API-202B Empirical Methods II

Session #13: Instrumental Variables Case Studies

miguel_santos@hks.harvard.edu
@miguelsantos12

Our class today

- Instrumental variables review (IV)
- Example 1: Colonial origins, institutions and growth
- Example 2: Minneapolis Domestic Violence Experiment (MDVE)
- Takeaways

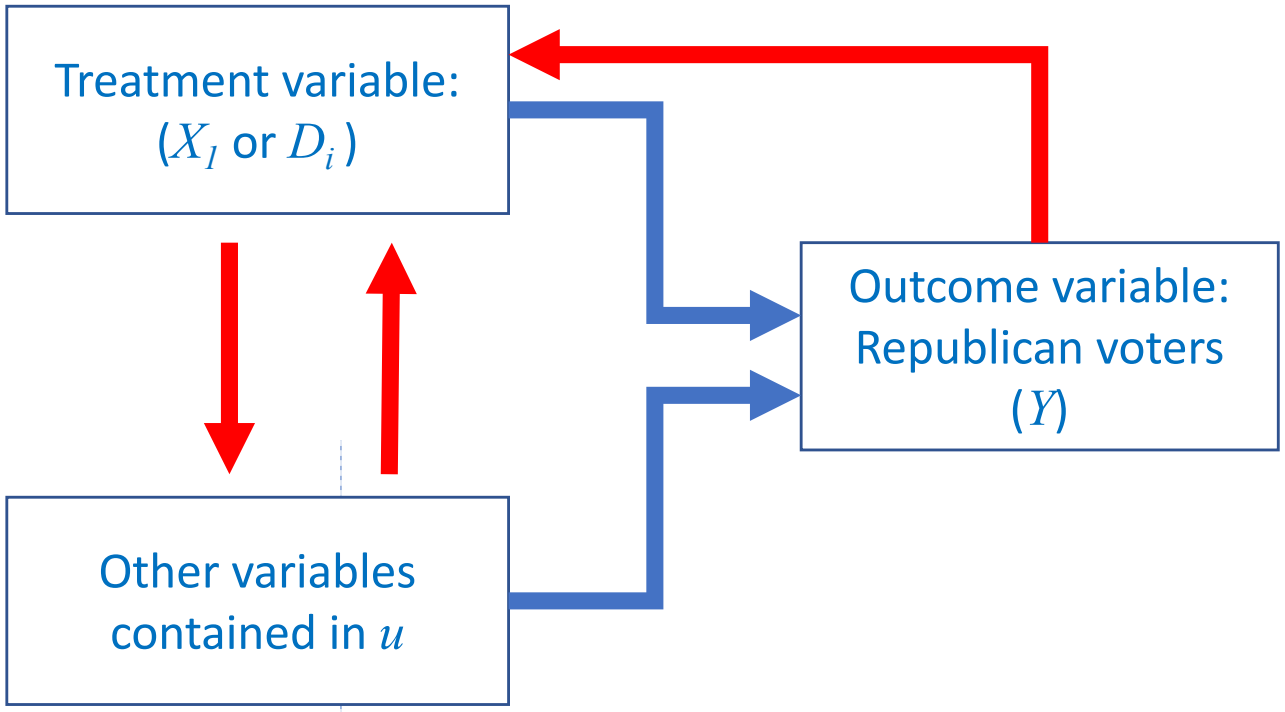
Instrumental Variables Review

- When dealing with instrumental variables, always look for:
  - An independent variable of interest **(Y)**
  - An endogenous regressor **(X)** or treatment variable **(D)**
  - An exogenous instrumental variable **(Z)**
    - **Exogenous:** Its value cannot plausibly depend on any other variable in the system

