[SQUEAKING] [RUSTLING] [CLICKING]

BEN OLKEN: OK, so just to remind you guys where we were last time-- So we're talking about-- in the middle of the stuff on how do we figure out who's poor for redistribution programs.

And what I was telling you about was this first experiment that we did in Indonesia looking at community-based approaches to identifying who is poor. So this idea is they're going to rank people or whatever.

So what I want to talk about is a little bit on this. And then I want to talk about the self-targeting paper. So one thing you can do is we can just-- so how do we evaluate this? What are the outcomes? So just to remind you guys where we were like Wednesday when we talked about this-- Wednesday, yeah-- was we had a baseline survey to measure the truth.

We then randomize villages into different ways of identifying poor households. And then we want to evaluate kind of which did better. So one simple thing you can do is you can say we can define mistargeting just to be a dummy variable if we got it wrong.

So if you're poor and we gave you the program, that's an error. If we're rich and you-- sorry, poor and you did not get the program, that's an error. Rich and you got that program is an error. That's kind of discrete.

But we may want to do something different. And in particular, one of the things we were interested in this paper was whether-- trying to evaluate these things on different welfare metrics. So in particular, we wanted to identify what not only did they do better on consumption, but did they do better on other alternative measures of well-being.

So we started off by saying consumption, just again, that the per capita consumption. We then wanted to say, well, are the communities, when they're ranking people, doing a better job kind of in that village of matching how people kind of privately rank each other in the privacy of their own home.

Like is there some generally accepted notion in this village of who's poor and who's not? And maybe that's what they're matching. Maybe they're matching how the elites in this village are matching people. And the other thing that they're doing is they're matching-- maybe they're matching people's self-assessment, their own kind of internal conceptions of poverty.

And so we measured all four of those things at base-- all four of these things actually at baseline. And so we wanted to compare which of these things does better at matching these four different welfare metrics.

And the thing I wanted to point out is one problem is these are on very different scales. So we wanted to find kind of a way of evaluating-- so we need to put all these things onto kind of a comparable metric to make sure we were kind of comparing comparable things.

And so what we did is we came up with a ranked correlation. So in particular, we took each of those four metrics, and those pretty much are ordering of people in the village, maybe with ties. So like Ben is the richest based on consumption, or Ben is the richest based on-- or the happiest based on self assessed kind of utility or whatever. So all those things generate rankings.

And then we can compute-- and each targeting method also generates a complete ranking. That was the ranking from the-- in this one, it's the ranking on the wall or on the clothes line. In the proxy means test, it's your predicted income score, it's also a ranking.

And then for each of the 600 villages in our experiment, we can compute we can compute a ranked correlation, which is the correlation of ranks between the targeting outcome and each of those four different kind of welfare metrics.

And so I just thought that was-- it's kind of a little bit unusual because we had to think about how do you take these things and put them on a common scale? So we're going to end up doing is creating essentially 600 times 4 different rank correlations because there's 4 different welfare metrics and 600 villages. So there'll be 2,400 different rank correlations.

And then we're going to regress for each different welfare metric this rank correlation itself on the targeting method used. So we're trying to predict which-- did this method generate a higher rank correlation? So it's a little unusual, which I just wanted to describe that in detail.

And so what do you find? Well, on the one hand, you do find that the community method made more errors. So this is regressing kind of like this dummy of mistargeting by basing assumption on the community and the hybrid-- I'm not going to talk much about the hybrid-- and predominantly kind of by people near the middle of the distribution.

But on the other hand, if you look at the 4 different rank-- if you switch to this table where I'm regressing the 4 rank correlations on the different treatments, what you can find is that actually that indeed consistent with what I just showed you, the rank correlation between what is produced by targeting and baseline welfare is lower if we use consumption as baseline welfare in the community treatment. But it's higher on basically everything else.

So it's not just that these communities are like systematically making mistakes or being totally captured by elites or whatever. No, what's actually is happening is, I think, to my mind, the best assessment, best thing that's happening is there are some local notion of what poverty is in this village, which is highly correlated with but not perfectly correlated with income.

And when you tell communities to tell you who's poor, that's what you get. And in particular, you strongly get that it's very correlated with-- what we have in communities is very correlated with kind of what you would get if you ask people about who is poor and who is rich in the privacy of your own home.

And it's also correlated with-- same thing if you ask the subvillage head, the neighborhood head, to do the same thing. It's very correlated with that. It's also pretty strongly correlated with just your own self-assessed poverty. This is just a ladder scale. On a scale of 1 to 6, where 6 is the richest people in society in 1 is the poorest, where you put yourself, that thing generates ranks.

And what you get out of the community treatment is like 10 percentage points more correlated with that thing in the community. So it seems like you're picking up a kind of a different welfare metric. So I think that's kind interesting that maybe these different local welfare metrics, and that's what you're picking up here.

And the other thing I'll just mention is that the communities all like this community-based approach much better. So another question is this-- and we measured this in two ways-- one is just some general satisfaction. Like in an online survey, are you satisfied with the way targeting was done in this village? Way more satisfied in the community treatment?

The other thing is we show them the list, and say, here's the list. Did we miss anybody? Anybody should be added? And they're making many few fewer edits to the list. So it seems like it's really picking up kind of that local welfare metric, and they're much happier with it.

So in general, kind of the point of that paper I think is that the community has a somewhat-- we started off by actually designing this whole project as like-- we thought it was a trade off between a leak capture and better information. That was the decision.

But actually, the third thing kind of emerged that we weren't really expecting, which it was really about this different well metric. And I think that's actually what I think is the most important thing. So we find kind of basically no elite capture. We went to all this effort to measure elites by like who's connected to the village head and who's family members of who. And so none of that seems to matter.

We do actually show that there is local information about consumption. They do have additional information that the government doesn't have. But they are tending to do something a little bit different.

OK, so any questions on this before I move on to the next paper? OK, so I think the conclusion is, if you want to use poverty headcount, minimize poverty headcount, maybe you'd want to use the database approach. But if your goal is to maximize utility and you believe that the social utility is a function of the individual utilities and you think that's what this is capturing, then this approach might work better.

So the paper that you guys actually read for last time that I want to spend a fair amount of time-- that's also going to be talked about in the problem site is this paper on self targeting.

So what is the idea? So the idea goes back to this paper by Nichols and Zeckhauser, which is that-- what they say, they say or deals can be used to target the poor. And what do they mean by that?

Well, here's an example. Suppose you need to wait in a really long line to get unemployment benefits. And maybe suppose the unemployment office is only open from 9:00 to 3:00 in the middle of the workday. What is that going to do? That's going to be not very costly-- I mean it's costly for everyone because no one wants to stand in line at the unemployment benefits.

But it's differentially costly for unemployed versus employed people because unemployed people don't really have that much to do between 9:00 and 3:00 on a workday. I mean they have other things they'd rather do, but don't have to go to a job because they're unemployed.

Whereas employed people would have to miss a day of work to do that. So it's going to be differential-- so the idea is this idea of having you have to show up every once a week to screen your unemployment benefits to recertify your unemployment is going to differentially screen between people who are really unemployed and people who are not.

And the-- one second-- and the downside of that approach is that it's imposing a cost on the unemployed people, which is, in principle, you think that cost is just totally kind of deadweight loss. Like we don't want to have to have the unemployed people have to do this. The only benefit is the screening cost.

