[SQUEAKING]

[RUSTLING]

[CLICKING]

ESTHER DUFLO: Today I want to switch back to empirical-- to examining empirical things. You guys have read my paper. I'm going to talk a bit about that. And I'm going to talk also about it on Monday. So some of the questions that you had-- many of the questions, comments you had were on the returns to education piece. And those we'll discuss on Monday. But there are a number of things that come up that I should be able to address today.

So if you remember, we ended up last time by saying-- or we at some point last-- on Monday, we had this equation that the level of education chosen by a family for their kids is going to be a function of the cost of education, the marginal cost for them of acquiring one more unit of human capital, and of the returns to education, both the immediate financial return in the form of wages as well as the human-- how is it going to acquire the human capital of the next generation.

So these are all things that people have looked at. So we should expect people to respond both to costs and benefits.

So the cost piece seems relatively straightforward. One would look for a setting where cost to education direct or indirect went down and then look at whether people respond by educating their kids more.

And then, the benefits is whether people respond to returns to education. That might be a little bit more tricky. And we are going to do a bit of that, both the actual benefits and the perceived benefits, which, of course, is in this equation. Implicitly, something we hardly discussed but just a little bit last time is whether the perceived benefit are the same as the real benefits.

So let's start with the cost which are the straightforward piece. So there are a number of papers that look at the reaction to the direct cost of education. So in particular, but this is a bit of a biased review of the literature, because it has mostly my papers.

[LAUGHTER]

But I have a paper with Pascaline Dupas and Michael Kremer where we provide the school uniforms for kids entering grade seven or grade-- yeah, grade six, or seven, or eight. And we found a pretty large effect there.

We then have a literature that we'll discuss in much more detail. We have a paper on the secondary school to education, which are much larger than the school uniform, which we'll discuss in detail on Monday and you're going to read, which again show pretty large impact.

In a sense, what is-- and Michael has other papers with other people on various forms of scholarships for schools. So this literature is interesting, in particular the uniform literature. Because on the one hand, you're seeing effect-- impact of paying for uniform on schooling. So that would suggest that people are appropriately responding to costs.

On the other hand, these impacts are really large-- really, really, really large. So you get, for example, in our uniform paper you get reduction in dropout of a few percentage point. And the cost of a school uniform is $5 or so. So which even for a Kenyan family is not that big.

So that is-- on the one hand, you get the magnitude rate. On the other hand, it is not really, I would say, a vindication of this model because the impacts are too large. In the sense that if the only reason why the cost to education enter is you're comparing them to the interest rate, then clearly-- or not to-- I mean, to the benefits and appropriately discounted by the interest rate, and that seems way too large. These are indicating huge credit constraint that they can't even $5, or something else is going on. Yeah.

STUDENT: When you were designing this experiment, why did you think about uniforms [INAUDIBLE] and cost of [INAUDIBLE]?

ESTHER DUFLO: Everybody says it's a big deal. Everybody says it's a big deal. In fact, we got to uniform-- this is a paper where we looked at HIV. We were very interested in testing the impact of the government HIV school intervention.

So the government in Kenya was introducing a new program for HIV, which we actually were reasonably skeptical about. This is called the ABCD Strategy, which was followed at the time by most countries in East Africa. That stands for abstain, be careful, use a condom, or you die.

[LAUGHTER]

That was relatively a kind of fear mongering among-- in this order. So abstain is the best. Be safe, use a condom, or you die. So this was an example of what's called a avoiding all risk-- risk avoidance strategy as opposed to risk mitigation strategy.

So this program was introduced among upper primary school students. This is grade six to eight in Kenya. They were interested in evaluating and we were interested in evaluating it. But at the same time, we were quite skeptical. And we wanted to compare that with the strategy that is just keeping girls in school as a direct protective measure as opposed to trying to teach them something on HIV when they are in school. So we wanted to have that.

And it was pretty obvious to us that for kids their age, the uniform is a big constraint. They cannot go to school without the uniform. The school send them back without the uniform. Initially, when we started the project, conceiving of the project, there were school fees. So we were going to pay for the school fee. And the school fee went away.

But we realized that the uniform were still a barrier. So that's why we did the program. Even though, if you think about it from the comfort of your office, people shouldn't really be responsive to the cost-- to the presence of the uniform. But as it turns out, they are.

So that's how we came to that. It was less of-- and the results are pretty interesting in the sense that keeping kids in school, keeping girls in school does reduce HIV. That program doesn't work. And in fact, that program undoes the effect of keeping girls in school.

So it is better to just pay for the uniform and do nothing else. That's worse. It is worse than trying to teach them the ABCD program while they're in school. For reasons that we probably will get into and we talk more about health. So that that's how this came about.

So this is something that everybody talks about, the school uniform, even though this is not consistent with this model. But that's probably a failure of the model as opposed to a failure of people's thinking.

So that's the first set of costs. So I would say, just as a vindication of this model, it's not very-- it's a bit half-half. Because on the one hand, you see people responsive to cost. On the other hand, they're way too responsive.

Then there are the indirect cost, which is the opportunity cost of being in school. So while you're in school, you don't make money for the family. And it is believed, in particular for slightly older students, middle school, high school, that it prevents some kids from being in school.

And this is one of the motivation of conditional cash transfer programs which were introduced initially in Latin America, which has a component that is conditional on sending kids to school. And the idea is that that's actually a fair amount of money that people get from the conditional cash transfer. And it generates-- so as long as your kids goes to school regularly and is not absent too much-- is enrolled and not absent too much-- you get a transfer monthly.

