

Federal Reserve Forum on Consumer Research & Testing: Tools for Evidence-based  
Policymaking in Financial Services, November 9, 2010  
Panel Two: Exploring research methodologies for consumer testing and studying consumer  
behavior

Adair Morse:

Okay, so my hero is Petra Todd. They put this up here 'cause if you need a reference, she has wonderful slides on methods. And Heckman and Rubin are brilliant; reading their papers can be challenging. Petra Todd does a beautiful job of writing summaries. So if you need that, go to that.

Can experimental methods advise policy outcomes both for ex ante policy implication and ex post can be--we're in a federalist system, you know, so you maybe want to evaluate state implementations and apply them at a federal level or not, right, or state level, so forth. So the types of policies, this list could be hugely extended. Disclosure is a big one right now, and I'm going to kinda focus on that. But limiting choices – we heard before that some comments of that. Opening and closing of a market education. We talked--Margaret talked about – lots more. Okay. So field experiments are in the Dodd Frank's, Jeanne sent me this, trial disclosures are written here in the bill saying, you know, we should test these things before just implementing. Alright. I'm not gonna--you can look that up if you want more on that.

Okay, so what I wanna do is I wanna lay out some identifying assumptions. In other words, if you're going to do experiments, whether field experiments or natural experiments, you wanna make sure that after all that effort, you got it right. Right? So that's the goal here. I'm gonna lay out these assumptions in words. I'm not gonna put any math up today. And you can go to Petra, she's good. In the process, we'll talk a little bit about behavior and heterogeneity, which has already come up as super important.

Okay, so here's a problem statement. Let's assume a policy has been implemented, and we observe the outcome, and I'm gonna use this treatment terminology, we observe an outcome of those with treatment and those without treatment. The problem is we don't observe the counterfactual of what those who were treated would've looked like had there been no treatment. Okay. It's the simple program evaluation literature. Right? So what the treated would have looked like, this--we refer to this is the average impact of treatment on the treated, and then we also really--what we're interested in as policy makers is this average treatment effect, what's--what the effect to the policy treatment would be on the population. And the difference in there--these two points, is that those treated may not be representative. This is a huge issue, right? That if you're going to implement a policy, you need to make sure that your design of evaluating either--say, one state's policy or the trial disclosure – that you're getting representative sample, and the treatment might have spillovers.

Okay, let me just give you two quick examples. Mortgage disclosure. We could change mortgage disclosure today and say, "Oh, well, let's match people up on income and then we

could compare them." Well, the problem obviously here is that well, the environment of what the disclosure has an impact now after the news reports about disclosure and the bad product, this and that, is none at all the same. So this comparison is completely invalid.

Second example, financial literacy. Okay, the research's long struggled with the causality of teaching finance in long-term outcomes. Alright. Lusardi's got work in this, and Margaret's done a lot of work, that--and it seems like--and I apologize—Margaret's work gets around some of these issues, but you know, we love the idea of financial literacy, but when we talk about doing testing, often we say, "Well, we're gonna have this program and you've got voluntary selection in." We know though that people that select in are not representative of the population at large, and that affects outcomes. Okay. So we have to deal with these things statistically before we implement, go--you know, gung-ho with these experiments. We really need to be serious about these things 'cause we want to get it right.

Okay, I'm gonna talk about three--I'm sorry I'm talking fast, but I wanna cover a bunch of things. I'm gonna talking about field experiments, which in my opinion is the best method, that statistically randomized control trials are the way to go if you could do it. I'm gonna talk about natural experiments which can be very appropriate and regression discontinuity if I get it--to it at the end.

Okay, field experiments, randomization. Randomization is beautiful. You can have a control group not exposed to the policy treatment. You can make sure that that control group is representative. Very, very important that this counterfactual of what they treated would have looked like had there been no entreatment. Critical to be again, precise upfront in the design and I'm gonna talk about in a second. Randomization, you know, there's some math behind this in determining how big of a trial you need, but this is not infinity, you know. So--but getting it right upfront, again, get the math done, so you have enough power to identify what you're trying to identify. Okay. So I'm not telling you how to do it, I'm just saying know that that's important.

Okay. Internal and external validity, here are the kind of the two biggies in terms of getting the field experiment right. Internal validity, must--the people, that are treated must have the same distribution as the people that are the control sample. Right. When I--I'm using this terminology as if we're doing drug testing, right, 'cause that's where it came from. But here's the hard part. There must be the same distribution on observables and unobservables. Okay. So matching on income and education, which is nice and lovely, it doesn't work unless you can make an argument that those are completely representative of the things you care about. I think I give an example maybe on the next slide of what--a case that that might not be true.

Okay, choice-based participation. You know, this is--I already kinda said this. It always struggles with representativeness is because the people that choose may look the same in income and education, but there's something about them, their behavior – we're talking a lot about behavior – that makes them agree to participate. And since now we're actually paying attention

to behavior, whatever that is, is going to--may matter. It may matter for--we were just talking about privacy, it probably matters for privacy, agreeing to participate. Uh-oh, wait, this doesn't count in my time.