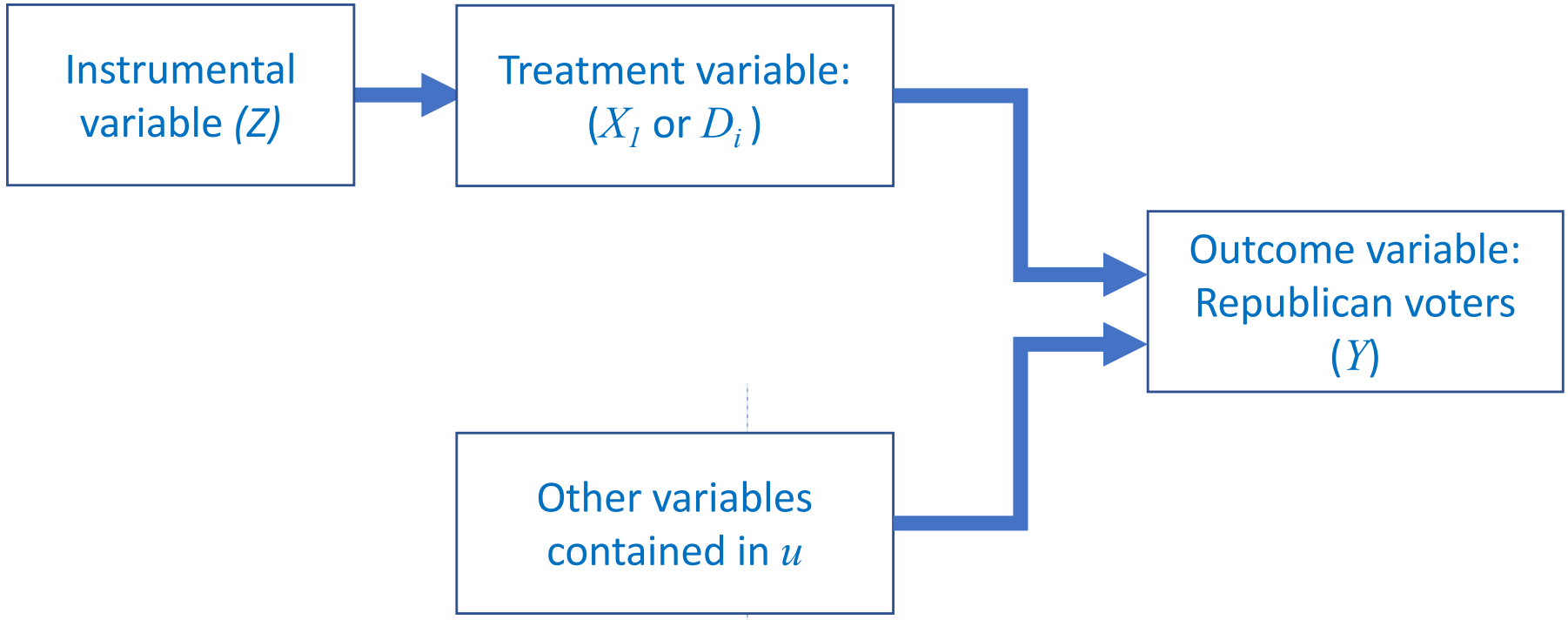
Instrumental Variables Review: What is the problem we are trying to solve?

Simultaneous causality:
X_1 causes Y, but plausibly Y can also be a driver of X_1

Endogeneity:
Explanatory variable is correlated with the error term



Instrumental Variables Review: What is the problem we are trying to solve?



Instrumental variables: First stage and second stage regressions

First stage regression:

$$X_1 = \alpha_0 + \alpha_1 Z + \varepsilon \quad \text{or} \quad D_i = \alpha_0 + \alpha_1 Z + \varepsilon$$

Second stage regression:

$$Y = \beta_0 + \beta_1 \hat{X}_1 + u \quad \text{or} \quad Y = \beta_0 + \beta_1 \hat{D}_i + u$$

Instrument Z must satisfy **two conditions:**

1. **Instrument relevance**: corr $(Z, X_1) \neq 0$. The instrument is correlated with X_1, the endogenous (non-random) regressor

2. **Instrument exogeneity (exclusion restriction)** corr $(Z, u) = 0$. The only way the instrument **Z** affects the outcome variable **Y** is via the endogenous variable (X_1 or D_i)