So that sort of compensate for not having their earnings if they worked on a market or on a farm or something like that. So that's also a way of manipulating the cost of education by creating a benefit that offsets an existing cost. So that's a form of indirect cash transfer.

There, again, I would say the results-- the evidence towards in favor of this model are relatively-- is mixed. Because on the one hand, people are very responsive to conditional cash transfer. They have now been evaluating in dozens of countries, starting with the evaluation in Mexico, which was one of the first big prominent city in development.

It was replicated in lots of countries. And maybe bizarrely enough, because the first one was done with an experiment. Everybody thought they had to have their experiment too. So it was experimented in many, many, many places. And you tend to always find an impact on education, particularly for these older kids. So that would be good that people respond to the cost of education.

However, they also responds when the cash transfer are unconditional. And there, again, remember, in our model an easier-- just a lump sum transfer should not have an impact. Well, you could say maybe it's credit-constrained or something like that. But it's certainly not directly a response to the cost of education. So people are just about-- in some studies you find them somewhat more responsive to conditional cash transfer. In many studies, there is no difference.

So we did one in Morocco, for example, where we found that conditional and unconditional transfer are more or less equivalent. With an unconditional cash transfer, possibly a little more effective, actually, than the conditional cash transfer. So again, that's kind of mixed. While people respond to things that are meant to send them to school, maybe meant to reduce the cost. But they seem to do that even though the transfer is not directly affected to the cost.

And then, of course, there is another form of reducing the indirect cost is reducing the distance to school. So some would ask, how do you think about this intervention-- the school construction intervention? And one very natural way to think about it is you reduce the cost of going to school by making the school very nearby so it's much easier to go there for the child and for the parents. So that's the idea of my paper in Indonesia.

So someone asked how I came about with this idea. So maybe that's relevant. Second-years I've already said that because I mentioned it in the paper writing class. But that turned out to be my job market paper although I didn't know it was going to be my job market paper.

But I was sitting in development and in labor. And in labor we only talked about instrumenting for education, saying it's so important to instrument for education because there might be ability biased people who are more likely to go to school are also maybe more competent in various ways. And therefore, they'll earn more money anyway, even if they don't go to school. So that's the ability bias.

I thought, well, in development that seems worse. Because it's not just your ability bias that, in fact, doesn't seem to really be there in rich countries. You don't seem to have a-- to change the return to education much when you control for this selection bias.

But I thought, probably in development things would be quite different. Because there are many reasons why kids would go to school, like in particular, the connections of their parents, the money of their parents, that might be related with someone's ability to get a job later. So I thought the ability bias would have to be larger.

And then I was very surprised that there was no study using instrumental variable in development. So I thought, someone should write this study.

[LAUGHTER]

And I was trying to think about what would be a good instrument, and thinking about distance to school being a good instrument. But I'm thinking, well, that's not going to work because if people are closer to school, or probably are in more urban setting. So what you really need is a change in the distance to school. Well, where would that come from? Well, if a country started building a lot of schools, that would reduce the distance to school. Is there a country that has built a lot of schools?

And so, in the back of my mind, I was looking for that country, which had built a lot of schools. Now, separately at the time, we didn't have a second-year paper. But we had to write an econometrics paper. And I had written mine on Indonesia because there was a good data set that a lot of people use to do any number of things called the Indonesian family life survey.

So I started with Indonesia. And I went to the library. And they have all the World Bank reports. One of you found this report that I found-- 1989 report that-- but they have a bunch on all of the countries and what they are doing every year. And I said, let me start with Indonesia since there is this data set which I already know.

And then I found that. I said, that's great. That's a lot of schools. That should reduce the number of schools a lot. So let me see whether I can do something with that. And then I quickly realized that, so first of all, I had to found a number of schools. And I had to find a large data set.

I figured that such large data set exists, maybe from that same report, that they were using it, this kind of large census and inter-census survey. So I knew that this data set existed. I didn't know that the number-- the data and the number of school existed. But I thought, probably it's somewhere. It has to be in some library in Indonesia.

So I went to Indonesia during the summer. And I hunted for the data which I knew existed. That's the outcome data. I had to go to the statistical office to find it. I think-- so anyway, I paid some money to someone, and they gave me the data.

[LAUGHTER]

And I didn't poke too much. But that happened.

[LAUGHTER]

Now it's much easier. The data, since then, the data from Indonesia is very, very easily available. The Australian National Agency Fund gave them a lot of data to make their data available. And then I went to the library. I knew who had implemented this program. There was a planning department. I went to the library of the planning department. I asked them, would you have this data?

And they gave me books which showed the number of schools that were built each year, the plan of the schools, et cetera. I made some photocopy, and I was done. And then I had two more weeks to spend in Indonesia. And I got very bored. That was the end of that. So that's how I came to that.

Everything lined up quite nicely. It doesn't have-- anyway, it's not always like that. There is any number of places where this project could have collapsed. But in this instance, I was lucky at every step.

Interestingly, I was initially interested-- so I had no idea how good an idea it was. I just thought, this is-- I gave this advice to someone yesterday. And I think it's good advice, that it's very hard when you start working on a research project to think about whether that's going to be an A-plus contribution, or a B-minus contribution, or no contribution at all. It's hard for you. And it's even hard for people who talk to you.

So I think the best guide that I have found is to write the paper that I wish I had read. I want to write the paper that I wish to read. And then you know that at least one person will like it.

[LAUGHTER]

And a pretty important person, because that's yourself, and you're going to spend a lot of time with that paper. And the rest works itself out. And I just had no idea that in and of itself doing all of these step was so key. It seemed to me very easy. I thought of it as a scales, like doing scales, to take a strategy that is very, very well-established somewhere and transpose it to another context.

