Policy Evaluation: Methods for Testing Household Programs & Interventions

Adair Morse
University of Chicago

Federal Reserve Forum on Consumer Research & Testing:
Tools for Evidence-based Policymaking in Financial Services

November 9, 2010

Goals – Policy Evaluation Pre-testing

- Acknowledgement: My hero is Petra Todd at Univ. of Pennsylvania. She posts wonderful slides and summaries of methods.

- Can experimental methods advise policy outcomes?
  - Methods for ex ante and ex post
  - Types of policies this might be relevant for in consumer finance:
    - Disclosure,
    - Limiting choices,
    - Opening/closing of a market,
    - Education & training,
    - Etc… many more

Field Experiments

Dodd Frank

Title 10, Subtitle C, Section 1032 (disclosures), (e) Trial Disclosure Programs:

(1) IN GENERAL.-The Bureau may permit a covered person to conduct a trial program that is limited in time and scope, <u>subject to specified standards and procedures</u>, for the purpose of providing trial disclosures to consumers that are designed to improve upon any model form issued pursuant to subsection (b)(l), or any other model form issued to implement an enumerated statute, as applicable.

This talk... Methods & Design

- Lay out some identifying assumptions (in words) needed <u>in order that the statistics one uses for policy evaluation are valid</u>
  - For math, go to Petra's website
  - Give examples of research (mine and others)

In the process, talk a little about

- behavior
- heterogeneity of response

Problem Statement (treatment terminology)

Assume a policy has been implemented…

- We observe outcome of those with treatment and outcome of those without treatment.

- But we need to observe the counterfactuals:
  - What treated would have looked like had their been no treatment
    (ATT – average impact of treatment on treated)

  - What the effect of the policy treatment would be on the population
    (ATE – average treatment effect)
    - Those treated may not be representative,
    - And treatment might have spillovers to untreated

Quick Examples

- Mortgage Disclosure…changes in disclosure requirements now
    - Cannot look at outcomes as comparable to pre-crisis.
    - Disclosure not identified vs. new environment

- Financial Literacy…
    - Research has long struggled with causality in teaching of finance and long-term outcomes
    - Long known that the voluntary selection into financial literacy program "treats" those needing it least
        - (Bruce Carlin has a nice new paper on the "finance fair" for high school students in L.A. No choice-based participation.)

Econometric Solutions

Talk about 3:

Field Experiments (best)

Natural Experiments (can be very appropriate)

Regression Discontinuity (tough, but good if assumptions hold)

Advantage of <u>field experiments</u> is randomization

- Control group not exposed to the policy treatment is representative of the counterfactual outcome of the treated, on average
  - Critical to be careful in not taking anything for granted in design

- Randomization…. Needs to be designed large enough in sample size to give power to estimators to get result,
  - BUT this is not infinity.
  - Depends on number of treatments, variance of outcomes without treatment

Internal & External Validity

- Internal Validity : Must have same distribution of observables & <u>unobservables</u>
  - Not sufficient to look the same on <u>important dimensions</u> (say, income or education)
  - <u>Choice-based participation</u> always struggles with representa-tiveness, even if match people on all possible demographics
    - Behavior of participation is usually correlated with something
    - Often better to have control group within the participants and then try to argue external validity

- External validity: Experiment forum must be representative across distribution (not just at the mean) of population of concern
  - Ie.. Can we apply results to those we care about?

What problems can arise in field experiments? (Petra)

- *Randomization bias* – occurs when introducing randomization changes the nature of the program

- *Contamination bias* – occurs when control group members seek alternative forms of treatment

- *Ethical considerations* – Politically: unfairness. Statistically: implementation opposition (because of unfairness) may mean some refusal to participate, resulting in bias. Morally: lowering people's utility

- *Dropout* – some of the treatment group members may drop out before completing the program

- *Sample attrition* – may have differential attrition between the treatment and control groups

"Information Disclosure, Cognitive Biases and Payday Borrowers" with Marianne Bertrand

Topic: Even if payday loans are priced fairly and non-predatory, one has to wonder whether cognitive limitations or biases by some borrowers explain the use of payday loans

<u>Idea (not just for this setting)</u> : Mandate disclosure that is

- Better informed as to what mistakes are being made
- Better targeted to de-bias potential cognitive biases causing these mistakes

<u>Field experiment</u> at national chain of payday stores
- Can we impact future borrowing with debiasing disclosure.