Okay, good. Alright. Alright, so yes, so it's often better rather than doing choice-based participation to have the control group be within those that agreed to participate. Okay. Now that you could say, "Well, wait a minute then, you still have this bias here of choice--you know, who's going to participate." Now, we're talking about external validity. External validity--or that -- does that group that chooses to participate, are they representative of the population at large. And this is certainly a problem. It's less of a problem than this is. So get this one right first. Okay. External validity certainly matters for--when you're going large, you know, the whole country with the program, but it's less crucial. This is--I borrowed most of this from Petra, this slide.

Problems that can arise, I'm not gonna go over these so much. There are lots of them but I do wanna go over this one, which I enhanced. Ethical concerns. So you know, withholding treatment, you know, if you've got the cure from cancer, the medical research struggles with this a lot. Right? It's hard. We're in the same boat here for policy--you know, political unfairness. If one jurisdiction's decides that they're not gonna withhold treatment from someone, in addition to getting into the morals of that, but statistically, you have implementation bias there as well. The one I really care about, the moral part of lowering people's utility.

So we talk a lot about innovations, and it's really nice, and I'm gonna pick on Hal for a second 'cause he's--I can apologize to him 'cause we're in the same city, but say, you show a picture of me, old, right? You show a picture of me, old, but he's gonna show you it's kinda cool stuff, so--and I--when I'm taking out cash or I don't know what, and--but it made me feel bad looking at me old. Now, you've lowered my utility. Right? And so, I think we need to think about these sorts of things about where we're going with consumer finance, with all these data, and things we're learning about people's behavior, 'cause utility is not just about money. Sorry Hal, I love your stuff. But sorry, I had to have an example.

Alright, I'm gonna cover this paper real quick 'cause I think it's relevant that I wrote with Marianne Bertrand. So, even--so the idea here is payday loans, even if payday loans--let's just forget about the debate about whether they're good or bad -- even if they're priced fairly, you have to think that cognitive limitations are biased by some borrowers might explain their use. Okay, so I the idea here is that rather than--our experiment is what if you mandate disclosure? There's a particular type. It's a de-biasing disclosure. So it's informed to what mistake is being made. This is the key content. You know, for a particular decision--this is not general, it's a particular decision -- what mistake is being made, you could target the information with some debiasing strategy, right? And so, what we do is we do a field experiment in national chain of payday stores and see if we can impact borrowing--future borrowing. So the idea is what's the mistake? Well, maybe the mistake is that people may not internalize the APR. Okay? So what

do we do? We make very salient the APR. Okay? Payday loans are expensive. And we put it in context in comparison to others, not that they can switch over to these other products, but just to make it look expensive. Okay?

Alright. Treatment number 2. Maybe people don't add up the cost of the decision over time. We know it's true. There's narrow bracketing, narrow framing, the psychologist would say Peanuts Effect from Markowitz. So we add up the cost of a payday loan. If you don't know a lot about this product, the fees add up from cycle--pay cycle to pay cycle, and in three months, if you took out a 300-dollar loan, you're paying 270 in fees. Okay, so we make that stark in terms of dollars.

Information treatment 3, maybe people fail to consider adequate variants and the future outcomes or they're overconfident about their abilities to pay back the loan. So we give them a distribution of refinancing to reset their expectations, so just to give them an idea of how other people struggle to pay back this loan. Okay? Now, the psychologist in a room will say fairly, which is true, "We've made--we've brought attention to the expense and offer it." Okay. So we can't disentangle that, but we're gonna look at the economic magnitude of one versus the other.

Random assignment, right? So this is a methods talk. Let me just mention here, we randomize--we pay a lot of attention to get this randomization right and we test characteristics, we do testing of whether this randomization across the different treatments was fair. Challenges, you know, in the paper, you kind of--you just write a line or two about--well, we did this and that. We struggled a lot. How do you get the clerks to be uniformly giving the information? We did lots of training. The stores are not comparable. So you cannot randomize by store, you can't randomize by customer because the clerk can't keep track of it, so you randomize by the day of the week, but you get a day of the week effect. The people that come in on Friday are not the same as Monday, right? I'm just trying to illustrate. It's hard.

Estimation details. Well, the observation-counts are not the same per store. You could weight the regressions by count, but also, stores have shocks that everyone in that neighborhood have--or face those shocks so it is not independent so we cluster by store, becomes very restrictive econometrically, just trying--should I--it's hard. Okay.

Results, what we find is the dollar information treatment, you know, that adding up, not the APR, the adding up is what lowers borrowing. And we do it on a number of methods. Now, again, what we cannot reject that bringing attention to one or the treatments or another is at the cost, but the magnitude is the largest and most consistent for the dollar information. I highlight this in red. If you're running different treatments, you have to pay attention, of course, to the fact that if you run a hundred treatments, one of them is gonna show up significant, or probably five, right? So you have to do some joint test. And in our first round, we forgot to do that and we had a good referee remind us.