And in fact, in my mind, it was going to help me think about non-parametric returns to education. I had read a paper which was very new at the time by Whitney Newey about estimating non-parametric system with triangular-- with this type of situation where you have an instrument for the variable. So I thought, that's where I'm going to get. And at least that's going to be something new.

Now, it turned out-- for a reason we're going to discuss very soon, or if not today or on Monday-- that it was completely inappropriate to do that. Not only that, but I could have intuited it from the beginning that if I had an instrument that affected people going to primary school, I was not going to be able to use it to extrapolate to the entire distribution of the returns across the entire education spectrum. So I had thought about it, I would have realized this was not-- this part of the project couldn't get done.

But it turned out that even without this part of the project that that was a good paper, or at least understood to be a good paper, got me a job, and so on and so forth. So all that to say that it's hard to-- it was a nice project because the steps were so clear to me. And then, I was lucky to get them done. I was not stopped by the fact that I didn't know if I would find the data. I would take some risk.

And then even though the steps were so clear, it was simple at the time, it was sufficient. Now, of course, everything needs to have more bells and whistles. But I think the general principle of not being discouraged in advance by this is too easy or this is too obvious still holds. So that's for the autobiography of related to this paper.

So it's a difference. In difference, we are going to see a lot of that. You're going to see a lot of that in your career. If you're an empirical person, at some point you'll probably estimate some version of an event study graph or a difference in diff and diff.

In the simpler setting, it's pretty straightforward. You have two groups and then two periods. And then the group gets treated in the post-period and not in the pre-period. And your identification assumption can be written in this way. Which is if you think of it in terms of your potential outcome and treat it, it's a function-- put a possibly very flexible function of anything could happen with respect to the group. It could be related to-- sorry, to the group, to the time period.

But what's not there that's going to help you for identification and [INAUDIBLE] education. What did I propose from this equation that determines-- yes.

STUDENT: Are there interaction between belonging to the group and--

[INTERPOSING VOICES]

ESTHER DUFLO: Exactly. So we assume that maybe people who are younger will get more education because they are younger. Countries is richer. People who are in some region will get more education than others. But there is no-- the younger people in the place that are more educated would not otherwise have been even more or even less educated. So that's the assumption we are making.

And that's the key identification assumption. And in a lot of difference in difference application, the trick is going to both justify this assumption from a priori principle and gather as much evidence as possible that the assumption is, in fact real. Once we have that, then we're in good shape.

So again, if we have two groups, two period, then the diff and diff estimates are simply-- or so the object that we are trying to estimate is the difference in mean between post and pre for the treated group minus the same thing for the control group. And there are the identification assumption that I laid down. That's going to give you the impact-- the average impact of the program.

Of course, we don't have these things. So we replace them by population averages or by sample averages. And the definitive estimator is now going to be-- this should be a hat there-- going to be the average treated group post minus treated group pre minus control group post minus control group pre.

I would like you to do three lines of algebra when you go home which will convince you that this and this is numerically equivalent to estimating an OLS regression on a group-- a treated dummy, a constant treated dummy, a post dummy, and the interaction of the two.

With undergraduates, I make the calculations. But I'm going to assume that you can make it. That's from the OLS formula. So it takes the lines of algebra. But it's interesting-- it's good to have in mind. You really need to know. And sometimes we ask this question in the exam and not everybody gets it.

You really need to know what this is. This is the difference between post and pre, beta 1. This is the difference between treated and control in the pre period numerically. And this is the diff and diff. This is something one needs to know in life because you get to.

So in our setting, we have a relatively swift school construction building program financed by the oil boom in 1973. So someone asked how it came about that they decided to use their oil money to do that. What's the political economy of that?

So the political economy of that is they were-- Suharto came to power after a pretty bloody civil war. Millions of communists were killed. The country was not in good shape either politically, or economically when he took and consolidated power. And his view for fixing that was to try and create a national identity, which was uniform for the whole country.

And to do that, he invented a whole ideology called the Pancasila. Really promoted the Bahasa Indonesia language, which interestingly is the language of no one in particular. It's a group-- it's a trader language that's not spoken or was not spoken, in particular, by the main ethnic group. And he thought schools would be a good way to diffuse that ideology and to diffuse the language, and therefore to create a sense of national unity after what they had experienced.

And before that, Indonesia is an oil-producing country. But before that, the oil proceeds used to go to the places that produced the oil. And another thing that he thought that he would consolidate both his power and the whole place was to use this proceeds in a more equitable way.

So he decided that all of the money from the oil had to be redistributed by the center in social programs. And one of-- and the school construction program. They are called this in-pres, in presidential instruction, by presidential instruction. So the in-pres-- the education-- the primary school program was the first of those. Then there was a water and sanitation program.

So but when they started, it was very small, because they started to-- the first year of this was 1972. And then the oil boom came. And then it became, I think, much bigger than what they had anticipated. And they stuck with it, both in terms of how to allocate the rules, how to allocate the schools, and to continue using all money to this purpose.

So it's very different to Brazil, for example, where the oil money is being spent in the provinces that extract it. Or in fact, a lot of the oil is in the sea. So it's extracted-- it's used in the provinces that are in line to the sea. And that money-- there is a whole literature showing that that money is not very well spent by those places. So that's the political economy of that.

Then they decided that they were going to allocate the schools because the objective was to get to 85% enrollment everywhere. So they were going to allocate the school as a function of the difference between 85% and wherever the place was. They had their census in 1971. So they were basically use that to say-- to vertically linear increase to 85%. And in principle, it should have stuck at 85%.

