ECON 326: Economics of Developing Countries
TA Session 2

Vaidehi Parameswaran (Northwestern Econ)

January 2025

Today's Agenda

- Interaction Terms
- IV Application: Acemoglu, Johnson, and Robinson (2001)
- Stata: `merge`, `reshape`

Interactions

Interaction Terms

- Ignore endogeneity concerns for now
- Consider the following estimating equation:

$$Y_i = \alpha + \beta X_i + \epsilon_i \tag{1}$$

- If treatment effects are heterogeneous across observations, we want β to be indexed by i.
- Estimating equation (1) yields the **average effect** of X on Y.
- Suppose we want to control for the effect of W on Y.
- We can include an interaction term between X and W in the estimating equation:

$$Y_i = \alpha + \beta X_i + \gamma W_i + \delta(X_i \times W_i) + \epsilon_i \tag{2}$$

Interaction Terms

- Take conditional expectations of Y given X and W:

$$E[Y_i|X_i, W_i] = \alpha + \beta X_i + \gamma W_i + \delta(X_i \times W_i) \tag{3}$$

- So now the effect of X on Y is $\beta + \delta W_i$.
- If X is continuous,

$$\frac{\partial E[Y_i|X_i, W_i]}{\partial X_i} = \beta + \delta W_i \tag{4}$$

- If X is binary,

$$E[Y_i|X_i = 1, W_i] - E[Y_i|X_i = 0, W_i] = \beta + \delta W_i \tag{5}$$

Interaction Terms

- Reading a table:

Table A8: Heterogeneity according to self-control and external pressure (untrimmed sample)
(compare Table 8)
Dependent variable: Real monthly profits

	Pooled								
	Men & Women			High Profit Women		Low Profit Women		Men	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Cash Treatment	-11.22	8.436	8.236	25.10	28.80	-5.537	-8.763	13.15	8.787
	(33.03)	(15.16)	(16.18)	(21.65)	(21.68)	(8.819)	(9.886)	(32.63)	(36.82)
In-kind Treatment	39.64**	37.65***	39.19***	72.06**	71.38***	4.832	4.273	45.82*	44.32*
	(18.66)	(12.27)	(12.72)	(27.88)	(27.14)	(7.562)	(8.049)	(23.65)	(24.07)
Cash Treatment*Low Digitspan Recall	26.03								
	(35.43)								
In-kind Treatment*Low Digitspan Recall	2.213								
	(24.76)								
Cash Treatment * Lack of Self-control		-24.42**	-24.32**	-40.96**	-40.66**	0.917	1.228	-43.47**	-43.48*
		(10.04)	(10.36)	(19.40)	(19.49)	(7.771)	(8.098)	(21.87)	(22.21)
In-kind Treatment*Lack of Self-control		-7.817	-7.067	-14.88	-5.347	-0.932	-1.407	-11.00	-13.95
		(8.281)	(8.653)	(21.12)	(22.99)	(6.488)	(6.706)	(15.65)	(16.30)
Cash Treatment * Narrow External Pressure		3.200		-2.008		13.12		-12.27	
		(9.935)		(12.70)		(9.720)		(22.27)	
In-kind Treatment * Narrow External Pressure		-8.274		-26.18		2.870		-12.87	
		(11.18)		(32.60)		(7.424)		(16.22)	
Cash Treatment * Broad External Pressure			1.095		22.37		9.773		-19.64
			(15.08)		(14.61)		(9.028)		(32.36)
In-kind Treatment * Broad External Pressure			-7.240		-27.86		3.228		-1.823
			(14.22)		(33.43)		(6.281)		(22.97)
Observations	4,221	3,969	3,838	927	904	1,465	1,412	1,577	1,522
Number of firms	767	690	667	160	156	256	247	274	264
P-values for testing cash=in-kind for:									
Low digitspan interaction	0.527								
Self-control interaction		0.091	0.085	0.238	0.133	0.838	0.778	0.111	0.148
Narrow external pressure interaction		0.652		0.721		0.365		0.705	
Broad external pressure interaction			0.836		0.135		0.555		0.805

Notes: results from fixed effects estimation
Randomization occurred within matched quadruplets.
Robust standard errors clustered at the firm level in parentheses.
*, ** and *** denote significance at the 10%, 5% and 1% levels.