So the way to think about this is, on the other hand, if we can screen out enough non-needy people, we could take all that savings and give the truly unemployed kind of more. So it could be actually beneficial on net if you can think about it through the fact that you can kind of increase the transfers you're giving to the really needy. But it does impose costs. Yeah, [INAUDIBLE].

AUDIENCE: Well, I was saying this isn't quite targeting the poor, right? This is only targeting people with low [INAUDIBLE] the cost of time, which if--

BEN OLKEN: Correct.

AUDIENCE: So if you were-- your income level was the same as somebody else's income level but you didn't have as much wealth as they did, it could be--

BEN OLKEN: Correct, yes, yes, yes, yes. So right. Hold that thought. Well, this is-- yes, so this one is to target unemployment benefits. So this is explicitly trying to target people who are not. This one, this example is trying to target people who are unemployed. So I think that example works pretty well.

But I think you are absolutely right that the general question whenever you have these kinds of mechanisms is, is it targeting the thing you want or not? And I'm going to talk about a bunch of reasons why that may or may not work out kind of in practice. Other questions?

OK, so here's a basic version of the model. So imagine that-- now, in this example, this is now from our paper. But this is the basic case where you have-- we have to walk some-- go some-- travel some distance to apply for benefits. And the thing we're trying to screen on is your opportunity cost of time.

So imagine that the cost is going to be kind of-- so basically people who are richer imagine they have a higher opportunity cost of time because their wage is higher. So imagine this is a single dimension, where it's your wage.

Right so in that case, if you had a close-- you don't have to walk very far, the cost will be downward sloping in consumption. But they will be more downward sloping in consumption if I increase the distance because the amount of time you have to spend walking there or whatever because that's going to be proportional to your wage.

And so that's imposing a cost on everyone. So here's the whole thing now but it should sit differentially for the poor or the rich. And therefore who's going to show up is going to shift to the left. In this example, everyone-- people will show up if they're above this line. So increasing the cost in this way is going to get you better targeting.

Now, in fact, it's more complicated than that for a bunch of reasons. I think you mentioned kind of one of them, which is that maybe the thing you're targeting is-- what you want to target, say, is maybe wealth. And what you're actually targeting is kind of wage. And those may not be perfectly correlated. So that's one example. Anyone remember some other examples we talk about of other reasons these things may or not be the same? Yeah.

AUDIENCE: Different costs and getting there, like having a car.

BEN OLKEN: Right, so it may be that this ordeal was setup for imagining everyone's walking, but actually we have different ways of dealing with the ordeal. So maybe we can drive. Maybe if you're really rich, you can hire someone to wait in line for you. You can pay an agent to do this thing for you. And so like actually maybe the rich have ways of undoing this ordeal. Other things?

So the so the one we were really worried about was this one, actually, which is what happens if you have really concave utility in a day's wage, for example. And this actually-- when we were pilot testing this, it was not ex ante obvious this was going to work. It was certainly we thought it might work, so that's why we did it. But there was reasons to think it might not.

And the one thing we heard from some qualitative reports is that some people said, look, I can't do this, I can't show up for a day, even though like there's like $1,000 of benefits on the line here, in NPV terms from my showing up in applying for this program. I'm going to miss a day's work if I do this.

And that's going to be-- basically, what I earn that day is what I eat that day. So like I'm going to basically have to go without food for a day. So it's like super costly for me to apply for this. I mean I could do it, but it's super costly.

Whereas, say, for richer people, they didn't have that. They could maybe smooth that, for example. So it may be that even though this previous one was in terms of income, net income. But if I start putting concave utility, if the thing is super duper concave, then actually the fact that this is shifted down is actually first order, could be really important, because actually these people could actually be losing a lot of utility over here.

And so if I put concave utility, maybe the thing looks like this, for example, because here we have-- this side, we have the standard effect we have had before. Over here we have the effect that utility is super duper concave over here. And these people are just experiencing-- even though maybe it's a smaller income utility loss, it's a larger utility income loss.

And so in this example, actually, maybe what happens is you have the people that apply shift in or whatever. So then it's not obvious. And then also, as you guys talked about, there was example of maybe you actually switch, kind of switch technologies.

So maybe for example, the rich have different technologies. And as the cost goes up, they switch the way they get there.

So the point of that was to say that actually-- the first point I think that we want to say is that, look, this Nichols Zeckhauser idea is a nice idea in theory, but it's not ex-ante totally obvious. And those are things that we wanted to test empirically in the data.

The second thing was-- and again, this is also part of our learning process in writing, in doing this project. When we set out to do this paper, what we were worried about was this stuff. Like this is what we had in mind as the reasons this might not work or the things that might be really important.

What we realized ex post after writing the paper is there was another factor, which was really important, which was the paper ends up being about, which is that in the context we're going to study, what you guys read about was it's not just that you do this thing and you get the benefit for sure.

What happens is you do this thing. And then you get some stochastic like-- you get the benefit with some probability. And the probability is the proxy means test that we talked about last time.

And so in the example here, after showing up and applying for benefits, they still proxy means test you. And you may or may not get it. And so in addition to a cost and benefits of applying, you also have to forecast, like am I likely to get this thing or not? And you're going to take that into account.

And that's going to add another complication wrinkle to the model, which is that sophisticated households are going to-- and there's two different ways have to model this. Number one is it may be that I fully understand what the government's going to do, in which case, actually, there's not a lot of uncertainty.

There may be still a little bit of uncertainty, but not a lot. And then the other case is, well, maybe I don't understand what the government's going to do.

And in that case, it's kind of interesting to say, look, if I'm-- and it's actually pretty reasonable, which is you say, look, I know my income. I have some rough sense of what other people with my income level is likely to happen to them. But I don't really understand the details of what the government is doing with that proxy means test.

And this is pretty plausible. There's like at least one of you in your comments was like, well, maybe they should know. But I think it's at least pretty plausible that they don't necessarily know exactly what's happening.

And that means that the-- so if this one's going on. If this sophisticated one is going on, that's still helpful for the government because they have to bother to screen people who aren't going to get it anyway.

But this one is now really valuable because I'm making my decision as to whether or not to apply based on my actual y, my actual income. But the government-- and that includes the part my income the government observes and the part of the income the government doesn't observe.

So I'm making the decision including my unobservable information, and in so doing, I'm kind of revealing that information of it to the government. And so the government is getting more information, in some sense, through that decision. And that's kind of what we think is helpful here.

And so just to show you what this looks like-- so this is if we plot in the data-- so in the paper, we know your baseline consumption score. And in our baseline survey, we also collected all the variables that are used by the government for the proxy means test screening.

And so we can replicate the proxy means test screening. We can compute your proxy means test score as you would have observed it. Suppose you knew the government's formula and knew the truth about you.

We can say, what's your proxy means test score? And then we can say for those who actually did show up, did they actually get the benefit or not? This should be a step function, right? You would think. Actually, the thresholds were a little different in each province. That's why there are different lines. So it should be like six step functions there were six-- sorry, each district, six districts. There should be six step functions in this graph.

Why are there not six step functions? Because when the government goes in measures your four different variables, or maybe a little bit of measurement error compared to when we measure the four different variables. So even this thing is not perfect. But it's pretty steep.

One second-- on the other hand, if we plot the same thing against your total income or your total consumption, it's still downward sloping, but it's a lot flatter. And that just captures the fact that I don't fully know, I still-- if I know-- so this is just saying, what do I infer about my probability of getting benefits if I'm, say, down here versus kind of over here or whatever.