That's another thing that I tried to use and didn't use to see whether after 85% it should have flattened. And so, one could use that for a neater identification. But it didn't quite-- it wasn't it was too smooth, so it didn't work.

So with that, just it depends on the pre-campaign enrollment. That gives me my difference in difference. So I'm going to define a group of treated people which are the young guys, and then treated in control region. To start with, we can say, well, let's be low above the median. And then we can do a simple diff and diff for year of education.

So these are sample average for the high program intensity and the young people, high intensity and the old people. The first thing we note is among the old people there was the education level, this is years of education attained by adults who were kids during this period is lower in the high intensity region.

That makes sense, because they were putting more schools in places that had less education to start with. And they are still lower in for the young kids, but less so. So the difference is positive although it's insignificant. So this is illustrative, but it's not very precise.

So that's good. So then now we can look at-- we can be a bit concerned that in even in the absence of the program, the young cohort would have converged maybe to the old cohorts because-- go back to our basic model. Things converge. There should be one steady state to where everybody is going and it's within the country. So we would expect convergence anywhere. So am I just capturing that?

So one easy way to do that is to compare the old to the very old. And we can do that with that control experiment. And then we can see that at least before it seems to be that the trend were in fact parallel between the two places. So that's encouraging. The old and-- these are already my old. And this is my very old. And if I had done the diff and diff for them, I would have found a much smaller number. And so, that goes in the right direction.

Now, this is all-- this is quite good. And we can do better than that. And that's where we have-- first of all, we have many groups. So in some places, they built one school per 2,000 kids, and two school per 1,000 kids, and 100 schools per 1,000 kids. Or well, it doesn't go to 100. But so we could think of several treatments.

So in our case, you could have one to three if you have not very many. Or you could have, in our case, we can say this is something discrete. And I'm going to replace my simple diff and diff by basically a regression pre and post on the number of education differences on the number of kids that were built.

So again, if I had only two cohorts and I only relax that, it gives me-- I'm now need to control. I need to put region-fixed effect. So that's control for any differences. And then I'm going to assume here-- I make the assumption that the effect is linear in the number of school. So I'm regressing on an interaction, not anymore on a treated dummy, but on a number of school per 1,000 kids dummy.

And that gives me this regression, again, that I can do by comparing just the young to the old. And so, now this is interpreted as saying that in the places that got one more school-- for each extra school per 1,000 kids, the increase in education between the young and the old cohort is 0.12 years faster. Does that make sense?

So this is how we read that. That's a regression. You could have read that-- you could have run that as a regression of the difference between the young and the old on the number of schools per capita that were built in their region of birth controlling for any differences between the region.

And we can do the same experiment between the old and the super old. And things are now much more precise, although they are too precise for reasons that I will come in a minute. My standard errors are completely off in that paper. I didn't know any better. But I now do know. So this standard error off for reasons I'll discuss in a bit. But at least it seems quite a bit more precise, or a lot more precise.

So that's good. It looks like every 1,000 schools increase the number of year by 0.12 years. And now we can do the-- we can use the fact that in fact, we have many cohorts and also that perhaps we have a prediction. So instead of controlling, first of all, instead of controlling for just the post, I could control for one demi-period. That's not going to change anything.

But more interestingly, I could, instead of having just an interaction for post and the number of schools, I'm going to now control-- I'm going to have an interaction for every year dummy and the number of schools, which you can think-- the way this is written, which starts from the absolutely youngest people or oldest people until the very end, you can think of running a series of diff and diff, where every time you're comparing all the people born in 19-- I think the oldest people are born 1950 or something.

You compare them to people who are born-- you do a diff and diff between the 51 cohort to 50-- the difference between 51 and 50 regressed on the number of schools that were built. And then the difference between 52 and 50 and the number of schools that were built, et cetera, et cetera.

And that builds you your-- so this first one is just treated. And this second one is combined with the treatment of schools. And that gives you this typical event study graph. So this is the one for education where I started here at zero, and for 24, for the oldest people. And so, this is a diff and diff, the continuous diff and diff for the 23, for the 22, et cetera. This is run as one regression. But this is each of these dot compares to the oldest people.

And I do have a prediction that there wasn't any school built that these guys could go to if they are older than 12 in 1974. So I shouldn't see-- all of these guys should be zero. And then as schools most schools get built and people are younger, so they get to enjoy them for a longer period of time, the prediction is that the treatment effect should increase.

So at that time, I think people really liked that graph. Since then they have become very, very, very common, and fashionable, and important. So this graph has a lot of issues. It's not really-- it's not the way that one would construct such a graph today. And I'll show you all of the things we would need to add.

But at this time, I was pretty satisfied that if you eyeball, this looks like a straight line. This looks like an increasing line. And that's what we would expect. So that's encouraging.

Once we have found this, we can say, well, now that I have convinced myself that this is, in fact, zero, I can impose it, which gives me a large control cohort. And now I can construct my diff and diff all with respect to the control cohort. So all of this is going to be normalized as having zero effect compared to them.

And then, this is going to be compared to the entire group. So someone asked, why would you do that? Well, you do that because you get much more precise results if you have a control group that's much larger. And oh.

So that's what I did in that paper where I still-- I estimate where-- that's what's done here. We are here running a regression with only 10 dummy or 11 dummies, which are all the people treated and their various treatment effect as they age.

And then, once we've done it for education, we can also do it for wages, which will lead to what we'll discuss next time for the. So that's all nice. This is a very typical example of a diff and diff two-way fixed effect that have become super popular in applied economics. And you're going to keep seeing them.