Instrumental Variables Review

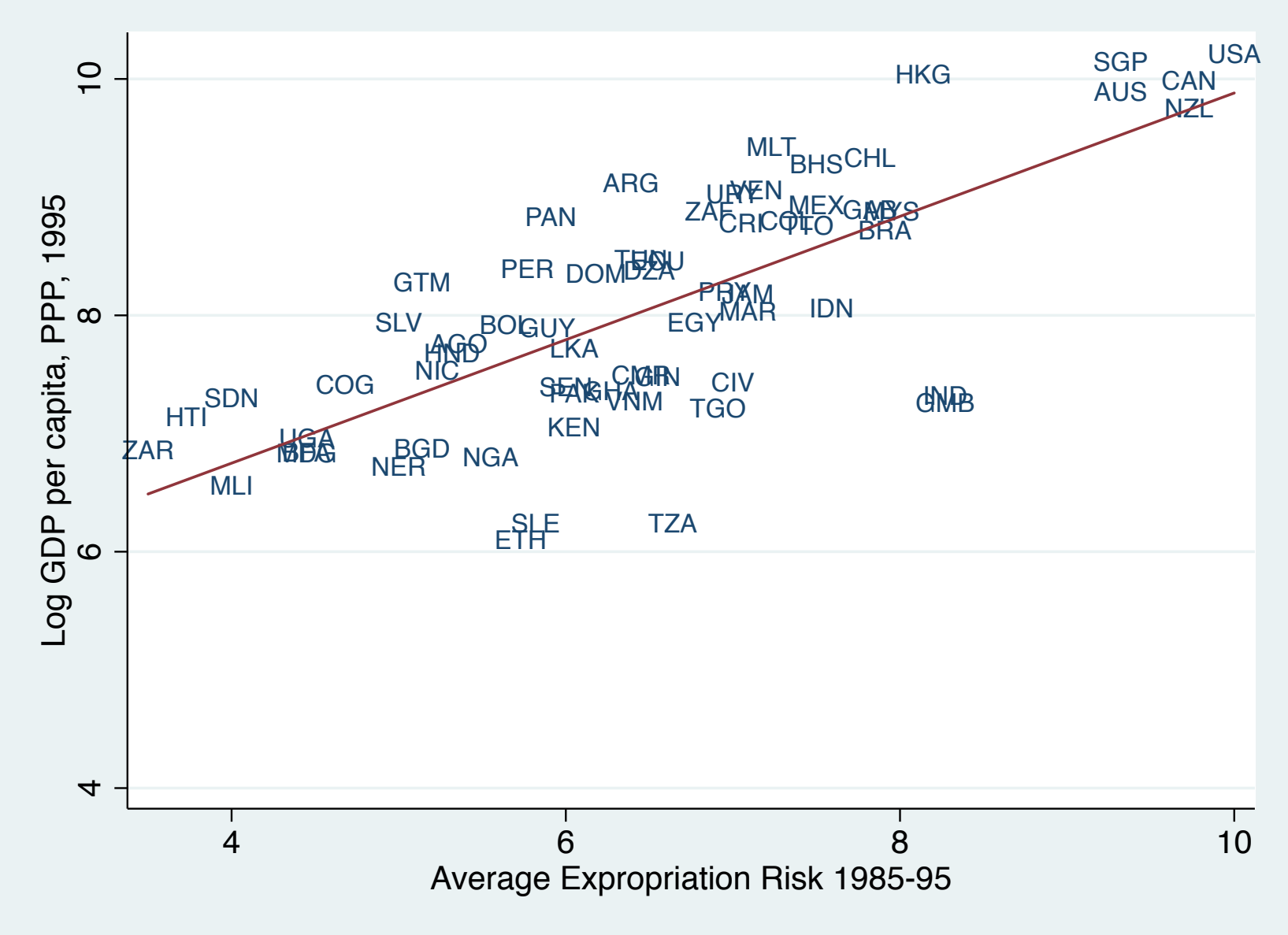
- The process through which we estimate the causal effect of endogenous regressor **(X)** or treatment variable **(D) on** the outcome variable **(Y)** using instrument **(Z)** is called **Two-Stage Least Squares (2SLS)**
- We can think of our endogenous regressor as having two parts:
  - A "good" source of variation: The variation in X or D that is exogenous, i.e. not influenced by the variable of interest Y
  - A "bad" source of variation: The variation in X or D that is caused by changes in Y. Reverse causality leads to endogeneity, biasing the OLS coefficients.
- How does 2SLS process help us leverage on the "good" variation while excluding "bad" variation?

The Colonial Origins of Comparative Development: An Empirical Investigation

By DARON ACEMOGLU, SIMON JOHNSON, AND JAMES A. ROBINSON*

We exploit differences in European mortality rates to estimate the effect of institutions on economic performance. Europeans adopted very different colonization policies in different colonies, with different associated institutions. In places where Europeans faced high mortality rates, they could not settle and were more likely to set up extractive institutions. These institutions persisted to the present. Exploiting differences in European mortality rates as an instrument for current institutions, we estimate large effects of institutions on income per capita. Once the effect of institutions is controlled for, countries in Africa or those closer to the equator do not have lower incomes. (*JEL* O11, P16, P51)