And it's still downward sloping, but it's not quite as flat because the government-- it's based on this unobservable component. And observable component, I don't fully forecast that. I'm sorry. Someone had their hand up. Yeah, Kadesh.

AUDIENCE: Should we be designing an experiment like this, how much are you thinking about theory paper that you referenced at the beginning?

BEN OLKEN: The Nichols and Zeckhauser one?

AUDIENCE: Yeah, how much are you just thinking about the local context?

BEN OLKEN: We were totally thinking of Nichols and Zeckhauser paper. So and a local context-- I mean, both. I think that-- we were thinking about the Nichols and Zeckhauser paper. And that's why we intentionally not only--

So stepping way back-- so what were we doing? So stepping way, way back-- we did the first paper. And we presented the results to the government. And they said, that's pretty interesting. It looks like this improving targeting thing has some potential.

But you did that one before in kind of a low stakes environment. We'd like you to try some stuff in a real environment now and see what actually works. So that whole-- so first we did little pilots. Then we did it in kind of a low stakes environment. Then they were sufficiently intrigued. They said let's do this in a high stakes environment.

And particularly, they said we're expanding the conditional cash transfer program nationwide, or not nationwide, to a bunch of new areas. They said there's 600 villages which haven't gotten the program before. And we'd like you to test in those 600 villages three different approaches to doing the targeting to see which one is going to work best to inform our future decision making.

So that's why we were able to do this for a high stakes-- like for real programs because the government wanted to understand what's going to happen.

So actually, there's a third piece of this, which I don't think we analyzed here, but we wrote up in a report, which we actually replicate the version of the community treatment in 200 other villages and show that it looks pretty similar, not exactly the same, but pretty similar.

That's not on this paper because we didn't-- that wasn't how we wrote this paper. But like we did-- actually, there were 600 villages. 200 of them go to replicating the community thing.

The other thing we said to the government was, look, there's this idea that maybe on-demand application-- that's kind of what they call this in the policy sphere-- can help can help you both reduce inclusion error and exclusion error.

Well, I can it reduce inclusion error, falsely including the rich. It's exactly for the self-screening and self-selection argument. So that we had in mind. Why could it potentially reduce exclusion error? That's because, for a totally-- that's a local context reason, which is that basically-- but I think it's actually generically actually quite true, which is you try to go door to door and find everybody, but you miss some people.

And in particular, poor people, because they live at the margin-- often live at the margins of society and the fringes of the village, or maybe they're not on everyone's radar screen, you may miss them.

And so opening up an on-demand application process allows people to make themselves known. So that piece was kind of a real world piece. So we had both a theory in mind and the kind of real world piece in mind.

But we also said, look, there are reasons to test this before you scale it up nationally because, for the reasons I was talking about, maybe the poor aren't going to be able to do this. Maybe they're going be too busy. So that's why we want to test it and see if it's going to work.

In terms of the theory, I think that we had this close versus far treatment, where we explicitly varied the extent, the distance you had to travel. And that was explicitly based on this theoretical idea that longer-- that making the ordeal marginally more difficult should improve the screening. And that's what we wanted to test.

So that close versus far-- so on-demand application was kind of a thing in general that they were interested in. But this close versus far treatment was explicitly designed from the theory.

Now, we did not have the full model written down here. And that came after. So I mean part of it came in parallel. But-- and it was-- let me actually defer a little bit the model part to when I get to the model part. I want to talk about that in a sec. But does that give a little bit of background?

And but I do think, to the extent, you can have a theory in mind as you're designing your experiments, all the better because there are things you will do that are going to be more precise. You do stuff that's more precise.

And in particular, actually one of the empirical challenges-- it's actually came up related to one of your comments, I forget which one. But one of the things that actually we found, where we had a theoretical idea in mind, which is, look, we want to actually have these tests whether these ordeals are improving screening.

It turns out that actually imposing ordeals on people for no good reason other than screening is something that I think just is contrary to human nature. And so we had this problem-- we had this situation which happened in the real world of like all kinds of spontaneous organizing in the field to make people's ordeal easier.

So for example, people would be like, oh, you're in line. You're number 74 in line. Here's a little chip that's at 74. You can just come back later when your turn is in line. Or oh, there's so many people in line. The line is so long, why don't we like divide this village into four sub villages and have neighborhood one come in this day, neighborhood two come in this day, number three come this day, and neighborhood four come this day, to make it like less costly for people.

So I think that's actually-- it's good in the sense of it's a great impulse that people are trying to make everyone's lives easier. It does make it harder to test some of these theoretical ideas.

AUDIENCE: So you have that in mind.

BEN OLKEN: Exactly. So the reason is-- the reason I mentioned that is because this was the thing we were trying to understand, it was important to monitor these issues and keep track of them.

And as I said, the point I want to underscore here is that the goal here is that even though you're imposing costs on people, the goal is if you can differentially screen out enough of the non-poor, you can gain sufficiently large amounts in transfers that you can actually deliver much more benefits ex-post to the poor.

So even though you're imposing these costs at the screening side, it can be wildly beneficial for the poor potentially because the costs relative to the program are-- the costs are, say, I don't know, half a day's wage just to stand there. And you can get $1,000 in the program. So if we can-- over $150 a year for six years.

So if you can screen out a bunch of rich people and save all that money, there's a huge-- you could easily imagine this could be way-- and use that money to increase the benefit size. You could make this a substantially better program for the poor, even with these costs.

So it's important to keep in mind. That's the rationale here, is imposing these costs to improve the screening, which can actually make programs kind of better for everybody. Other questions?

OK, so how does the model work? So let me say-- yeah let me say one other thing actually before I get into the details of the model.

This is an example where I think we learned a lot from the model. And in particular, it's kind of could be a fun exercise. I don't know if I still have it. If the very first draft of we wrote a wrote of this paper was not like this, did not highlight this whole like forecasting my lambda thing.

We thought this paper was much more about the Nichols and Zeckhauser basic kind of screening and opportunity costs. And we looked at the fact, they said, oh, look, poor versus-- the close versus the far, seems like it's doing a little bit of screening, seems like everything's going on.

And we actually thought this paper was about that. And it wasn't until we actually wrote down and tried to calibrate a model of this that we realized that those things, those forces were just quantitatively too small to be explaining the kind of magnitudes of things that we were getting. And this other force seemed to be really kind of driving things.

So if you compare kind of our very first draft of slides on this paper, even once we saw the results itself, targeting kind of improved targeting by a lot, we had a simple theory. We explained why it could be. And it was based on this Nichols and Zeckhauser idea. We had results saying, look, self targeting works. And it's doing this thing.

But until we actually wrote down and estimated the model, we didn't realize that those effects that we were talking about, where they were qualitatively in the right direction, weren't quantitatively large enough. So it was an example where I think we actually learned a lot and changed kind of our views and what the paper really meant by going through this exercise and trying to quantify some of the stuff. So I just wanted to mention that as well.

So what's the model? So in each period, households get a utility from consumption x. Preferences are additively separable in their present and future utility-- discount factor delta. Their flow income is y in each period. There's no savings. y0 is the share observed by the government.

So we're going to calculate-- y0, we can calculate that as before because we know the proxy means test. So y is your consumption. y0 is a proxy means test predicted consumption. And yu, the unobservable part, is y minus y0.