IV Application - AJR (2001)

Institutions and Development

- Institutions: set the "rules of the game" in a society
- So agents have to make decisions under these constraints
- Intuitive that having good institutions is better for development
- But causality is hard to establish

AJR (2001): Measurement of Institutions

- Institution quality measured as an index of protection against expropriation risk
- Constructed from political risk rating agencies
- Issues: subjective, volatile

AJR (2001): Correlation

- Consider the regression of GDP per capita on institutional quality

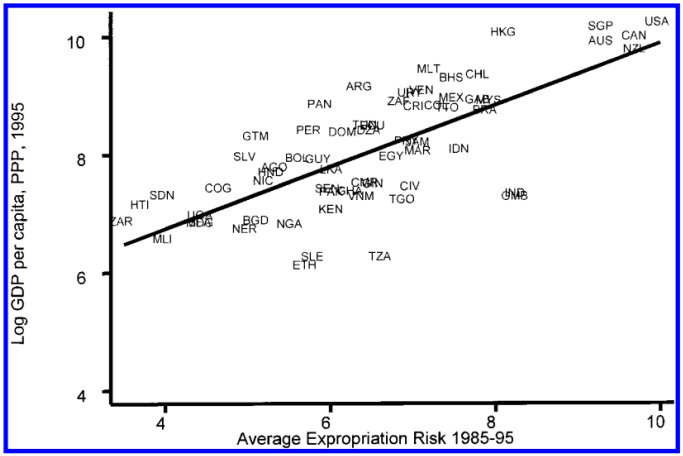


FIGURE 2. OLS RELATIONSHIP BETWEEN EXPROPRIATION RISK AND INCOME

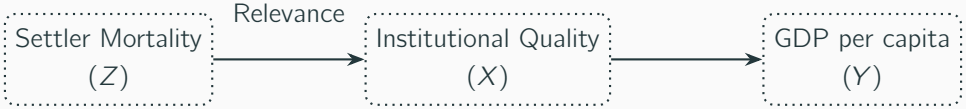
AJR (2001): Correlation

- But ... is it causal?
  - Reverse causality?
  - Omitted variable bias?
- Strategy: Instrumental Variables (IVs)
- Recall the conditions for the IV to be appropriate:
  - Relevance: $Cov(Z, X) \neq 0$
  - Exogeneity: $\mathbb{E}(Z, \epsilon) = 0$

AJR (2001): IV

- AJR use **settler mortality**
  - Relevance? Extractive institutions vs inclusive institutions
  - Exogeneity? Plausibly governed by geographical factors that are no longer relevant to GDP today
- We see some of these institutional differences even today

AJR (2001): IV



- We want to say X causes Y
- But we have endogeneity
- As an IV, we want Z to affect X
- And we want Z to affect Y but only through X