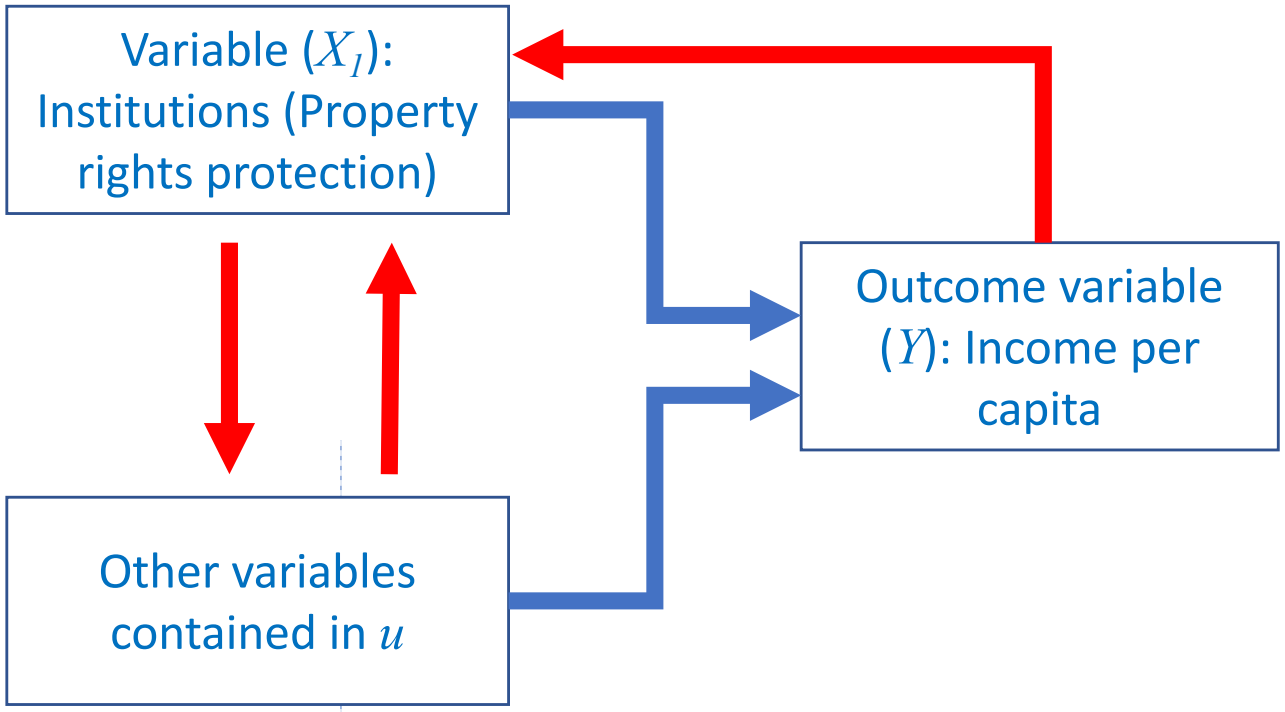
Institutions (Average Expropriation Risk) and Income per capita



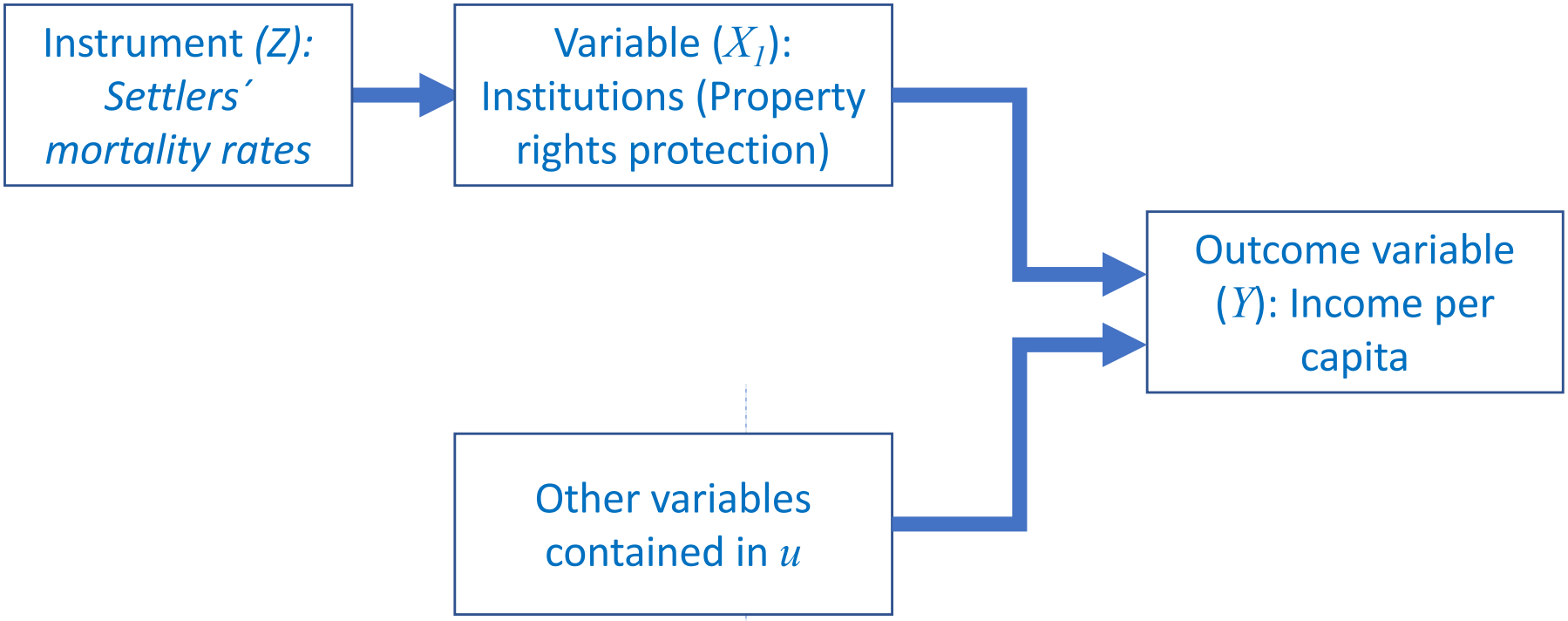
IV Example 1: Institutions and economic performance: What is the problem we are trying to solve?