So households decide whether to sign up or not by balancing costs versus expected discount of benefits. So the cost of signing up is c of l y, where l is the distance to the place you have to sign up and y is your income. But we're going to model that as proportional to your wage and maybe how you get there.

For sophisticated households, you're going to say I get benefit B with probability mu of y0, where this is mu. So we're going to give them correct beliefs. And for unsophisticated households, when they up, they get it with probability lambda of y, which is this thing over here.

So the mu people-- the sophisticated people know the steep curve. And the mu ones only know-- the [INAUDIBLE] the other one. And there are some utility shock measure, which is f of epsilon.

So then what's my expected benefit from showing up? Well, for the sophisticated ones, it's just the cost of showing up plus the expected benefits of getting the program using kind of the mu one, the true one. And for the naive ones, it's the same thing, but using the lambda thing there.

And we'll imagine that share alpha of households are sophisticated. 1 minus alpha are unsophisticated. And the final thing, which is just a little technical detail but in the paper is this lambda-- this mu thing is just a raw function of the underlying government's technology.

Lambda is actually an equilibrium phenomenon because it depends on who's actually showing up in person. So there's a condition-- there's some kind of I don't integral-- I think it's equation 8-- there's some integral in the paper which basically states that thousands of rational beliefs in equilibrium just to close the model.

So what was-- OK, so that's the model. So what was the experiment? So we had 400 villages who did this. This program was just-- this is a constraint that we're working with. This is a program targeted to really the very, very poor. So this is the government is aiming for the bottom 10% of the population in terms of--

And so in the villages were randomized into proxy means test are self targeting. And the self targeting is you had to go to the self target-- the way self targeting works is you went to the central meeting place and you said I want to apply. You to wait in some line. And then the government numerators kind of took all your information, put it into a computer.

And then if-- the way it worked was if your answers say you're out, then you're out. If your answers say you're in and the government has data before which also says you're-- from its previous targeted survey that also says you're in and not even close to the threshold, then you're in. If you're at all close or if the government doesn't data on you, they actually do a field visit to verify that you're just not making stuff up.

And I mentioned we varied the distance to the application site. There was also a treatment where we varied the opportunity cost of applying by actually requiring that both spouses show up at the same time as a way of varying the opportunity costs.

That one, if I remember correctly, we had to put in exceptions in case, because reasonably, like what would happen if your spouse was kind of outside-- this is actually what we learned in the pilot process.

People say, well, what happens if my spouse is out of the city? Like I shouldn't be denied benefits of the program. And of course that's fair. So there had to be an exception process by saying that the village head or whatever, the neighborhood head could sign a letter saying your spouse is out working somewhere else and can't come.

And I think that one actually didn't do very much in practice because basically anyone whose spouse couldn't come could just get this letter. So I don't think it actually made that much of a difference in terms of the actual opportunity cost. So that's why we don't analyze that one as much in the paper.

So we're going to investigate who signs up, compare them to the experimentally the PMT, and estimate the model structurally to tease out these mechanisms. So this is the probability of showing up as a function of your per capita consumption.

So as an aside, in the paper, we talk about Fan regressions or non-parametric regressions. Have anyone seen these before? Anyone have any ideas? The first point is to clear up a confusion from when I was a student. These are not fans like this way, like a fan like you wave. This is like Mr. Fan. So just clear that up, clear up years of confusion on my part.

Second of all, what is this? So these are basically local regressions. So what you basically do is suppose you have some x variable over here and you have some like dot cloud of outcomes.

And if you put a line through this thing it's going to look like that. And that's not really what you want to capture. So you want to summarize this thing in some non-parametric way. So what these regressions do is they say for every point, let's divide our x-axis into say 100 equal bins, let's just say, whatever, some number of bins.

And for each bin, we're going to start over here. And we're going to run a regression where each point is weighted based on the distance from this bin. And there are different weights. The weight you use are called kernels. So there are different kernels. So a simple one is a triangular kernel, which for example looks like that.

So what we'll do over here is we'll go over here and we'll run a local regression where this point gets weighted a lot, and this gets weighed a little less, and this gets weighed a little less, and so on. And we'll get a slope it looks like this.

And then we're going to do that for each of these 100 points, like that. And then we'll connect the dots. OK, so that's what that is. So in theta, l poly does this. I'm sure there's an r version that does that as well. But I'm dating myself because I know this data command, not the r command.

And I just I wanted to mention this because I think these are really nice. And then you can bootstrap, by the way to get confidence intervals. So have we talked about bootstraps? Yes?

All right so that's-- I find these very nice because they summarize the data in ways that are a little hard to see with dot clouds, especially because if this is a binary variable, if you try to just plot this thing as a scatter, it doesn't look like this. Right if this is a binary-- if you plot this to a scatter, what it looks is like a bunch of x's and some y's and it's not very easily visualizable as a scatter since it's a binary variable.

So I find these kinds of non-parametric, locally weighted regressions really helpful as a way of summarizing the data. End of aside. Questions?

And then you can-- by the way, the other thing about this is can play with the bandwidth. So the wider you make this thing, the more you're going to be imposing smoothing on it. If you make it super narrow, then you just get the dot cloud back. Eventually, it ends up looking like a binned scatter. So you have to play with the-- there may be-- I don't know if there's an optimal bandwidth for these.

I don't think there is that I know of. But there may be. But you can certainly play with the bandwidth to get a reasonable degree of smoothing when you make these graphs. And the standard errors tend to blow up with the tails because you don't have that much data over here to estimate them.

So what does that say? So this is like downward sloping income, great. But the more interesting point is that we can then decompose this based on your observables and unobservable consumption. And you get it downsloping in observable consumption. And then this says, look, maybe these households kind of forecast they're not going to get it and they don't apply.

The kind of I think really exciting part from a targeting perspective is that it's actually downward sloping in the unobservable component of component of consumption, which is what this is. So this is y minus the observable-- y minus y hat on the x-axis. And whether you show up and asked to be screen on the y-axis. And this is also downward sloping. And that says the government is kind of potentially getting some information.

And so and this just shows you this in regression form, which is that your probability of showing up depends both on the observable piece and the unobservable piece of your consumption.

OK, so then we can look at the results. So this is a CDF of the results. So another point is I really like showing CDFs of results for two reasons. I don't know if you guys have talked about this nugget. So there's two advantages. First of all, someone want to interpret this? What do we take away from this? Wesley had his hand up. Go ahead, Wesley.

AUDIENCE: So under self-targeting, the log per capita consumption, so you'll get to 100% of the-- you'll get to the full range of individuals applying earlier in the long per capita consumption under self-targeting [INAUDIBLE].

BEN OLKEN: Yeah, [INAUDIBLE], what were you going to say?

AUDIENCE: I guess the one that's further to the left is doing better.

BEN OLKEN: In terms of?

AUDIENCE: In terms of targeting.

BEN OLKEN: Yeah, identifying poor households-- Right. And just to be clear, this line first order is this line first order stochastically dominates this line. So it's like it's to the left kind of everywhere on the distribution.

So why do I like plotting CDFs as opposed to PDFs? There's two reasons. Number one, if you want to show your data non-parametrically, there's no statistics in creating this graph. There's no-- this is literally just the data, because why? How do you make-- what's a CDF?

Well a CDF is for every point, you just calculate what is like-- for each x, we're just calculating what is the probability that say income is less than some x, conditional on treatment equals 1 or 0. So we're going to do this for each of those things and just go over the whole range of different values. That's what a CDF is.

