps those who are healthier, know their health outcomes would not improve from enrolling in health insurance. These individuals would not enroll even if they were admitted. Are β_{1i} and π_{1i} positive, zero, or negative for this group? Call these β_1^B and π_1^B.

IV and ATE [SA8-Q2d]

If we use Z as an instrument for enrollment, would the resulting estimate identify the average treatment effect (ATE) of health insurance on health outcomes? Please explain your reasoning.

Solution No. In general, the 2SLS estimate identifies a local average treatment effect (LATE). The LATE equals the ATE only if at least one of the following cases holds: (i) the treatment effect is the same for all individuals, (ii) the instrument affects all individuals equally, or (iii) heterogeneity in the treatment effect and heterogeneity in the effect of the instrument are uncorrelated.

Based on (a) and (b), the treatment effect and the effect of the instrument are different for the two groups, ruling out cases (i) and (ii). For group (a), $\beta_{1i}^{A} > 0$ and $\pi_{1i}^{A} > 0$, whereas for group (b) $\beta_{1i}^{B} \leq 0$ and $\pi_{1i}^{B} = 0$. This implies that $\text{cov}\left(\beta_{1i}, \pi_{1i}\right) > 0$, ruling out case (iii).

Calculating LATE [SA8-Q2e]

Derive and interpret the local average treatment effect (LATE). How does the LATE compare to the ATE?

Calculating LATE [SA8-Q2e]

Solution

When both the treatment effect and the effect of the instrument are heterogeneous, $\widehat{\beta}^{IV}$ estimates the following:

$$\widehat{\beta}^{IV} = \frac{E\left[\beta_{1i}\pi_{1i}\right]}{E\left[\pi_{1i}\right]} = \text{ LATE}$$

Let's evaluate this:

$$E\left[\pi_{1i}\right] = 0.5 * \pi_1^A + 0.5 * \pi_1^B = 0.5 * \pi_1^A$$

$$E\left[\beta_{1i}\pi_{1i}\right] = 0.5 * \beta_1^A\pi_1^A + 0.5 * \beta_1^B\pi_1^B = 0.5 * \beta_1^A\pi_1^A$$

Plugging these into the expression for the LATE, we get that the LATE identifies β_1^A, the causal effect for individuals who might enroll in health insurance. It gives zero weight to individuals who would never enroll. Because in this case $\beta_1^A > \beta_1^B$, the LATE exceeds the average treatment effect.

Any questions?

... Remember – Every question is useful!