Simultaneous causality:
X_1 causes Y, but plausibly Y can also be a driver of X_1

Endogeneity:
Explanatory variable is correlated with the error term



IV Example 1: Institutions and economic performance



First stage regression

- Is the instrument relevant?

In all three specifications the instrument (settlers mortality) is relevant, as it significantly associated to institutional development (protection of property rights). In cases (1) and (2) we can reject the null hypothesis of non-association between the instrument and variable X at the 99% confidence level (t=4.83 and t=3.67), whereas in (3) it is close to the 95% threshold (t=1.92)

. reg protection lmortality latitude asia africa other

	(1)	(2)	(3)
	(1)	(2)	(3)
lmortality	-0.61329	-0.51720	-0.34995
	(0.12694)	(0.14086)	(0.18160)
Abs(latitude ~90		2.00722	2.00057
		(1.32989)	(1.38339)
asia			0.46739
			(0.50160)
africa			-0.25669
			(0.40631)
other			1.04590
			(0.84187)
Constant	9.36589	8.55590	7.77345
	(0.61059)	(0.80828)	(0.94819)
Observations	64	64	64
r2_a	0.26179	0.27670	0.27324
rmse	1.26202	1.24921	1.25219

Standard errors in parentheses

IV Example 1: Institutions and economic performance

- **Second-stage regression**

Higher average protection to private property - as instrumented by settlers mortality – causes statistically significant higher income per capita. T-statistics of average protection:

(1) One unit increase in protection (as instrumented by settlers mortality) is associated to 92% higher income (t-statistic: 6.15)