This one is a simple one because everyone got treated at the same time. The only complication is that the effect-- the treatment is introduced progressively. So we would not expect no effect and then a program effect. Because we expect-- the treatment itself, the intensity keeps increasing. So the impact should also keep increasing.

There are also many cases where states are adopting policies in a staggered fashion, or regions are adopting a policy in a staggered fashion. So for example, you could have a change in the minimum wage laws that one state approves, and then another one, and then another one, and another one.

And then you have-- you can also compare, even in these settings, even if everyone eventually adopts, you can still estimate the impact of the policy by comparing, for example, people who haven't adopted yet to people who have already adopted. And that group keeps changing over time. So in this setting, so that's called the staggered design.

So this is a one-time, one shot policy stuff. Those examples are staggered design. And there is a whole literature for how to do these things right. The first thing with staggered design is that you want to put them in the right place. You want to relabel them such that zero is the beginning of time-- is the time where a particular state introduces the policy.

And everything is-- you will draw your event study graph not in calendar time but in pre and post the event time, event being the policies being adopted in a particular place. Does that make sense? I mention that because I'm going to show you an event that is graphed before, [INAUDIBLE] understand what it means when the policies are staggered.

So there is an active literature on how to do it right, which means that this paper is not right anymore. Fortunately, the findings are still mostly right. But there is a lot of things you would do differently.

First of all is the standard error. So we discussed at some point very briefly the idea of clustering standard errors. So at a minimum, how would you cluster the standard error in a diff and diff program like this one where you have individual data but the treatment, the program is allocated at the district level.

STUDENT: [INAUDIBLE]

ESTHER DUFLO: At the minimum, you want to cluster at the district [INAUDIBLE] by 10 times year. And that because that's where everyone in a particular district in a particular year has the same number of schools. So that's one, which is not done in that paper. But is easy to do, doesn't change much, but could have if the outcomes between people in a given district were [INAUDIBLE] recorded.

Then you need to-- then there is one more worry, which is that suppose that instead of having annual data. In fact, I think in that data set, I have people exact birth date. So I aggregated it by year of birth.

But in truth people's months of births also affected. So instead of aggregating the data, running the regression at the district time year level, I could have said, let me run it at the district time month. Why not? It should be more precise and going to give you more data.

And suppose that I very seriously cluster my standard error at the district time month. Suddenly, I have many more clusters than in district time year. So I'm going to get results that should be much more precise. Does that sound right?

We could move further. I could say, well, I'm going to aggregate at the district time day of birth. Probably I'm back to very small clusters now. A lot of very small clusters should become very precise. Should I want to do that? What sounds odd? Yeah.

STUDENT: Never mind.

ESTHER DUFLO: Yes.

STUDENT: You eventually run out of freedom degrees, degrees of freedom?

ESTHER DUFLO: No, because I'm not running-- I'm still running exactly the same regression. So in terms of the number of-- or I'm going to have more dummies for day of birth. But that's not really a problem, because I also have a lot of--

[INTERPOSING VOICES]

STUDENT: But not have treated and control people inside the same cluster?

ESTHER DUFLO: Yeah, I could have very small clusters. But that's fine. I'm going to have-- the smaller the cluster, the better, per se. Yeah.

STUDENT: Because you should have the treatment is correlated across time [INAUDIBLE].

ESTHER DUFLO: Exactly. It sounds a bit like I'm manufacturing data. I'm not manufacturing data because a number of observations will be the same. But I'm reducing the cluster size. We agreed that I needed to cluster at the level of the unit of the treatment.

And we said, well, that's about-- that's year time district. And then now I'm making it-- well, year times month time district. So I have more cluster, basically more effective observation. It sounds like I'm creating precision that seems too good to be true.

And the issue is that the treatment is very correlated by your year of birth in particular. Someone born in February or March is treated almost the same. Their level of treatment is almost the same.

So that would in and of itself be fine except, that the outcomes might also be correlated because they are in the state-- they are in the labor market roughly at the same time, et cetera. So yeah.

STUDENT: But what if the-- it's, I guess, a similar scenario. What if we observe multiple observations of the same person?

ESTHER DUFLO: Yeah, likewise. If we observe multiple observations of the same person, we wouldn't want to treat them as individual observations.

STUDENT: So we cluster by individuals?

ESTHER DUFLO: Exactly. So the question-- in that case, we would cluster by individual. We would say, they all come from the same person. We would either make an assumption about what the time series of this era is. Or we would just cluster by individual saying, I'm acknowledging that they're all coming from the same individual they are not independent observation. Therefore, I need to take that into account in my standard error.

Similarly, for people who live in the same district year we say, well, they live together, or they do a lot of things in common. We need to cluster at that level because the treatment is also cluster at that level.

And then, similarly, as you both are getting to, people who are living in-- who are from the same year of birth, similar years of births, similar months of births have experienced similar education condition, labor market condition. Their outcome is likely to be correlated. And their treatment is very correlated.

In fact, with diff and diff, it's really extreme. Because once someone is treated, they stay treated forever. And therefore, there is a mechanical very strong correlation between treatment status of people who are born in related years because it's not-- it doesn't go on and off. Therefore, we need to take that serial correlation over time into account.

The simplest way is to do exactly what you're saying, which is instead of clustering at the district times year level, we are going to cluster at the district level to take into account the correlation in the outcomes of people who live in the same district and are born in different years. The clustering-- so the correlation is not one.

But the clustering-- we let the data figure out roughly what it is. Unless we have very few clusters, in which case you need to do other things. Yeah.

STUDENT: And if you have data, like monthly data of the construction of schools, would it make sense to have clusters that depend on the time as well?

