
API-202B Empirical Methods II

Session #20: Final Exercise + Closing

miguel_santos@hks.harvard.edu
@miguelsantos12

Logistics

- I will return your graded final exercise next week
- Final exam Friday May 4th 2-5 PM, [rooms L230 \(A-M\) & L280 \(N-Z\)](#)
 - Focus on material since midterm
 - Review problem sets with solution sets
 - I will continue to run office hours and will run a review session (hopefully the week before the exam or early that same week)
- Today
 - Final exercise debrief
 - What have we learnt

 - Course evaluations

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)
```

A simple approach... South America...

	(1)	(2)	(3)	(4)
	Teen fertility rate			
HS enrollment rate	-0.490*** (-24.59)	-0.435*** (-21.90)	-0.108** (-2.34)	-0.0405 (-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"		

A simple approach... East Africa...

	(1)	(2)	(3)	(4)
	Teen fertility rate			
HS enrollment rate	-1.147*** (-10.88)	-1.130*** (-14.78)	-0.592*** (-4.09)	-0.913*** (-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"		

A simple approach... Asia and the Pacific (excluding high countries)

	(1)	(2)	(3)	(4)
	Teen fertility rate			
HS enrollment rate	-0.362*** (-6.31)	-0.362*** (-6.31)	-0.480 (.)	-0.480 (.)
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"		

A simple approach... Asia and Pacific adding data for each country...

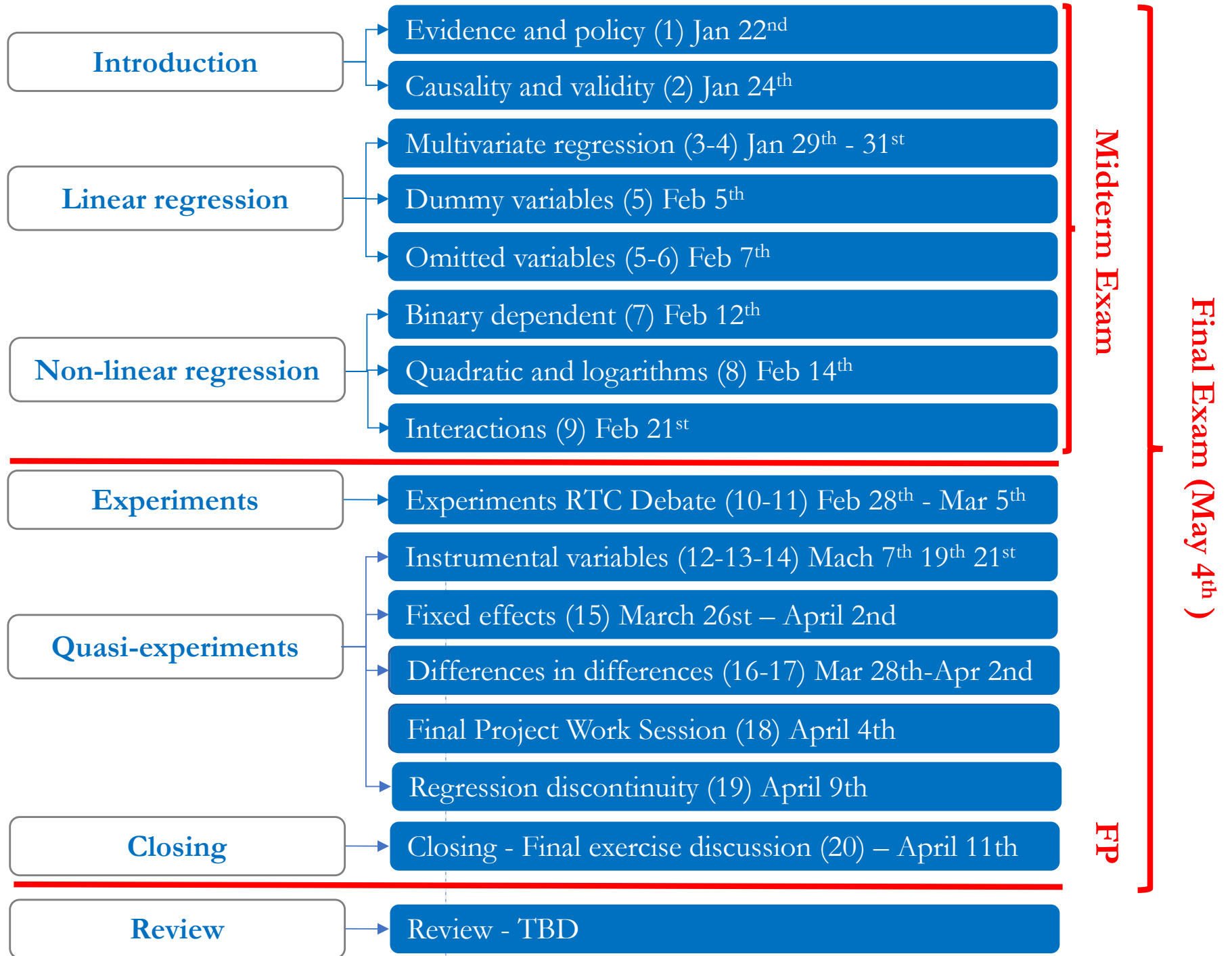
	(1)	(2)	(3)	(4)
	Teen fertility rate			
HS enrollment rate	-1.542*** (-14.71)	-1.508*** (-21.57)	-0.619*** (-4.84)	-0.365*** (-3.30)
N	250	250	250	250
Country fixed effects		X	X	X
Year fixed effects			X	X
GDPPC				X
t statistics in parentheses				
= " * p < .10	** p < .05	*** p < .01"		

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)

API202B



What have we learned?

What have we learned?

- Compelling **causal inference** involves eliminating **omitted variable bias**, to the extent possible
- **Randomized experiments** do this best (though can suffer from non-compliance, spillover effects, etc.)
- Without an RCT, quasi-experimental methods can help:
 - Use **regression discontinuity**: if there's a threshold
 - Use **difference-in-differences**: if policy change affects only some units
 - Use **fixed effects**: if within-group comparisons can eliminate major OVB
 - Use **instrumental variables**: if you find a great source of variation that is both relevant and exogenous
- If no experiment or quasi-experiment exists, use **multivariate regression** to control for potential major sources of OVB

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.

Two final words of advice...

Always be aware of the difference between what you see and what you want to see ...



‘I’m going to tell you my secret now... I see dead people... Walking around like regular people... They don’t see each other... They only see what they want to see... They don’t know they are dead...

<https://www.youtube.com/watch?v=QUYKSWQmkerq>

Rigorous research design can help you in decision making on public policy; and yet you can almost always “prove” any point you want to make by tweaking the research design.

Be honest to yourself: Learn to differentiate what you see from what you want to see.

Life will take you to places...

Wherever you are... that is where you are...

Be where you are

Thanks!

Course evaluations!

- Please take this seriously. The course you've just experienced was substantially improved by previous cohorts' comments. Pay it forward.
- Make your comments constructive. How can we improve?
- Write about any aspect of the course that matters to you, but I'd be particularly interested in:
 - Are lectures/handouts helpful?
 - How's the balance between theory and examples?
 - Do you want more or fewer case studies and readings of papers?
 - Do you want spend more or less time learning Stata?
 - Are there topics we should eliminate? Ones we should add?
 - For those who attended Fridays, give Evgenii useful feedback
 - Give Armando & Maggie useful feedback in main evaluation form