Alright, conclusion of that paper. De-biasing failure to add up over time reduced borrowing. The dollar treatment--then we do heterogeneities that--I'm not gonna cover here, but the dollar treatment is most effective on those without higher education and the self-controlled were able to act on that information. So the paper generally advocates for really understanding that that bias, whatever it is, that's affecting the decision at that point in time and then having de-biasing disclosure, okay?

Alright. Alternative methods--I have 5 minutes. So I wanna just quickly talk about natural experiments and regression discontinuity. Alright. So, natural experiments are observational data and so, within the population of interest, a sample is exposed to a treatment. So the--in the context here, maybe one state has already implemented a different privacy law or something like that, okay? And the paper I wrote about payday loans is about natural disasters and using natural disasters as a natural experiment for financial distress. So it wasn't caused by your decisions or by what's going on in the community. There are buzzwords here, difference in differences, triple propensity score matching, all these things, right? Let me just go through a quick example. So a state implements a policy. Difference in differences would be to compare that state to another state over time and see if their trends are impacted by that. The problem with that is, of course, these things may not be comparable. So here I have--for example, there's a disclosure change in New Mexico or Delaware and say they match on observables, which they don't. But we know for a fact, you know, New Mexico and Delaware are gonna have different culture, have different perceptions of regulation, different social provision of good, so it's hard to do these differences just matching at state level. You really have to pay attention to the details. Propensity score matching is just a method to match on all observables. And you have to then assert that there're no missing factors. It's hard. If you can do it, it could be very powerful, and of course, cheap, 'cause it's observationally--it's observational data. Okay. Yeah, I'm not negative about this, I'm just--it's just hard.

In fact--so my other paper which was actually my dissertation and I got a job off this, so it seems okay, that--what I did was to look at payday lenders again in California, and I did this triple differencing thing with around the national disasters. You know, the problem is, of course, locations with payday loan stores are very different than locations without payday loan stores, even in natural disasters, they're exogenous. So I needed this triple difference because, basically--I don't have time. Let me just explain. So basically, what I assert is that if you have a trend, you can get the trend between two locations with payday loan stores. And you could say, "Well, the change in welfare is a pattern, and those two haven't been hit--those two assert the pattern of the trend, the payday loan stores and welfare." I'm not saying this well. But when--what I'm differencing around is the disaster, the difference in the pattern. Let me show you an equation. So for example that we can look at, think of this "W" as foreclosures, that foreclosures in locations that have a payday loan store that are impacted by a disaster compared to places that are impacted by a disaster that don't have a payday loan store, but we need to benchmark that against the pattern that we would've seen for locations that don't have disasters but with and

without payday loan stores. So I apologize if I didn't say that. Well, the idea here though is that we have this natural disaster, but it's not enough. You know, you really have to pay attention to what the patterns would be to get the natural experiment going. By the way, I find that for foreclosures in small crime that the existence of a payday lender helps people in disasters. Now, you--and that--I usually get an uproar from that. The--you know, there are people that borrow from payday lenders that are not in financial distress. I'm not speaking to those people, they use the finance in that setting, but at least, provide--this provides evidence that we need to think about alternatives for people in distress before going after generally, that this is crazy in terms of finance, okay?

Alright. Regression discontinuity, which I can do in one minute. Alright. So assumptions, the idea here is--and I'm gonna show just a picture rather than talk about the switch, I know it's fuzzy, but I borrowed this, okay? So the idea of regression continuity is there's some rule. So one of the famous starting papers was--the rule was about college admissions. You didn't know what score on the SAT would get you into Dartmouth but when you're sending your application, but there was a distinct rule there. And so, then looking at just the people right around that rule, because they were close on all other observables, right? And so, it's--this right around the rule, what we're looking for here is the discreet jump in securitization of mortgages right around the rule of the FICO scores at 620. This is from my colleague [inaudible], as well as [inaudible] from the board here. And so, you see at 620--I know it's fuzzy, but I had to paste it--that there is a jump in the rate of securitization of mortgages. And so, they can look at just these people and speak to the loan performance of just around this rule, okay? So regression discontinuity provides another kind of observational way to think about evaluating policies, either before or as you're implementing, if you can think about what--where there might be a rule that doesn't affect the consumer decision, but it's going to affect the outcome, right?

And I'm sorry I went so fast, but I wanted to cover those three things. Field experiments, just to wrap up. I know, stop. I know. Field experiments allow for randomized controlled trials, which I think are really valuable. Policy implementation is too important 'cause it's hard to change policy. So it's too important to run--to not run trials. We're in this second best behavioral world, and honestly, it's hard for economists to guess outcomes. I mean, it's just true. And this, you know--if you're thinking about the cost, it's cheap relative to missed policy. Okay? But if you're gonna do this, take the time to get the econometrics right before you start and make sure the implementers don't have an agenda, okay?

Alright. Don't exclude other--these other possibilities that I squeezed in, the natural experiments and regression discontinuity because they can be valuable, but again, you know, you need to make sure the identification restrictions. Alright, I know.