Development Economics 1

Jerry Montonen

Aalto University

October 7, 2021

Outline

- Feedback on homework 1

- Homework 2

  - Motivation and Context

  - Regression examples

  - Questions and Clarifications

- World Bank Data

- Basic Stata tutorial

Polls on homework 1

- How difficult was the homework 1 for you?

- How time consuming was homework 1 for you?

Feedback from Homework 1

Next, we will go through some general feedback on HW1 given by Miri

General comments

Remember: the imagined audience of the referee report is someone who has **not read the paper that you are refereeing**.

$\Rightarrow$ your summary of the paper, as well as your main and minor comments, should help the reader gain a complete picture of the paper, and should be possible to follow also without reading the paper.

The summary

- Should give an overview of what the paper does:

  - what is the research question

  - what is the research design - with enough detail so the reader can evaluate if that sounds like a good design/setup.

  - briefly describe the main findings

- It should also include a motivation, for why this paper is interesting and important.

- The summary should be **short:** about 0.5 page,

Summary: the methodology

Include enough detail so that a reader can understand if the methodology makes sense, while still keeping it short.

- The two key aspects of the methodology to describe in the summary are:

    ▶ what was the identification strategy?
    Field experiment (Deutschmann et al); DID (Zipfel)

    ▶ what variation was used?
    the variation between randomly assigned treat and control groups (Deutschmann et al); the variation across areas with different levels of exposure AND between cohorts with different levels of exposure (Zipfel)

General vs. specific comments

- Comments that could be made about any paper, e.g. "DID assumptions may be violated" are less convincing than comments that require close reading of the paper, e.g. "The parallel trends assumptions may fail if trends for eligible and ineligible cohorts are different between highly treated and less treated areas."

- Even better is if you can give an example about what could lead to such a violation: "since compared areas were different before the reform (highly treated areas had lower enrolment rate before") we may expect them to develop differently after the reform because..."

Constructive comments and arguments

When you point to a problem, always try to clarify:

1. Why is this a problem, i.e. why do I think that this issue poses a threat for either internal or external validity of the study?

2. Can this problem be solved or addressed in the context of **this paper**, and how?

This would (in theory) help the authors address your comments.

A constructive suggestion can be simple, such as suggesting the author to discuss a certain issue in a different way.

Advice such as "future studies should do X" is not constructive for how to improve **this** paper.

Specific details on Zipfel paper

Estimating equation on page 6 explains what is the main variation used in the paper to identify and isolate the effect:

$$Y_{idt} = \alpha_t + \delta_d + \beta \, Treatment_{id} \times Post_{it} + \epsilon_{idt}$$

the coefficient of interest is β and the variable of interest is the interaction term of *Treat* and *Post*.

- *Treatment* is related to the location: some districts have more intense treatment than other districts*

- *Post* is related to the *cohort*: some cohorts are eligible (young enough at the time when the reform is passed), other cohorts are partially eligible or ineligible.

- Note: in a basic DID regression we would also see the dummies $Post_{it}$ and $Treatment_{id}$, but these are captured by the location and cohort fixed effects: δ_d and α_t.

---

*the exact definition of treatment is related to the pre-reform school completion rate in the district, see page 6

Homework 2

- Question 1 is about identification strategies, interpreting regression coefficients and understanding when results are biased or not.

- Question 2 is a bit empirical, and requires you to discuss the results

Motivation and Context

- The impact of antiretroviral treatment to combat HIV on earnings is something that has been studied by several researchers.

- Tompsett (2020) studies the issue on a macro level

  - She finds out that an expansion of ARV treatment leads to economic growth

  - Tompsett uses the sharp fall in ARV prices in 2001 as a source of exogenous variation in her IV approach

Table 4

Impact of increasing ARV coverage on life expectancy and growth in per capita GDP.