So there's no-- this is just like tabbing your data. There's nothing new to estimate here. When you want to start estimating PDFs, you have to-- a PDF is, as I'm sure you know the derivative of PDF. But you also need to smooth those things in practice. And again, that can depend on your smoothing parameters. This is literally just the data. OK, so that's nice because it's just kind of a way of showing your data.

The other thing is this shows you a lot of information because this also shows you the difference in your treatments at every quantile. And so to read that, you just read this off like, so what is the difference in between treatment and control evaluated at the 50th percentile of the outcome distribution?

Well, I just go to-- I go right here and I can read it the crossover here. And this gap here is like the 50th quantile. And so I can say, for example, is this having effects kind of lower or higher in the income distribution or whatever.

So that's not necessarily the most important in this context. But in a lot of other contexts, that kind of additional information can be really valuable. So anyway so that's the CDF.

So that tells you the difference between-- so you can see that the people who are selected are kind of uniformly left, they're poorer. This is showing you the probability of obtaining benefits. So this is graphing-- this is now estimating non-parametrically, what is the probability that you obtained benefits at each kind of level.

And what you see is you see two things. So the desired targeting range is over here. You see two things going on. Number one, this black line is basically much, much below the orange line. Why is that? That's because the rich people are not bothering to-- the rich people who might get through the proxy means test like by mistake are not bothering to show up.

And that's kind of the key-- I think the key thing that's going on in a lot of the paper is that these people are just not bothering to show up. And therefore even though they might have made it, the probability ex ante that they would have made it really small that they kind of decided not to do it.

The other thing which is interesting is the other point, which we did not expect, actually, is that the black line is greater than the orange line. And that's for this other reason at the very far left. That's this other reason I mentioned, which is that you try to figure out who should be screened in a door-to-door approach. But maybe you don't get everyone necessarily.

Whereas in an on-demand approach, people can make themselves be known and say, look, I'm poor. Please include me in this program. And so it turns out that actually this approach has both of these benefits of excluding the rich who you didn't mean to be-- who you didn't intend to target, and getting the-- allowing the poor to include themselves.

And the other thing I'll say is that relative to other kinds of approaches that use kind of costly ordeals as targeting. Some approaches are like workfare programs. So for example, probably the largest program in the developing world, anti-poverty program is the MNREGA program in India right now. And it's the largest-- it's one of the largest programs, very large program, 140 million people, I think. Certainly in the order of 100 million people or more in this program.

And what do they have to do in MNREGA? They have to go basically work. If you want to get paid, you have to go work through manual labor for a project for however many hours or whatever. And you get paid minimum wage.

That's a costly ordeal because they could have other things they could be doing with their time. Or they might not like manual labor or other things it's a very costly thing to do. It takes all of your time for 100 days. This, by comparison, is a super small ordeal. It's like half a day of your time.

And so I think what's really attractive here is to say, look, the Nichols Zeckhauser idea says that we have these really large ordeals we're going to do it. But here by going back to the pre-period and leveraging this uncertainty that people are going to have about the application process, maybe we can actually get a lot of those screening benefits with pretty small ordeals. And that's kind of what this shows.

AUDIENCE: One thing that I would be curious with this paper is even at the peak of the [INAUDIBLE], you have less than 20%.

BEN OLKEN: Yes, a couple of you guys mentioned this, yes.

AUDIENCE: So maybe qualitatively, do you have a sense of whether that's due to this issue you were talking before where still there's people who aren't announcing themselves, even with this self-targeting treatment or there's potentially stigma or--

BEN OLKEN: So I think there's two things-- so I think you're totally right. Why is this only 20%? That's not great. So I think there's a couple of things going on here. The first is the show up rate at the bottom is about 60%.

So part of it is 40% of the people here are not applying. That's not all. That's part of it. And I don't know if this is necessarily stigma. It could be some of the people actually do have challenges showing up. I think that's real. I'm not sure stigma was a huge problem here in this context, actually. But that could be.

But I don't think it's a real problem. I think it's more just people are busy. They have other things to do. There was a lot of effort to get the information out. Maybe not everyone got it. So I think that's right.

And I think that maybe doing this kind of multiple times maybe it might be really important to get this number higher. But the second point is that proxy means tests are imperfect. And I think that's the other piece of it. What's the probability you get benefits?

I didn't bring it in as a function of total. Oh, yeah, here, what's the probability you get benefits as a function of your total y? Well, it's only like 0.3. And that's because it's imperfect. And it's also imperfect precisely because they're trying to target a really low fraction of people. They're setting this really tight.

And that's kind of a policy decision. There's a broader question of this policy trade of, like do you want to target really tight? And you're going to make fewer inclusion errors, but also a lot more exclusion errors. Or do you want to target more broadly? And you're going to make fewer exclusion areas and more inclusion errors. And the particular program is targeted really tight.

So I think it's the combination of this fact and this fact. So I think it's more about the imperfections in the proxy means test fundamentally at targeting people who are this poor is I think more.

And I think understanding how to fix that problem and reduce exclusion is a first order issue that I don't think we have a good handle on. And actually, to me, to my mind, actually, I think one of the biggest issues that I don't think we really understand is how do you-- how do you reduce exclusion error in these contexts?

And I think that this approach says actually, despite the fact that only 60% of people or 60% or 70% of people applying here, on net, this approach actually reduces the exclusion error because it has this other benefit of at least we allow everyone the option to apply. So on net, this is better than the status quo.

But I completely agree with you. There's a ton of exclusionary here. And I think that how do you reduce that is really important. So I completely agree with that question. Other questions?

So the final point-- so then we want to understand the different theoretical mechanisms. So we're going to estimate the model. So how do we do that? So what do we have to do?

Well, the first thing is we have to parameterize some stuff. OK, we have to take those vague things like c of l and y and turn them into actual numbers that we can estimate, plug into a model in the computer. So parameterize some things.

So we're going to parameterize the cost as follows. We're going to say your wage-- the cost is equal to your wage times the travel time plus the average amount of time that you have to wait there. This is your expected cost. It's not your realization of the cost. The expected cost. So it should be-- it's how much you forecast it's going to cost you to apply for this thing, plus any monetary cost of traveling there.

And by the way, we were thinking about this in advance, so we collected the variables we needed to get this, estimate this. We just knew the distance. And we knew how people would typically get to these places and whatever. And we collected that in the baseline surveys because we had an idea we were going to interested in this question.

That's another, going back to the theory point, we thought this was going to be this costing thing was different, cost was important. So we tried to collect some data that was going to be useful to estimate it.

One thing, which by the way, is first order here, is consumption is probably measured with error. And so we have to deal with that. So we were going to parameterize the measurement error here. That's fine. We have a different data set with a short panel. And so we're going to use that where we measure consumption twice in a row for the same people and use that to parameterize measurement error.

And the reason for that is that we observe consumption measured with error. But we assume people make their decisions based on true consumption. So we want to parameterize that. Otherwise we're going to load too much stuff on the error term, basically, if you don't incorporate that.

We need this lambda term. We'll just make that-- we'll assume that's a probit. It seems sensible. And so the unknown parameters, we have to estimate our alpha, which is the share who are sophisticated. We have no idea.

The mean and the variance of the utility shocks epsilon and the parameters of the probe of the lambda distribution, which is people's-- which is what are the unsophisticated people going to think?