AJR (2001): Results

- First-stage result

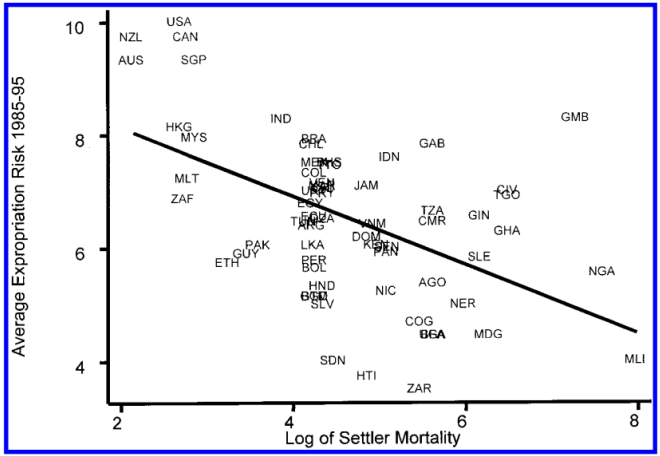


FIGURE 3. FIRST-STAGE RELATIONSHIP BETWEEN SETTLER MORTALITY AND EXPROPRIATION RISK

AJR (2001): Results

- Reduced form

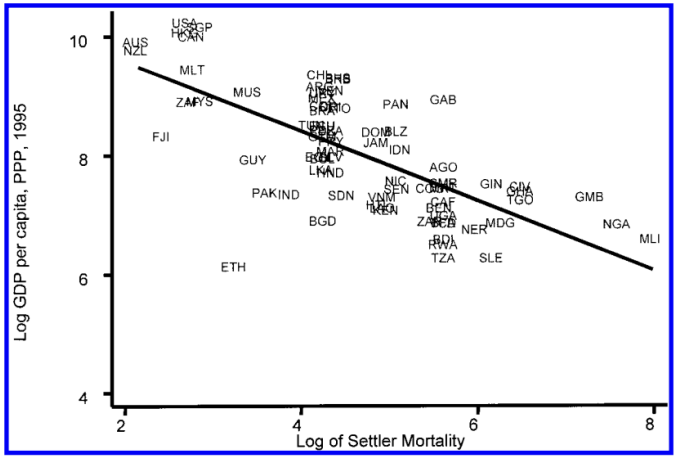


FIGURE 1. REDUCED-FORM RELATIONSHIP BETWEEN INCOME AND SETTLER MORTALITY

AJR (2001): Results

- IV estimates:

TABLE 4—IV REGRESSIONS OF LOG GDP PER CAPITA

	Base sample (1)	Base sample (2)	Base sample without Neo-Europes (3)	Base sample without Neo-Europes (4)	Base sample without Africa (5)	Base sample without Africa (6)	Base sample with continent dummies (7)	Base sample with continent dummies (8)	Base sample, dependent variable is log output per worker (9)
Panel A: Two-Stage Least Squares									
Average protection against expropriation risk 1985–1995	0.94 (0.16)	1.00 (0.22)	1.28 (0.36)	1.21 (0.35)	0.58 (0.10)	0.58 (0.12)	0.98 (0.30)	1.10 (0.46)	0.98 (0.17)
Latitude		−0.65 (1.34)		0.94 (1.46)		0.04 (0.84)		−1.20 (1.8)	
Asia dummy							−0.92 (0.40)	−1.10 (0.52)	
Africa dummy							−0.46 (0.36)	−0.44 (0.42)	
"Other" continent dummy							−0.94 (0.85)	−0.99 (1.0)	
Panel B: First Stage for Average Protection Against Expropriation Risk in 1985–1995									
Log European settler mortality	−0.61 (0.13)	−0.51 (0.14)	−0.39 (0.13)	−0.39 (0.14)	−1.20 (0.22)	−1.10 (0.24)	−0.43 (0.17)	−0.34 (0.18)	−0.63 (0.13)
Latitude		2.00 (1.34)		−0.11 (1.50)		0.99 (1.43)		2.00 (1.40)	
Asia dummy							0.33 (0.49)	0.47 (0.50)	
Africa dummy							−0.27 (0.41)	−0.26 (0.41)	
"Other" continent dummy							1.24 (0.84)	1.1 (0.84)	
R^2	0.27	0.30	0.13	0.13	0.47	0.47	0.30	0.33	0.28
Panel C: Ordinary Least Squares									
Average protection against expropriation risk 1985–1995	0.52 (0.06)	0.47 (0.06)	0.49 (0.08)	0.47 (0.07)	0.48 (0.07)	0.47 (0.07)	0.42 (0.06)	0.40 (0.06)	0.46 (0.06)
Number of observations	64	64	60	60	37	37	64	64	61

Notes: The dependent variable in columns (1)–(8) is log GDP per capita in 1995, PPP basis. The dependent variable in column (9) is log output per worker, from Hall and Jones (1999). "Average protection against expropriation risk 1985–1995" is measured on a scale from 0 to 10, where a higher score means more protection against risk of expropriation of investment by the government, from Political Risk Services. Panel A reports the two-stage least-squares estimates, instrumenting for protection against expropriation risk using log settler mortality; Panel B reports the corresponding first stage. Panel C reports the coefficient from an OLS regression of the dependent variable against average protection against expropriation risk. Standard errors are in parentheses. In regressions with continent dummies, the dummy for America is omitted. See Appendix Table A1 for more detailed variable descriptions and sources.