ESTHER DUFLO: Yes. You could do that. You could say, if you had monthly data on school construction, you would have your treatment defined at the month time place. But then you would also want to cluster at the place level, because for sure there is a lot of autocorrelation in the number of schools constructed, because it keeps going up. It doesn't go down. And the outcomes might also be correlated for people in similar cohort over times.

So that's the first thing we have to do. But this we've known-- yeah.

STUDENT: Would it also-- I mean, this would still be, I think, you would still have the autocorrelation that you can handle once you cluster at the district level. But would it be slightly comparable to calendar year to use a school year so you're getting the cohort born into the same school-year?

ESTHER DUFLO: Yeah. You could have a bit more precision that, maybe if you could try playing with that. I do think the calendar year is-- the school year is calendar in this particular instance. But I'm not positive. Yes. So that doesn't-- that's not about-- that's not a standard error problem. It's define your treatment in a more precise way. So that's going to give you more precision that is going to be reflected in your scenario.

So first thing that I didn't do that I now know, you have to cluster at the district level. And when you do that, actually, the estimates are less precise. You still have a significant impact on education. But the impact on wages are a bit dicey.

Then you have this series of people that are much more recent that you're going to see more in recitation and in the problem set. But I'm going to make the list for you. These are my understanding of the papers that needs to be read now if you're interested in ever doing an event study or difference in difference in your life.

One is the Goodman-Bacon point, which makes this point of stacking the data appropriately if the event is staggered. One is-- so this idea that it's a bit loose that, oh, is there a pre-trend that everybody is testing?

This paper takes that idea more seriously in saying, first of all, how would I test for pre-trend. Second of all, you could not reject the pre-trend, but they are there anyway, so we should control for them anyways. So it gives you some tips for controlling for any pre-trends.

This paper is great. I'm going to show you two graphs from this paper. But it's fantastic. I think eventually it probably everything you need to know is in here. I'm willing to bet that that becomes the standard and everybody else is forgotten. It's a review paper on how to do an event study graph as well as how to control for the pre-trend. And I'll show you their event study graph. We can compare it to mine.

But it's a review for the annual review or something like that. So it's very pedagogical. It's well done, and there are [INAUDIBLE] codes to it. So my sense is it's your one-stop shop. I just found out about it recently, otherwise it would be in the syllabus with three stars.

Then another issue that you might get into is when-- if the treatment effect are heterogeneous, then what you get from just a simple difference in difference in the stack design is a weighted effect of all the treatment effects.

And usually, that's fine. But if you're unlucky, some weights could be negative, in which case you're a bit screwed. So there is some work to correct for that-- construct the right standard error. It's an issue probably to check. I would, in my sense, probably less a central issue than doing that properly.

So here is their event study graph after all of the suggestions have been implemented. So maybe you don't even need to read the paper. You can refer to this very neat summary of it. I do think you should read the paper.

But what do we need to do? Well, this is our event study. So it's zero is the point. In our case-- in my case, it would be 12 in 1974 would be my zero. So what did I do wrong?

So first of all, they suggest that it is better to normalize the year before. I normalized here, at the beginning of my period. But of course, that doesn't make much sense. You would want to normalize the year before the study. So I should have used as my omitted category people who were 11 in 1972. So you normalize at minus 1.

Then I have no confidence interval in any of my graphs. Actually, some of them have, but not too much. So you should have the confidence interval that everybody is now doing. What they suggest adding as well are these longer confidence interval, which is a uniform confidence band. So it takes into account the fact that you are not just testing one by one by one, but you are testing all of them. So the standard error are a bit wider.

Then here, they put the average of all this point. So it's basically the average of your outcome in the-- it's normalized as zero in the regression. But it's not zero, of course. There is a-- so it's an average of your outcome in the pre-period. So it gives you a sense of magnitude and how the magnitude of the point estimate compare.

What else do they propose? They propose also to plot what you think is the more restricted model. So in this case, the assumption was that there is an effect, there is a pre-period, and then there is an impact. So that's the restricted model.

So you can see the extent to which there is deviation from that restricted model. In my case, I would have built-- I would have drawn this. Or in fact, I could have drawn something somewhere. At some point in the paper, I had drawn the number of school built.

So you could build-- you could construct for every child how many years of effective school they would enjoy. And that would be your restricted effect if the effect were only proportional to the number of school multiplied by the number of years you have to enjoy. So you could compare it to the shape of the coefficient. That makes it easier for the reader to do that.

Finally, you could introduce-- you could put in the legend of the graph-- so here, what they suggest in the paper but for some reason it's not here, is a P-value of all these guys being negative. So here you would not be able to reject. I think they are nice and negative. So a valid test of all these things being negative. So that's the test-- a formal test of-- sorry, being zero-- formal test of zero pre-trend.

And then, a formal test of the restricting model. In this case, the restricted model is that the treatment effect is constant once it starts. And here, you can see that it looks like it's not. And in fact, they introduce here the test for constant effect post and the rejected. So that's in the-- that you can put that in the legend. So I put both the pre-trend and the constant treatment effect. But for some reason, on this graph they only have that.

So that's your state of the art event study graph. If you're a second-year student, you're writing your second-year paper, you want-- and you are planning on an event study graph, it needs to looks like that.

And that's another suggestion they have is to plot the least wiggly line that fits all the point in the graph. So here are the least wiggly line that fits all the point in the graph is a straight line. So that basically is a way to tell you it doesn't fit. It looks like this is all pre-trend.

And here, this one, the least wiggly line possible is quite wiggly. So it seems very unlikely that this is due to some other thing that is moving in this funny way.