And so we're going to estimate-- so what we get out of the model is that for any given individual, i, your probability of applying for benefits is, well, if you're sophisticated, it's given by this-- what's the probability that your gain from applying as a function of the observables is greater than the-- less than the-- whatever, greater than the epsilon, greater than the epsilon cost, and maybe off by minus error. And the same thing for the unobservables.

So that's what the model gives us as a function of the unknown parameters. And so that's what we try to estimate. And so you went over a GMM last week. So this should all be familiar for you.

But I just want to say a word about GMM, which is you hear a lot, a lot, a lot about structural models in like many classes and many papers. And I think they are over-- like they have this whole hype around them. And they're not actually that complicated. And so I just wanted to spend one slide doing my best to demystify them because it took me a long time to figure this out. So I want to save you guys that time.

So what are we doing when we estimate a structural model? So we write down a model, which is like a theoretical model, which has some unknown parameters in it. OK, so in our case, I just wrote down a model, which has unknown parameters, which generates a probability that I show up as a function of individualized characteristics and some unknown parameters.

In our case, it was like alpha and the utility shock distribution and the data distribution or lambda distribution. What the structural estimation says is I would like to know these unknown parameters. So what values of those unknown parameters get the models predicted show up probabilities to match the actual show up probabilities in the data for similar people.

So the model is like probability of showing up for individual i is a function of these various things. And I just want to choose the values of those unknown parameters so that the models predicted show up and the actual show up in some data are the same.

OK, so to do this, you define moments. Moments are statistics of the data that you can also calculate in the model. And you need to have at least as many moments as unknown parameters. Or you can have more. More is fine, but you can't have fewer.

OK, and you want to say you then search for the particular thing you do is you just do a search over these unknown parameter vector, in this case five-- five-dimensional parameter vector, so that the moments from the model match as close as possible the moments in the data.

And that's why it's called the method of moments. So one sec. And that's it. That's the basic idea. Set up a model, the function of unknown parameters, calculate some moments in that unknown parameters. And search-- and keep mucking around with the values of the parameters until the values of your five moments hit the values of the similar five moments at the target five ones from the data.

So that's the main idea. So I said that's it. All the rest is commentary. So one important piece of commentary is, well, what happens if you have more parameters than moments. If you have more parameters than moments, you can't do anything.

If you have more moments than parameters, now I won't be able to generically hit all the moments because I don't have enough degrees of freedom. So one of the many handy things that GMM does is how should I optimally kind of weight the different moments to try to match the parameters as best I can. OK, sorry.

AUDIENCE: Oh, yeah. I was wondering how-- because I guess the thing we're matching here is probability. We all just learned [INAUDIBLE] models or probit models and metrics. I guess we're doing a very similar thing. I'm wondering like just conceptually, maybe like how does this compare to when I run a probit model in Stata.

BEN OLKEN: So those are maximum likelihood models. So there's two ways you can estimate a model. Sorry, this is the GMM version. The other version is maximum likelihood, where I specify the full kind of likelihood function from the data, and say, what was my full data generating process, and search over the-- and you're basically saying in maximum likelihood, you're essentially saying, what is the underlying set of parameters such that to maximize the likelihood that the data I observe actually came from that distribution.

AUDIENCE: But I can also fit my probit model to GMM.

BEN OLKEN: You can fit a binary variable model with GMM, yes. But the basic model you're learning in metrics is an [INAUDIBLE].

AUDIENCE: I guess maybe I misphrased the question. So it makes sense.

BEN OLKEN: And, Ed, by the way correct me if I'm saying it wrong.

AUDIENCE: That, what you just described is something that's a one liner in Stata, right?

BEN OLKEN: Which?

AUDIENCE: Like running the maximum likelihood of y on x. Obviously, this is not a one liner in Stata.

BEN OLKEN: No, it's not that many lines-- maybe, depending on the p set.

AUDIENCE: You don't necessarily have the likelihood function. So you have more information when you have the likelihood function.

BEN OLKEN: Correct.

AUDIENCE: OK.

AUDIENCE: So you wouldn't necessarily want to put a probit model in the GMM because if you're already assuming I know everything about the data generating process, then you can use all the information of the likelihood.

AUDIENCE: I see, I see. Whereas here we don't know how the parameters map [INAUDIBLE].

BEN OLKEN: We don't know the full data generating process.

AUDIENCE: But you do a couple of points [INAUDIBLE] it has to satisfy.

BEN OLKEN: Correct.

AUDIENCE: You're assuming [INAUDIBLE].

BEN OLKEN: I agree with that. It's a less restrictive-- you know less.

AUDIENCE: One point I was confused about is--

BEN OLKEN: Oh, sorry, let me say one thing before a question. But often writing down the whole likelihood function is very difficult. Whereas this thing, you can kind of usually generically do. Not always, you have to-- there are conditions for making sure that your model is identified, which is this idea that you have that you in fact can recover, that it's not the case that multiple different combinations of parameters would give you the same moments. That's an identification problem. But other than that, there's lots of cases in which you can actually do this. Sorry, go ahead.

AUDIENCE: Yeah just trying to understand, what you think about when you need experimental or [INAUDIBLE]?

BEN OLKEN: That is a great question. So where do the moments come from? And are the moments-- so I think that-- so the way to think about that I think is that your--

Let me see if I can craft a good answer for this question. In general, if you have a-- let me see if I can come up with a good answer, sorry, hold one second.

As I'm changing things in the moment distribution-- so imagine I'm computing to different moments for tall people and short people for example. In the model, I'm going to assert that the only differences between these people are short versus tall. And in the data, I'm going to just observe short people and tall people, short and tall, right?

If that variation between short and tall is essentially randomly-- imagine that was randomized, then in fact, those differences I'm going to observe, kind of those real moments, are in fact, going to be the true things that I should be plugging into the model because it's going to be the true, the causal difference between moving from short to tall.

Whereas in the real world, you might imagine that short is correlated with other things. And tall is correlated with other things. And so those differences between the short and the tall may not just reflect kind of the short differences and the tall differences in the modeling respect kind of other unobserved characteristics.

So in the same way that analyzing kind of a reduced form thing, you might be worried about omitted variable bias, you could be worried about the same things in structural estimation.

In practice, people don't always worry about that in the same way. But I think in general, it's the same kind of concerns. And so as you can-- as you actually see in this paper, a lot of what we're trying to do are things identify the model on things that look kind of experimental movements.

AUDIENCE: So kind of when you're writing a model, you're solving theoretical model should be thinking about what assumptions you can make.

BEN OLKEN: No, sorry, I don't understand the last part.

AUDIENCE: Like as in the time in which you're thinking about where you'd need experimental variation, that's when you're solving the model, you're thinking what assumptions you're making, like in the kind of functional forms you're choosing.

BEN OLKEN: Yeah, or another way of saying it is is like you want to say like what-- I mean another way of thinking about it is like from a design perspective. So here's a model. Here are the key unknown parameters I'd like to identify. What you want are you would like to have experimental variation in the real world that if I ran the same experiment in my model would give me really different answers depending on the parameter vector.

So if I really want to identify-- that's I think the best way to think about it. So if I really want to identify some parameter, I would take my model. And I would say, well, what if I change this in the world would I get really different answers depending on different levels of that parameter vector?

And that's going to say that this experiment is going to intuitively identify that parameter. And so I think technically what this is if you-- it's the-- I can never remember it. I think it's-- what you want is like d theta hat d moment, moment i.