Figure 5: Source: Acemoglu, Johnson, and Robinson (2001)

Glaeser's Critique

- Glaeser et al. (2004) argue that:
  1. Institutions are mismeasured
  2. Z is correlated with Y through human capital, so exogeneity fails
- GLLS propose human capital as an explanation for cross-country income differences

Issue 1: Measurement

- Subjective measure of quality which could be affected by GDP itself
  - Richer countries may have higher ratings
- The index reflects outcomes
  - The index is lower in countries where expropriations have happened
  - But expropriations themselves are a function of the constraints and a choice variable
  - Two countries with the same set of constraints may have different outcomes because of different choices by leaders

Issue 2: Exogeneity

- AJR's key idea: if Europeans want to settle somewhere, they bring good institutions
- But they could have also brought with them good human capital
- So settler mortality and human capital maybe correlated
- Human capital could affect today's GDP
- Thus the instrument Z affects Y through the channel of human capital, not just X
- This is a violation of the exclusion restriction
- And thus a threat to identification

GLLS (2004)

- AJR's IV can predict human capital since the 1900s

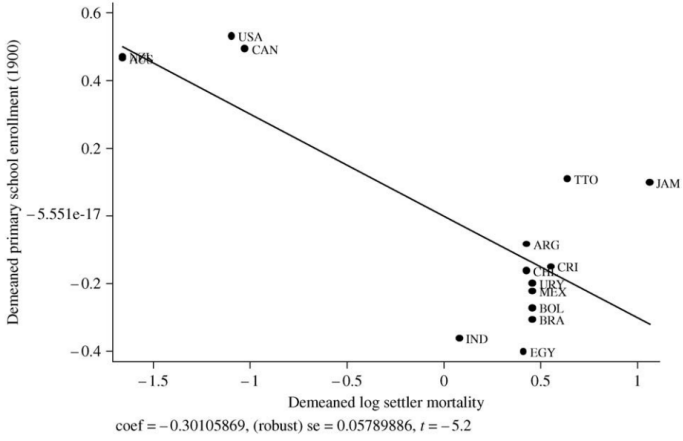


Figure 7. Primary school enrollment (1900) and log settler mortality.

Figure 6: Source: Glaeser et al. (2004)

GLLS (2004)

- So which is it? How can we square AJR and GLLS?
- Do institutions cause growth?
- Does human capital cause growth?
- Or is it both?
- GLLS attempt to identify the effects of both channels
- They find that human capital is a source of growth rather than institutions

Stata

Stata: `merge`

- Merging datasets is probably one of the most common tasks in data management
- And it's one thing Stata does easily and well
- Why merge? Typically to bring in additional variables
- What you need? 2 datasets and a common identifier variable in both
- The command: `merge TYPE IDVAR using filename`
- Example: `merge 1:1 id using filename`

Stata: `merge`

- Types of merges:
  - `1:1`: Unique identifier in both datasets
  - `1:m`: A merge where the first dataset has one observation for each unique identifier, and the second dataset has multiple observations for each unique identifier
  - `m:1`: The opposite of `1:m`
  - `m:m`: You NEVER want to do this
- Make sure the merging variable has the same name in both datasets
- It's also good practice to make sure it has the same variable type

Stata: `merge`

- Reading merge results:

Stata: `reshape`

- Suppose you have a dataset with variable X_ij
- So say i is the individual and j is the year, and X is income
- Such a dataset can be stored in two different formats in Stata:
  - Wide format: Each individual has a row and each year has a column
  - Long format: Each individual-year pair has a row

- What it looks like:

i	j		X_{ij}
id	year	sex	inc
1	80	0	5000
1	81	0	5500
1	82	0	6000
2	80	1	2000
2	81	1	2200
2	82	1	3300
3	80	0	3000
3	81	0	2000
3	82	0	1000

(a) Long format

i		..... X_{ij}		
id	sex	inc80	inc81	inc82
1	0	5000	5500	6000
2	1	2000	2200	3300
3	0	3000	2000	1000

(b) Wide format

Stata: `reshape`

- To reshape from long to wide, use the command `reshape wide` *varlist*, `i(i) j(j)`
  - i is the identifier variable
  - j is the variable that will be spread out (in this case, year)
  - *varlist* is the list of variables that will be spread out
  - you need to include all the variables that vary at the $i - j$ level in your *varlist*
- To reshape from wide to long, use the command `reshape long` *varlist*, `i(i) j(j)`
  - i is the identifier variable
  - j is a new variable that will be created (in this case, year)
  - *varlist* is the variable prefix (*inc* in this case)
- Since `sex` is common within i, it doesn't need to be included in the *varlist*

See you next time!