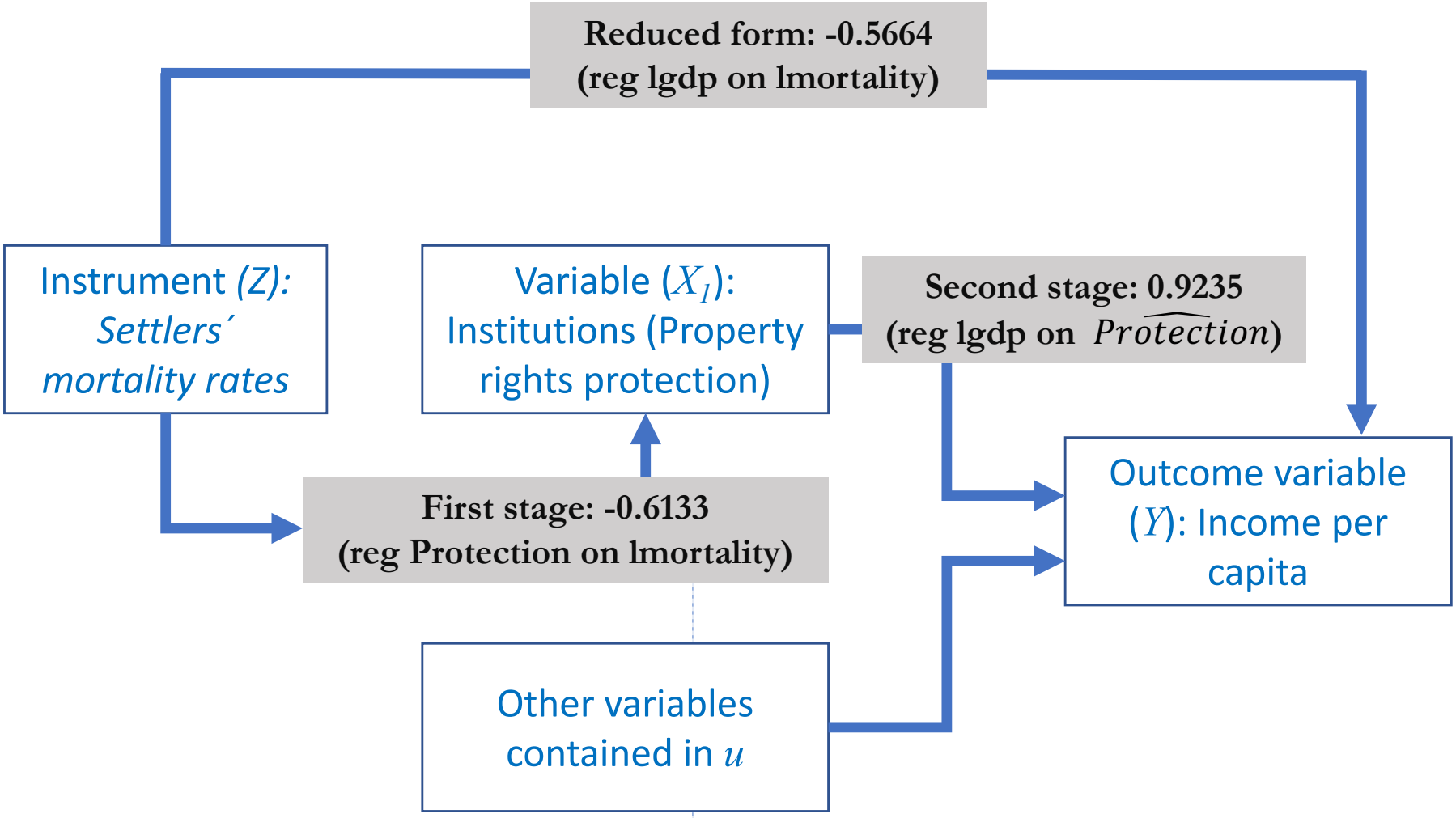
(2) t-statistic: 4.651

(3) t-statistic: 2.612

. ivreg2 lgdp (protection=lmortality) latitude asia africa other

	(1)	(2)	(3)
	(1)	(2)	(3)
Average Protec~o	**0.92352**	**0.96284**	**1.02786**
	(0.14995)	**(0.20703)**	**(0.39351)**
Abs(latitude ~90		**-0.50370**	**-0.94127**
		(1.25479)	**(1.51061)**
asia			**-0.99432**
			(0.45531)
africa			**-0.48546**
			(0.36718)
other			**-0.86717**
			(0.86165)
Constant	**2.04476**	**1.87979**	**1.92064**
	(0.98373)	**(1.20804)**	**(2.41212)**
Observations	**64**	**64**	**64**
r2_a	**0.20792**	**0.13143**	**0.08668**
rmse	**0.91425**	**0.94963**	**0.94954**

IV Example 1: Institutions and economic performance



IV Example 1: Institutions and economic performance

- Is the instrument exogenous?

- Can you think of why the instrument would not be exogenous?
  - What if Europeans decided to send their fittest, healthier and best equipped men to richer areas?
  - What if the disease environment at the time of colonization is related to the disease environment today, which in turn is associated with lower income per capita? To what extent can we isolate the disease environment back then from economic performance today?
  - What if Europeans today are an important share of the population of those countries where Europeans settled? In that case, their values and attitudes toward work may influence current income.

IV Example 2: Minneapolis Domestic Violence Experiment (MDVE)

- Instrumental variables can be used to estimate treatment effects in randomized experiments, when we suspect that actual application of treatment is influenced by the intensity of the outcome variable
- MDVE (1984):
  - Study to determine the impact of arresting batterers in the likelihood of reoccurrence of domestic violence within six months
  - Two treatments to suspected offenders:
    - Arrest (usually resulting in a night in jail)
    - Minor punishment (separation from premises for 8 hours and counseling intervention)
  - Treatment randomized by means of color-coded forms in report pads
  - Case of life threatening or severe injuries – felony assaults –excluded
  - Why would we be suspicious that random assignment was imperfect?