Information Treatment 1

Potential problem : People may not internalize APR because focus in store is the dollar fee structure on the wall.

Treatment: Reinforce understanding of APR by presenting it next to other (smaller) APRs.

Annual interest rates on different types of loans

	Median Annual Interest % (from government surveys)
Payday Loan	443%
Installment Car Loans	18%
Credit Card	16%
Subprime Mortgages	10%

Information Treatment 2

Potential Problem: People fail to add up cost of single decision over time

Behavioral Concepts:
      Peanuts Effect (Markowitz, 1952)
      Narrow Bracketing
      Narrow Framing

Treatment 2: Present additive dollar costs of payday loan fees into future

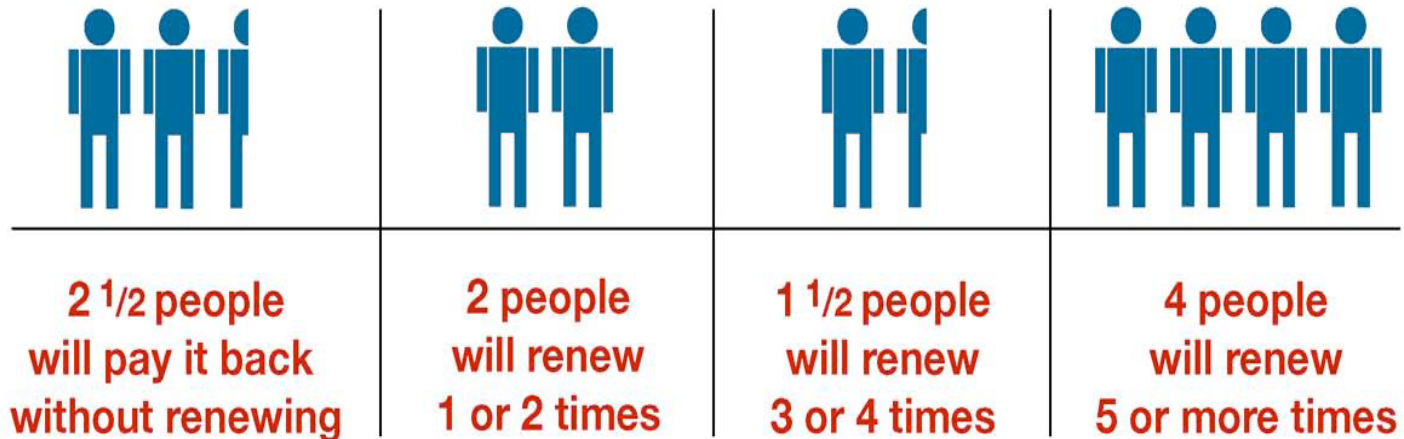## How much it will cost in fees or interest if you borrow $300

PAYDAY LENDER (assuming fee is $15 per $100 loan)		CREDIT CARD (assuming a 20% APR)	
If you repay in:		If you repay in:	
2 weeks	$45	2 weeks	$2.50
1 month	$90	1 month	$5
2 months	$180	2 months	$10
3 months	$270	3 months	$15

Information Treatment 3

Potential Problem: People fail to consider adequate variance in future outcomes/ People are overconfident about their ability to pay back loan quickly

Treatment: 3 Present distribution of refinancings to reset expectations

Out of 10 typical people taking out a new payday loan...

| 2 1/2 people will pay it back without renewing | 2 people will renew 1 or 2 times | 1 1/2 people will renew 3 or 4 times | 4 people will renew 5 or more times |

Random Assignment...

- Randomize at the store-day level
  - 77 stores, 12 days/store
  - Difficult to randomize at person level: errors by busy clerks in recording who got what treatments

- We test whether treatment is correlated with a host of pre-treatment characteristics
  - Income, age, amount borrowed, frequency of borrowing, education, etc..
  - Results consistent with randomly applied treatment

Challenges of implementation (a sample of headaches)

- Implementation challenges:
  - Training store clerks to be uniform!