So you want to choose experimental moments, things are going to actually vary in the real world that are going to give you-- that are going to identify the parameter vector theta that you're interested in. And by the way, Ed, if you want to add any commentary on this, since you just did a whole recitation on this, I'm open to suggestions.

AUDIENCE: Just to give you an objective point that is not that different ways of thinking about it.

BEN OLKEN: Yeah. OK, other questions?

AUDIENCE: I would say in the Bergquist and [? Bennerstein ?] papers [INAUDIBLE].

BEN OLKEN: What? Sorry.

AUDIENCE: In the recitation, we did Bergquist and [? Bennerstein. ?] And there were exactly two specific objects that you needed to estimate and design the experiment.

BEN OLKEN: And the experiment estimate them-- yeah. Another paper was also kind of around the same time is Gabriel Kreindler's job market paper from when he was a student here, which is a similar feature of he wrote down-- this is a transportation paper. So we're not teaching it in this class.

But had a couple-- the paper had a couple-- he had a model and a couple of unknown parameters. And he said, look, I really want to go design an experiment that's going to really give me variation that I need to estimate these unknown parameters.

And so it was kind of a case where he wrote down the model, figured out what kind of the unknown parameters were, and then designed an experiment that was explicitly going to get that kind of variation. So that's another nice one that is taking this relationship between model structural estimation to the ex ante phase and designing experiments that's going to estimate-- that's going to generate the variation you need to estimate the model clearly.

AUDIENCE: I don't know if this falls in the explanation part of the article part. You mentioned that, when you illustrate a model, you found that ordeal mechanism size wasn't the big deal and it was actually the probability.

BEN OLKEN: Where does that come from this slide. Not on this slide. It's going to come next.

AUDIENCE: OK, all right. But it's like could you have written down the model in a way that didn't allow you to distinguish those two. How did you know to write down the model in the way that you were able to [INAUDIBLE]?

BEN OLKEN: So what do you do once you have the model? I haven't forgotten your question. What do you do once you have the-- this is more about the calibration of the model. I'm going to actually-- this is all how do we-- oh, no, sorry, this is actually relevant.

So what do we do? So now we have the model. Now we can do different counterfactuals. So what does that mean is we can take the model with the parameters and start changing different things about the model and see how would the world behave in different counterfactual worlds.

So the first thing we did is we were like, OK, we think it's this whole different technology-- so the first-- backing up. We found selection. We didn't find very much between the close and the far treatment. That was weird. We were like expecting to find that. We didn't find that.

So then we were like, what's going on here? So our first hypothesis was this whole like different technologies for dealing with the ordeal. The rich take the bus. The rich the rich take the bus or take the motorcycle and the poor walk. That was kind of our idea.

So we're like OK, let's play with that in the model. We didn't actually have random variation in like forcing them to take the same technology. We probably couldn't have done that. But we said, well, can this be-- clearly and this was kind of qualitatively in there, in terms of how did you actually get to the place, or how would you get to this place. And they were doing different things.

But is this large enough to actually make a difference? So what do we do? We said, OK, let's parameterize this travel cost thing. And let's rerun the model taking this thing out. So we're going to rerun the model, but we're going to give everyone the exact same transportation technology.

And the answer was it made no difference, essentially. And that's how we concluded that this thing, even though it was qualitatively in the direction we were thinking about, was not quantitatively large enough to explain anything, at least in the context of our model.

On the other hand, we took the exact same model and we took out lambda and put in lambda bar. So we say let's assume everyone has the same beliefs about passing the thing. And that basically explains everything.

So this was quantitatively-- so both of these were like qualitatively in the right direction. But we needed some way of quantifying kind of which of these different things were large or small. And so by taking the estimated model and rerunning it under different kind of counterfactuals, we could say, well, which of these different factors are really important? Because they're all kind of in there. And you take them out kind of one by one in the model. And that's how we figured it out.

And this is what I was saying of like we didn't know this, but really we didn't know. We really wrote a paper like a PowerPoint presentation or a Beamer presentation somewhere on my computer, which is all about this one because it was like qualitatively there. And we actually had evidence showing they were taking different technologies. It was really great.

It just turned out that once we wrote down the model, we're like, oh, this is like two orders of magnitude too small or something to explain anything. So that was the point. Other comments?

And so this shows you a different counterfactuals or whatever. So for example, this is the coefficient on log consumption. This is the true kind of experimental coefficient, whatever consumption negatively predicts whether you apply. This is the estimated model. It's also kind of similarly negative, not so surprising because we fit the model to the data. Right it's not so surprising there.

But then we can see what happens we turned different things off. And like for example, if you turn off the-- if you give everyone the same kind of beliefs about the probability of passing, this thing just totally goes away. But you give everyone the same kind of technology of traveling or whatever, then that makes much of a difference.

OK, any other-- I'm going a bit slower than I had hoped-- but any other-- guys had lots of great comments. Oh, so let me hit a couple of the questions that you guys raised from your response on this.

So one was, I think we talked about how do you pick the moments. You want to pick moments where this is going to be true, and the moments you think are kind of well identified, where there's the difference in the moments is not correlated kind of [INAUDIBLE] ideally.

Are there any other questions that you guys-- we talked about the exclusionary point. Are there any other questions that you guys wanted to raise from your reading about this? I know it was like Wednesday. Sorry. Any other comments on this paper or questions?

So now you're going to actually practice doing this on the problem set. All right, so what I also want to say about redistribution, a few things.

So one is now that we know something about targeting, how do we think about welfare under this? How do we actually make some welfare judgments on this? So in order to do that, you have to take a stand on how do you think about exclusion error and inclusion error. And you have to be specific about writing on a model of how you think about those two different kinds of mistakes.

So here's one example of how we did it. But you may choose different welfare functions. But the basic point is you have to write down a welfare function. And you also have to specify how your programs change. So we made the following assumptions.

We said let's suppose everyone has CRRA utility where this is your baseline income and this is the benefits you get. And we'll assume some parameter value. I think we do three I think in the paper. Maybe we do two. I actually don't remember, two or three.

And we're also going to assume a fixed budget b. So there's a fixed amount of money the government wants to spend on this program. And we're going to say as we make the program more narrowly targeted, fewer people are going to get it. But the amount they're going to get is going to go up proportionately.

And we're also going to use the actual kind of proxy means tests to think about exclusion inclusion error. So what I mean by that is we go way back to-- way back, way back, further back, here.

So as I-- basically we're going to say, look, this is a fundamental technology that we're dealing with here, which says, here's your predicted consumption as a function of your actual consumption. And what different programs look like is we're going to draw the line in different places.

So we can have a really narrowly targeted program drawing a line here, like a universal program draws the line over here. The government can't necessarily-- we're going to hold this cloud as fixed and we're going to see what happens as we kind of move the line to make it more or less narrowly targeted, holding this technology as fixed.

Now I have to zoom forward again. Almost there. OK, so that's what we're going-- we kind of run a counterfactual. What we then do is say, well, what is the total kind of-- under this welfare function, which is-- you can make a different welfare function-- what is the total social utility taking these into consideration.

And in this case-- so, sorry, we're going to plot this as a function of inclusion error. So how much-- this is the most inclusive program this is the least inclusive program. Maybe we should have chosen a different-- we chose this to make it look like an ROC curve in statistics. But maybe we could have chosen different axes.