IV Example 2 MDVE: Treatment and results by group

# of cases	Delivered treatment		
Assigned treatment	Arrest	Separation / Counseling	Total
Arrest	91	1	92
Separation-Counseling	45	177	222
Total	136	178	314

% of cases	Delivered treatment		
Assigned treatment	Arrest	Separation / Counseling	Total
Arrest	98.9%	1.1%	29.3%
Separation-Counseling	20.3%	79.7%	70.7%
Total	43.4%	56.6%	100.0%

Reoccurrence: 17.8%
Reoccurrence in cases **assigned to** Separation/Counseling: 21.1%
Reoccurrence in cases **assigned to** where Arrest: 9.7%
Reoccurrence in cases **where** Separation/Counseling **was delivered**: 21.6%
Reoccurrence in cases **where** Arrest **was delivered**: 12.9%

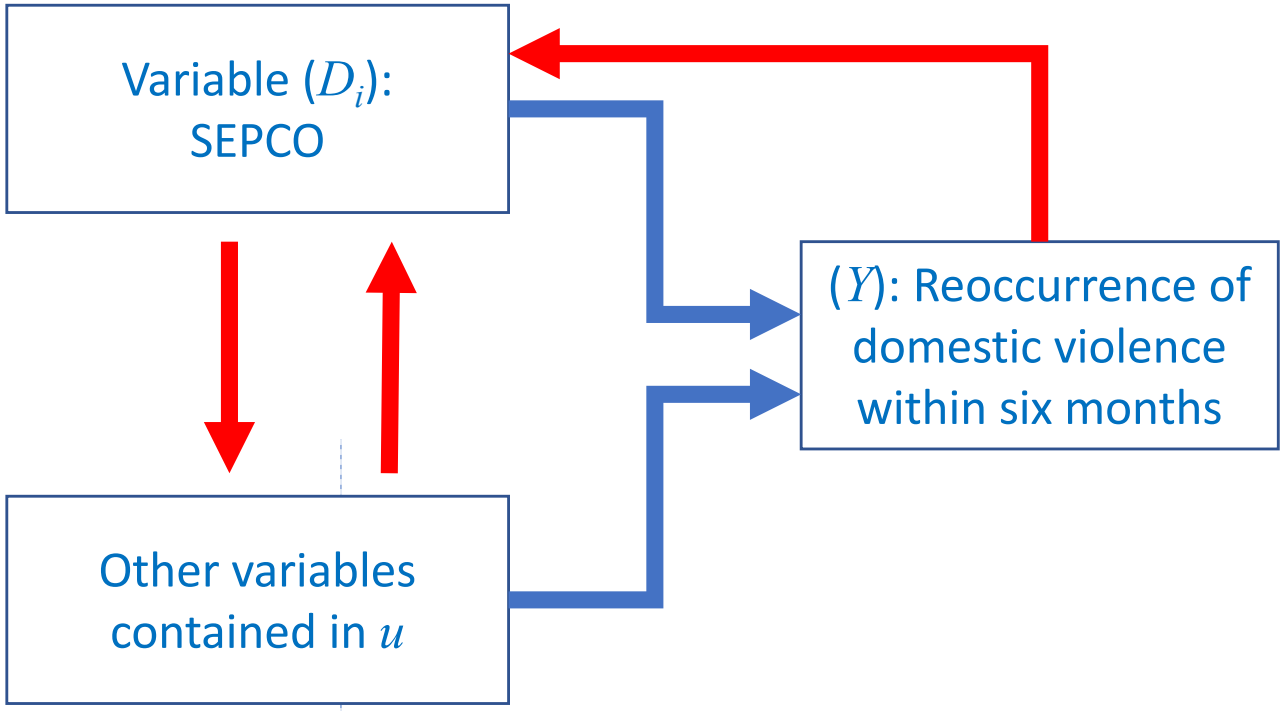
IV Example 2: Minneapolis Domestic Violence Experiment (MDVE)

- What would be the impact of SEPCO if the randomization protocol would have been followed?
- The impact would be given by the differences in probability of reoccurrence between **assignment to treatment** (SEPCO) and control (Arrest): 21.1% - 9.7% = 11.4% (Intention to treat effect or ITT).
- We could also estimate impact by the differences in probability of reoccurrence between **actual treatment** (SEPCO) and control (Arrest): 21.6% - 12.9% = 8.7% (Treatment effects on treated or TOT).
- What would that not be a good estimate of the impact of SEPCO in this case?
- ITT ignores the fact that some of the suspects assigned to SEPCO were arrested, or more generally, that the intensity of violence (presumably associated to the probability of reoccurrence) might have influenced the application of treatment.