So in our case, in the case of the in-pres program, the least wiggly line that fits all the point would probably have been nice and like that, and therefore quite unlikely to be explainable by something else. But it's a way of making this argument in a less-- in an impressionistic way.

So that's the state of the art of difference in difference. Any question on that? I thought it would be a good occasion to put all of this on slides. Then you'll have a convenient way to refer to them later.

So that's also what I have to say also on substance in terms of what's concerned the people's reaction to the cost of education. We'll go back to this paper. So many people have asked about general equilibrium effect. And I'll get back to that next time.

Actually, I did write a paper about it. So I have a paper that looks at the impact of school construction on people who are not affected by school construction to get at the get at the equilibrium effect.

Let me spend the last 10 minutes we have-- 10, 15 minutes we have thinking about the other aspects of what else enters in the education production function, what people seem to be responding to. So one very nice feature of a very specific feature of this model is people should respond to the perceived return to education. So how one would go about testing that?

So how Jensen first exploited the fact that people make mistakes on what they think the return to education are. And therefore, you can just tell them the truth. And by telling them the truth, you are affecting their perceived return. And you can see whether therefore they react by going to school more.

So he did this project in the Dominican Republic where the completion rate for high school is low. And he first went to the students and asked them their view of return to education by comparing-- by asking them to say how much someone with-- for those who have complete primary school versus those who have completed high school, what's the difference?

By the way, so let me finish that, and then I ask you. At baseline, people seem to be-- people here seem to be underestimating the return to education. Which, by the way, is really not a generalized finding. In my experience, people widely overestimate the returns to education, at least to secondary education.

But in this instance, they were underestimating them. They saw that you would get only a 10% increase from moving from the eighth to the 12th when, in fact, it's much bigger than that.

And then what they did is that they told them the truth in some schools to the Mincerian return are more like 10% per year. So 40% for going from eighth grade to high school. And then, after some time they compared who stays in school. First of all, they show that they update their beliefs. And I'll show you that in a minute. And then they investigate the impact on education.

So before I go further, what's the issue with doing that? Or in fact, even with asking them in this way, the difference they thought between in the income of someone who has completed versus not completed? What is potential concern in doing this exercise? Yeah.

STUDENT: Well, I just wanted [INAUDIBLE] don't totally understand what that means, and I just also [INAUDIBLE].

ESTHER DUFLO: Yeah. They might not understand what that means in some sense. Yeah.

STUDENT: It's depending on how the question was asked, it's not clear if it's like the return for an individual who chooses to [INAUDIBLE] primary school [INAUDIBLE] but also taking into account the ability.

ESTHER DUFLO: Yes, exactly. Exactly. And in what is given to them is full of the ability bias potentially, or the selection bias of any kind. You're telling them the difference between learning or someone who has completed and not completed education. That's not the returns to education. That's not the causal effect of education.

So potentially they are lying to people, which is a bit problematic. Or not lying, but giving them an information that's not accurate. It's not accurate for the average. It might-- and then the other part of your answer is that it's not even the-- it's about an individual. It's not necessarily the relevant information for any particular individual, which, of course, you don't know.

So that's a little bit of a dicey project. But again, maybe we didn't know until-- this might be 2020 hindsight. But that suddenly something that people criticized. They do find an impact on education. They find that people generally update their belief positively. And they find an average impact on education.

Let me show you the table. So perceived return have increased. On average these treatments regress for a bunch of stuff, doesn't matter. This is RCT, you shouldn't need any of that.

And then, so this is the average effect on passive return. And then the average effect on various things, whether they return, whether they finish the years of schooling. And then they actually find a larger effect for the rich household than the poor household, possibly because in addition, people they are also maybe credit-constrained, people can't do it immediately. So return is not the only thing that enter the decision.

So if you had to run this regression-- having done this experiment, if you had to run this regression, would you run it this way, or would there be another way to run it that would make more sense. Yes.

STUDENT: Wouldn't we need to take into account that the treatment effect is different depending on how [INAUDIBLE]?

ESTHER DUFLO: Exactly. So at the minimum, you should think that the treatment doesn't have a monotonous effect on these other things. Because some people, even though on average people were underestimating, it is likely that some people were overestimating and some people were underestimating.

So at the minimum, when you are doing this kind of debiasing treatment, and they have become popular and we are going to see other examples of that, you should look at-- you should separate the sample by the people whose prior you're moving up to the people who prior you're moving down, because you would expect the effect to go in opposite direction.

Well, that's just a note of how one could have done that. So given-- so this paper was criticized for what I'm telling you, which is that basically you're telling people some junk. That's a little bit troubling. Maybe we should do that. And then there are other paper that are trying to use paper that are trying to use the real data-- data from the real world, to look at how people respond to the return to education.

An older paper by Foster and Rosenzweig in India, quite nice, looks at the impact of high-yielding variety, and argue that high-yielding variety grain has a large impact on returns to education. Because basically, you need to be able to read the package, and understand what it is. And then they have some suggestive evidence that the return to high-yielding variety is very low for people who are not educated and very high for people who are educated.

So the introduction, which is a bit sequential, as well of high-yielding variety across villages has different impact as-- increases the return to education. So that's the first thing that they say. And then, as a result, when you see high-yielding varieties being introduced, that should increase people educating their kids because the return to education have now gone up. And they find that.

Now, one issue with that exercise is that at the same time that it increased the returns to education, it should increase people's income. And we've seen that in a richer model of the world, income effect would also increase education in and of itself. That's not an issue for them because they don't believe in income effect. So the only thing that could possibly have happened is the response to returns to education. But one could be worried about that.