The point I want to make is in this example, it turns out kind of relatively narrowly targeted programs look kind of welfare optimal because you're targeting-- even though you have more exclusion error, you're targeting so much more to the poor in this simulation. But you could do it in different way.

The more general point I want to make is that you do have to-- if you're making some trade offs, you have to take a stand on how you evaluate the different trade offs. And then you can evaluate your decision based on that welfare assumption.

And you can choose, as I said, like if you're going to choose as we increase row here and put more and more and more weight on the very poorest, that trade off may end up looking different. So you can do different-- you have to take a stand on the welfare function and then you can say some things.

The other thing I'll note, though, is that you have the other problem here, which is not captured in this utility function, is horizontal inequity. So this is the fact that different like people, inequity is the probability that people who are alike are being treated differently.

So we compute that. What is the probability that people who are treated kind of-- who are about the same-- sorry-- about the same in terms of their consumption get different outcomes. And actually these narrowly targeted programs, they have a lot of horizontal inequity precisely because you're targeting in the thick part of that cloud of the PMT. And so you're making a lot of errors.

And so I think in general, you want to think about, from just maximizing kind of utilitarian social welfare metric, these narrowly targeted programs are great. But they have this real problem of horizontal inequity. And I think one of the advantages of the community-based approach is it can help reduce some of the challenges with horizontal inequity because at least people have some local buy in as to who's getting the programs and who is not.

OK, so that's what I wanted to say about targeting. Any last questions on that? So I wanted to say a little bit about the form of transfers. And then I guess I we'll talk about tax starting next time.

So there's now like a whole big literature about how do you think about designing transfer programs in developing country contexts? And how do you think about those issues?

And so I guess I wanted to highlight I think two issues. And then a third one is this whole question of poverty traps, which Esther talked about earlier in the semester. So I'm not going to talk about that.

But there's a bunch of questions like, should you give cash? Should you give in-kind stuff? Should you give productive assets? Should you give-- should you make transfers conditional on activities or not? Should you give a large, one-time transfer or lots of small transfers?

Should you give people workfare programs like the MNREGA program I was talking about? Should you give them cash? How should you run the program? Should you give cash versus electronic payments? How do you identify people for these programs? Should you use smart cards? So do you use biometric authentication, and so on and so forth.

I think all of this is actually a pretty active research area. I think this was kind of has been an active research area for the past 10 years or so. I think this has become a really active research area in the past couple of years.

But I think that both of these areas are areas where people are thinking a lot about how do we effectively design and deliver these programs in a developing country context, where we have a lot of information challenges of the sort we saw in targeting, where we have a lot of governance challenges of the type we talk about in the political economy classes, and so on and so forth. Like how do you actually do this effectively?

So I just want to highlight I think two examples or three here. So one is about cash versus in-kind. So what does this mean? Cash is they literally just give you cash. And in-kind is I give you stuff. So how do you-- before I get into it, how you think about the difference between cash and in-kind? Like put on your economist hats, which I should always be on in this class. Like how do you think about the difference between cash and in-kind? What are some of the kinds of things you think about?

AUDIENCE: If people are rational, we should be giving them cash because you can't prescribe the highest utility [INAUDIBLE].

BEN OLKEN: Right, so economists love cash for the reason Rega just said, which is that we should be like trusting people to make their own decisions. And I don't know what they want, but they should know what they want. And with cash, they can just buy it.

AUDIENCE: Public support for a program that is in-kind might be higher. So if you need to [INAUDIBLE].

BEN OLKEN: Why?

AUDIENCE: If people find out-- if people don't agree with the types of things that people are choosing to buy, even if it is maximizing for them, society might feel like they should have programs, so to the extent that-- especially in the United States. That's why we would restrict people's food stamps because even if buying alcohol might make people feel better or enjoy it more.

BEN OLKEN: Yes, exactly. People may have paternalistic pressure-- political-- political reason people want to-- only want them to use the money on certain things. Patrick.

AUDIENCE: You might be able to address some of the targeting issues with [INAUDIBLE] cards.

BEN OLKEN: What, sorry?

AUDIENCE: You might be able to address some of the targeting--

BEN OLKEN: How?

AUDIENCE: If you target inferior products.

BEN OLKEN: Yeah.

AUDIENCE: [INAUDIBLE]

BEN OLKEN: Yeah, exactly. So if I give out inferior goods, then basically rich people may not want them. That actually may have good targeting properties. Absolutely. Other things you might want to think about? Yeah.

AUDIENCE: If we do an in-kind transfer, it could also help the suppliers. Like if we really like to want to help some rice company and give a lot of poor people rice, the government would be helping the rice industry.

BEN OLKEN: What are the conditions for which that would have to be true?

AUDIENCE: It would have to produce the supply shock that wouldn't otherwise happen.

BEN OLKEN: Right. So yes. So you have to have a supply shock. So when is it going have a supply shock?

AUDIENCE: Like if--

BEN OLKEN: So, sorry, let me back up. Are there some conditions where an in-kind transfer would have not generate a supply shock?

AUDIENCE: Like if they were going to buy it with the cash.

BEN OLKEN: Yeah, they're going to buy with cash anyway. Is that what you're going to say? Yeah. So imagine that I'm going to buy-- like in Indonesia, for example, everyone eats a lot of rice.

But suppose everyone needs 30 kilos of rice. And I hand out everyone 5 kilos of rice. That should be totally irrelevant in terms of supply because people are going to buy that rice anyway.

That's true as long as the market-- as local markets have elastic supply. If the markets are themselves have not inelastic supply, even if people are going to be consuming all that rice, you might be like moving physical rice and that actually may change through the aggregate supply in an area.

So it's actually a little bit complicated. So number one is, on the demand side, it's not-- on the demand side, if people were going to consume all that anyway, it's not automatically going to be supply shock. But weather there is kind of depends on what the underlying market structure looks like. So there may or may not be a supply shock.

Any other issues you might think about?

AUDIENCE: Insurance issues, so you're saying that safer food, like to give it in-kind, then people don't have to worry about fluctuating food prices over seasons. I guess it's like a form of insurance.

BEN OLKEN: Exactly. And there's a brand new paper on exactly this point called in-kind transfers as a form of insurance, which came out like in the last year or so, Was actually presented at the NBR meetings just a few weeks ago.

And which makes exactly this point that typically when we fix transfer amounts in-kind, we say we're giving you a certain quantity of stuff, 10 kilos of rice, an apartment, whatever. Whereas when we fix things in cash, we typically say we're going to give you a certain amount of cash.

Now, in principle, you could adjust it either way. I could say I'm going to give you the equivalent of 10 kgs worth of rice at some-- or whatever. Or I'm going to give you the equivalent amount of rice that I can buy with 100,000 rupiah at market prices. And I can move the quantities around.

Or vice versa, I can index the transfer to inflation, for example, or some particular types of goods. But in practice, that's imperfect. And therefore, I think they do actually provide that in and of that kind of in-kind assurance.

I think I'm out of time. Let me stop here. I'm going to talk about one issue next time, which is the supply shock issue. But all of these are kind of relevant. And then I'll talk about-- the last two papers I'm going to talk about in this section before I talk about tax, I'm going to talk about this one, the Cunha et al. paper on supply shocks and in-kind transfers. I'm going to talk about the cash versus condition paper by Baird, et al. on conditional cash transfers.

And then I promise we'll switch to tax next time. And so you shouldn't be-- I think the reading for today was the tax reading. So we'll talk about that one next time. OK, Thanks.