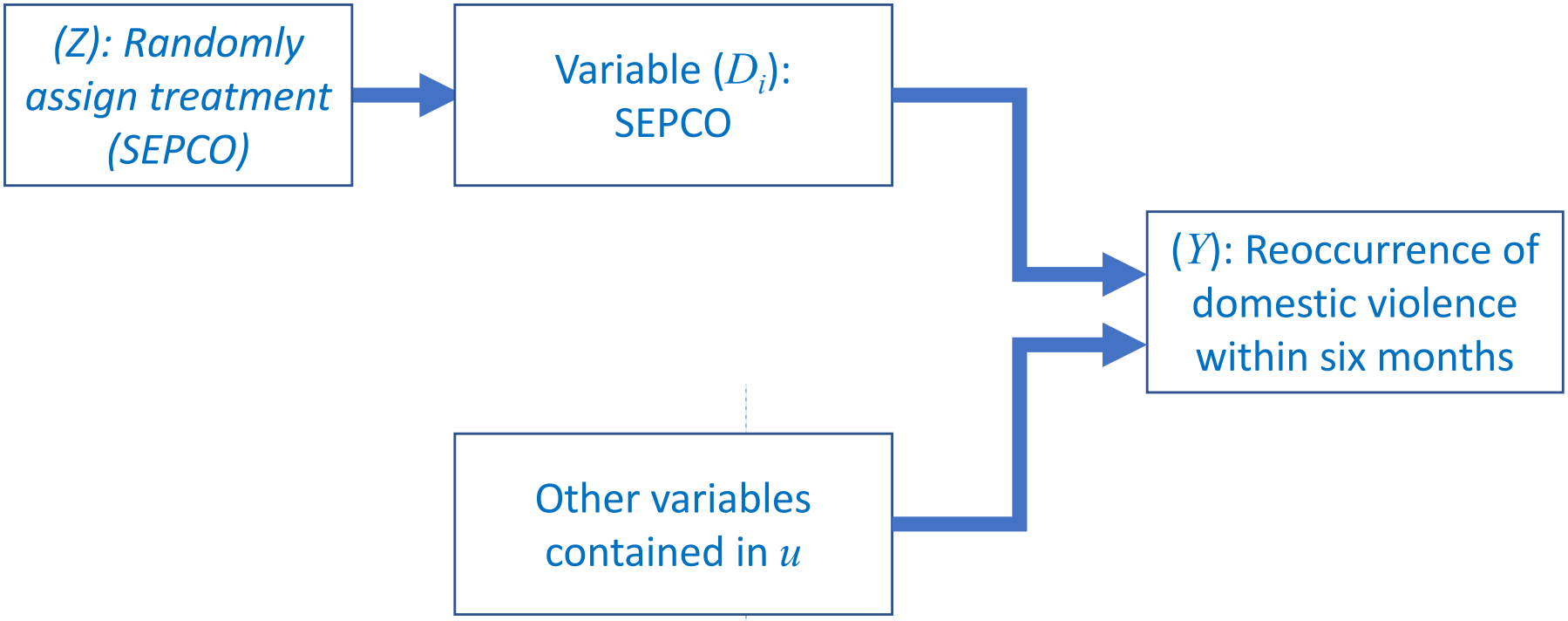
IV Example 2 MDVE: What is the problem we are trying to solve?

Simultaneous causality: ***In some case at police discretion, suspects were arrested even when random assignment called for separation/counseling***

Endogeneity:
Explanatory variable is correlated with the error term



IV Example 1: Institutions and economic performance



Instrumental variables: First stage and second stage regressions

First stage regression (regress D on Z):

$$SEPCO = \alpha_0 + \alpha_1 Assigned\ to\ SEPCO + \varepsilon$$

$$\widehat{SEPCO} = \underline{1.1}\ \% + \underline{78.6}\ \%\ Assigned\ to\ SEPCO$$

Reduced form (regress Y on Z):

$$Reocurrence = \gamma_0 + \gamma_1\ \text{Assigned to SEPCO} + v$$

$$\widehat{Reoccurrence} = \underline{9.7}\ \% + \underline{11.4}\ \%\ \text{Assigned to SEPCO}$$

Second stage regression (regress Y on $\hat{D}$, that is D instrument by Z)

$$Reocurrence = \beta_0 + \beta_1\ \widehat{SEPCO} + u$$

We will start Wednesday 21st class at this point