Two more recent paper that have the same-- those two papers have a little bit the same mechanism at hand-- is a David Atkin paper and Sara Hernandez paper. Sara was a graduate student here and now teaches at Northwestern. They look at expert jobs and export jobs coming from the garment sector, mostly in Mexico, and the flower sector in Colombia.

And they show that when these jobs are introduced it leads to dropouts of girls despite the fact that people aren't richer. Which they explain by the fact that the return to education are now lower because you don't need to be educated to get any of these jobs. So you can get a good job working in a garment factory, or you can get a good job working in a flower sector without education. So people dropped out.

In the example of David, they got screwed for it because then the sector disappeared and they found themselves without jobs and without an education. That's the full story of that paper. But so the advantage-- "advantage"-- the advantage of that example is that the return to education go down at the same time as the income go up. So we don't have the income effect and the substitution effect going in the same direction. So if we find education dropping, it seems likely that it's because of that.

But then Jensen went back one step further and say, well, you didn't like my change in the belief. So I'm going to change the return. That seems ambitious to say, are you able to change the return to education.

And he had a great, great idea, which is to participate in a recruiting campaign for call centers, these type of things when you call for your computer, and typically it's a nice young person from the suburb of Delhi, where suburb is intended in a very broad sense who is answering the phone.

So early on they were really recruiting mostly in Goregaon, very close to Delhi. And then they started expanding. And he worked with them at the beginning of expansion to lead them to villages where they had never been in Haryana, or in UP, so basically a broader sector are in UP.

He connects the BPO, the back office processing operation with the villages. Someone goes to the village, advertises the possibility, and then sees what happens. So some girls who were the right age and had the right education, which is secondary education with some English went. And he's interested in what happens to the younger girls who could not go yet. They were little. Are they more likely to stay in school?

So what he finds is that women are indeed-- younger women are indeed more likely to be working in a BPO. So that's good. There was no such effect for men. And that both younger and-- both these women but also the younger women, the girls were more likely to be in school. And even younger are better fed.

So basically, the parents-- the story is, parents realized that-- there is more in terms to girls. In general, they become more valuable in general, and in particular are more valuable if they are educated. So the return to education for women have gone up. And they start educating them more.

So interestingly, there is no effect for men. It's not discussed much in this paper. But I have another paper where they separate the boys between people who at baseline were described to be the one who are going to take care of the family farm.

So usually the oldest boy, versus the younger boy. And when they separate that, they find that actually when the BPO equipment started for the boys who are supposed to work in the family farm the education went down. And for other boys, it's more like girls.

And so the idea there is that you're very scared to have return to education go up for the boys because you don't want them to leave because someone needs to stay to take care of you and to take care of the farm. And therefore, you're going to suppress education so that they don't-- here the return to education for you are negative from the point of view of the parents.

So that leads a little bit to the conversation we had with [INAUDIBLE] and others about the contract between the family. That if you cannot contract to get the money back from your boy, then you are not sending them.

So there are two other things that I wanted to do that I didn't have time to do. One of them I'll do later when we talk about the family. That's fine. We can take that. That's the Rebecca Dizon-Ross paper about the fact that parents have no idea what their kids are up to.

But this one I want to briefly mention. It's a paper by Seema Jayachandran and Adriana Lleras-Muney who is looking at another way of affecting the return to education or the way returns are affecting is simply the fact that how long are you going to live?

So if you expect to live for a very short period of time, then your education is not very worth it. Or if you expect your girl to live for a very short time because she's going to die in childbirth anyway, then she's not going to experience the return for a very short time. So that's a very economist way to think. But it is quite interesting that it works here.

So what they are looking for, to know whether people-- so in principle, the implication of this model is that if people expect to live longer they should-- or expect their daughters to live longer, they should educate them more because they are going to enjoy the return of this education over a longer time period.

So you need an instrument that affect the number of years someone can expect to live without affecting anything else, like the skill premium or the low skill wage, which seems like a very hard problem.

And what they find is in Sri Lanka a program that introduces ambulance in all the villages. And when they did that, infant mortality collapsed. They did that, infant mortality collapsed. And in particular, the drop in infant mortality was much larger in places where infant mortality was very high.

So basically, they're putting the ambulance, flattened out the infant mortality curve, leading to huge decline in infant mortality in places that were really far from the hospital and people couldn't get to the hospital. So the ambulance eliminate that.

So now they use that in saying, well, that affects the life expectancy of the women but not the men. So now I'm going to use basically the pre-ambulance maternal mortality rate as an instrument for the decline in maternal mortality rate over the period to see whether people seem to respond by educating their girls more relative to the boy.

So now what we have here is a triple difference specification. I just write down-- put the specification down-- where we can have a fixed effect-- we can have a two-way fixed effect whichever way, district, year, district time year, and gender time district, and gender time year.

So all of the double interaction can be there. Because what we are going to use is maternal mortality in a particular year, like the time when the parents could observe it, times the female [INAUDIBLE]. And they find that more maternal mortality, less education. So using just that, they find that more maternal mortality leads to less education.

So that's kind of a nice paper because I think many people were kind of intuiting that something like that might be going on. But since everything is usually correlated with everything it's very hard to test. And this ambulance program that had such a massive effect on maternal mortality reduction gave them a chance to test the idea.

All right. So sorry-- so we're done for today. Sorry again for the slow start. And on Monday, we'll continue talking about the Indonesia paper. But we'll talk about this the estimating the return to education. We also talk about the Ghana paper, which I'm asking you to read as you go along, which is another take at estimating returns to education or impact of education.