HANDOUT 20 – FINAL EXERCISE + CLOSING

AGENDA

- Final Exercise
- What have we learned
- Big picture lessons

FINAL EXERCISE

- What did you learn? What part of the exercise was more challenging?

- What empirical approach did you use? Did you change? Why?

- Do you think you have found solid empirical evidence to unequivocally assess the impact of improving high secondary school enrollment in lowering fertility rates?

- A simple approach:

FinalExercise

```
// Loading Data

clear all
use "/Users/msantos/Dropbox/wdi_extended.dta"

// Keep variables we want to use and rename

keep sp_ado_tfrt se_sec_enrr ny_gdp_pcap_cd countryname year
rename sp_ado_tfrt teenfertility
rename se_sec_enrr hsrate
rename ny_gdp_pcap_cd gdppc

// Keep countries with non-missing data

keep if (teenfertility!=.) & (hsrate!=.) & (gdppc!=.)
keep if countryname=="Argentina" || countryname=="Bolivia" || countryname=="Brasil" ///
|| countryname=="Colombia" || countryname=="Chile" || countryname=="Ecuador" ///
|| countryname=="Peru" || countryname=="Paraguay" || countryname=="Uruguay" || countryname=="Venezuela"

// Make a numeric version of a country variable

egen country=group(countryname)

// Run regression with increasing controls and store estimates

reg teenfertility hsrate, robust
est sto a1

reg teenfertility hsrate i.country, robust
est sto a2

reg teenfertility hsrate i.country i.year, robust
est sto a3

reg teenfertility hsrate i.country i.year gdppc, robust
est sto a4

// Output regression results into a table

esttab a1 a2 a3 a4 using table.csv, keep(hsrate) coef(hsrate "HS enrollment rate") ///
star(* .10 ** .05 *** .01)
```

  - South America:

	(1)	(2)	(3)	(4)
	Teen fertility rate			
HS enrollment rate	-0.490***	-0.435***	-0.108**	-0.0405
	(-24.59)	(-21.90)	(-2.34)	(-0.92)
N	315	315	315	315
Country fixed effects		X	X	X
Year fixed effects			X	X
GDPPC				X
t statistics in parentheses				
="* p<.10	* p<.05	*** p<.01"		

- East Africa:

	(1)	(2)	(3)	(4)
	Teen fertility rate			
HS enrollment rate	-1.147***	-1.130***	-0.592***	-0.913***
	(-10.88)	(-14.78)	(-4.09)	(-6.22)
N	260	260	260	260
Country fixed effects		X	X	X
Year fixed effects			X	X
GDPPC				X
t statistics in parentheses				
="* p<.10	** p<.05	*** p<.01"		

- Asia and the Pacific (excluding high countries):

	(1)	(2)	(3)	(4)
	Teen fertility rate			
HS enrollment rate	-0.362***	-0.362***	-0.480	-0.480
	(-6.31)	(-6.31)	(.)	(.)
N	45	45	45	45
Country fixed effects		X	X	X
Year fixed effects			X	X
GDPPC				X
t statistics in parentheses				
="* p<.10	** p<.05	*** p<.01"		

- Asia and Pacific adding data for each country:

	(1)	(2)	(3)	(4)
	Teen fertility rate			
HS enrollment rate	-1.542***	-1.508***	-0.619***	-0.365***
	(-14.71)	(-21.57)	(-4.84)	(-3.30)
N	250	250	250	250
Country fixed effects		X	X	X
Year fixed effects			X	X
GDPPC				X
t statistics in parentheses				
="* p<.10	** p<.05	*** p<.01"		

Why exercise of this type?

- We gave you this assignment because it mimics challenges you may face with your policy exercises and later in your careers:
    - Data is messy, with lots of missing data or ambiguities.
    - Tasks can be open-ended in terms of:
        - The subset of data to use (countries, years, variables).
        - The empirical strategy to employ.
    - The most convincing cases may use multiple approaches and weave a coherent narrative out of them.
    - Analysis may result in certainty or uncertainty. Both are fine.
    - Detailed analyses should be run but final product should tell a **succinct story: make it easy for your audience to read** (without oversimplifying).

WHAT HAVE WE LEARNED?

BIG PICTURE LESSONS

- **Correlation is not causation.**
- Much as we emphasize causal inference in this class, **descriptive (i.e. non-causal) research** is helpful for motivating policy questions.
    - What basic pattern in the data suggests a question is worth exploring?
- When framing a policy question make sure you **have a theory of change** and be clear about the **counterfactual**.
- People tell stories with data. **Is their story the right or only one?**
    - Be critical (but not cynical) consumers of research.
- No one study is definitive and very few findings apply to all contexts.
    - It's the (slow) accumulation of knowledge that counts.