	Main sample				Sub-Saharan Africa			
	OLS (1)	OLS (2)	OLS (3)	IV (4)	OLS (5)	OLS (6)	OLS (7)	IV (8)
Panel A: Log life expectancy								
% of population	−2.29*	5.57***		6.67***	1.39	4.96***		6.10***
treated with ARVs	(1.21)	(0.76)		(1.26)	(0.91)	(0.74)		(1.34)
HIV Prevalence$_{2001}$ ×			4.85***				4.72***	
global ARV coverage$_t$			(0.52)				(0.62)	
First-stage F-statistic				116				98
Hausman test p value				0.047				0.094
N	2146	2146	2160	2146	926	926	936	926
Panel B: Change in log GDP per capita								
% of population	0.24**	1.14**		1.40**	0.36***	0.97*		1.25*
treated with ARVs	(0.11)	(0.52)		(0.69)	(0.13)	(0.52)		(0.73)
HIV Prevalence$_{2001}$ ×			0.99**				0.94*	
global ARV coverage$_t$			(0.44)				(0.49)	
First-stage F-statistic				124				106
Hausman test p value				0.365				0.401
N	2222	2222	2236	2222	964	964	974	964
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes

Note: Coefficients from regressions of outcome variable on listed variables. Data from 1990 to 2014. Main sample consists of 90 low- and middle-income countries (39 in sub-Saharan Africa). Data are from 1990 to 2014. Controls comprise year and country fixed effects, and country-specific linear trends. In columns 4 and 8, % of population treated by ARVs is instrumented with predicted ARV therapy coverage, equivalent to global coverage of ARVs interacted with HIV prevalence in 2001. Hausman test tests exogeneity of observed ARV therapy coverage. Standard errors clustered by country and shown in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.

Motivation and Context

- Increased health spending has been found to decrease child mortality in developing countries.

- Examples of studies that find such results are Novignon and Lawanson (2017) and Nyamuranga and Shin (2019).

- One issue with these studies on this area is that they often tend to be descriptive

  - The above studies use fixed effects and GMM regressions on large panel datasets

  - Are the results really reliable then?

Regression Examples

- In all economics, you have to know how to interpret regression coefficients in models

- I will now go through two examples of regression equations, which will (hopefully) help you in solving the exercises.

Regression Examples

- Let's assume that you want to study the effect of having a tractor on agricultural output in developing country X. To do this, you have access to a panel data set including many variables.

- To do that, you run the regression:

$$Y_{itr} = \alpha + \beta_1 Tractor_{itr} + \beta_2 Land_{itr} + \beta_3 HHsize_{itr} + \theta_r + \gamma_t + u_{itr} \quad (1)$$

  - Y_{itr} is the agricultural output for household i in year t in region r

  - $Tractor_{itr}$ is a dummy taking value 1 if the household has a tractor in year t

  - $Land_{itr}$ measures the amount of land household i has in year t

  - $HHsize_{itr}$ measures the number of people in household i in year t

Regression Examples continued

$$Y_{itr} = \alpha + \beta_1 Tractor_{itr} + \beta_2 Land_{itr} + \beta_3 HHsize_{itr} + \theta_r + \gamma_t + u_{itr} \quad (2)$$

- θ_r is region fixed effects

- γ_t is year fixed effects

  - What do these imply?

    - They control away all effects that are region and/or year specific

- u_{itr} is the error term.

What problems do you encounter when running this regression?

Regression Examples continued

- Let's say that we would additionally have data on average rainfall (over the sample period) in each region. Would adding this make the model better?

$$Y_{itr} = \alpha + \beta_1 Tractor_{itr} + \beta_2 Land_{itr} + \beta_3 HHsize_{itr} + \beta_4 rain_r + \theta_r + \gamma_t + u_{itr} \tag{3}$$

- What if we had data on average rainfall for each year for the whole country?

- What if we had data on rainfall per region per year?

Regression Examples continued

- Now what if some organization held a lottery, where the winning farmers would receive a tractor?

- How would the reliability of results change?

- What about external validity?

- Spillover effects? Would neighbours sharing (won) tractors be a problem?

Questions and Clarifications

- Now is the time for you to ask about everything that is unclear for you regarding homework 2

- Remember, the deadline is **Friday 15.10 at 6pm**

- Please submit the assignment **in pairs**

  - You still have until 1 pm today to sign up for the pair matching if you're missing a pair

World Bank Data

- Next, I will give a short introduction to world bank data, where to get it, how to download it, find definitions of variables etc.

- The world bank is a very useful and reliable data source.

- Check it out and play around with the data to get a hang of what the possibilities are

Basic Stata Tutorial

- I will now talk about loading data into stata and calculating correlations

- For those of you familiar with stata, this will be easy. Feel free to leave if you know the basics

  - However, I will stay for a while after this to answer questions