- Randomization details matter:
  - Stores not comparable, cannot randomize implementation by store.
  - Cannot randomize by customer, impossible for clerk to keep track
  - Randomize by day of the week, but need distribution across days of the week, because borrowers on different days of the week not random

- Estimation details matter:
  - Observation counts not the same per store
    - Could weight regression by counts
    - But also may not be independent (same shocks faced by location)
    - Cluster by store – very restrictive on econometrics

Effect of Information Treatments on Indicator for Taking out a Loan

	No Individual Fixed Effects	Individual Fixed Effects	Individual FE & 1 Year Prior Only
Savings Planner	0.006	-0.011	-0.020
	[0.022]	[0.020]	[0.019]
Dollar Information	-0.067**	-0.059**	-0.050**
	[0.027]	[0.023]	[0.022]
APR Information	-0.035	-0.035	-0.051**
	[0.022]	[0.023]	[0.025]
Refinancing Inform.	-0.038	-0.037	-0.034
	[0.028]	[0.028]	[0.029]
Post Intervention	0.030	0.038*	0.045**
	[0.021]	[0.021]	[0.021]
Observations	229,862	229,862	46,194
R-squared	0.197	0.401	0.381
All three: columns Errors Clustered by Store. Day of the Week and Store*Year Dummies.			
Tests for Joint Significance of Information Treatments			
p-Value	0.075	0.053	0.094

Conclusions of paper

- De-biasing failure to add up over time reduces borrowing by 10%+

- <u>Heterogeneities:</u>
  - Dollar Treatment most effective on…
    - Those without high education (experience most new information?)
    - Self controlled (able to act on information?)

- Paper advocates for
  - Understanding the specific cognitive biases that may lead to suboptimal decision-making
  - And subsequently designing some correcting or "de-biasing" information disclosure

Alternative Methods

Natural Experiments and Regression Discontinuity

Example: Some States have implemented a policy. How can we precisely evaluate its impact at large to a federal rollout.

Note: Any State assessment of its own policy is by definition biased.

Natural Experiments

- Nature provides an experiment in observational data
- Within the population of interest, a sample is exposed to a treatment (a policy change, a shock to accessibility of finance, a nature-caused distress, etc)

- Buzz words… difference-in-differences, triple differencing and propensity score matching
  - Important to make sure identifying causally

Examples to highlight models

- A state implements a policy
  - "Difference in differences"
    - Compare to other states over time
    - Need to be able to assert (hard):
      - Change in outcome, say mortgage defaults, in the state with the policy would have been the same as the comparison state had no policy change been implemented in the treatment state.

      - Cont..

Matched difference in differences

- Match policy change state with a state that matches on observables that affect the change in outcome
  - Propensity score matching: a method to match on all observables
  - Need to assert… no unobservable factor missing
    - Eg…. New Mexico and Delaware may match on observables,(they don't), but culture, perception of regulation, social provision of goods, and… affect behavior
  - Also need to assert that
    - Policy caused no spillover effect on untreated
    - Policy effect in New Mexico is representative of population that policy designed to help (United States)
- I'm not negative on this… it is just hard.

Matched triple difference paper:
"Payday Lenders: Heroes or Villains"

- Look at communities (zip codes) in California with and without payday lending to see how the financial distress of natural disasters affects them

  - Fraught with endogeneities of disentangling impact of payday lending from intrinsic community economic variables that determine welfare outcomes

- Test welfare outcomes as measured by foreclosures and crime

Empirical Methodology

- Imagine a model of access to payday loans:

$$Payday\ Effect = \overline{\omega}^{Payday} - \overline{\omega}^{NoPayday}$$

- <u>Concerns</u>
  - Location decision of payday stores endogenous
  - Community economics can impact welfare directly, not just through payday effects

Exogenous Shock
Treat Communities with Natural Disasters

- Natural Disasters are exogenous shock to Distress and are unrelated to location

- Look at changes in welfare after the disaster.

$$Payday\ Effect = \Delta\overline{\omega}_{Disaster}^{Payday} - \Delta\overline{\omega}_{Disaster}^{NoPayday}$$

- But Conditional Mean Independence Assumption not met
  - Communities with lenders have different pattern of welfare changes

Make Communities Comparable

- Location of payday lender is endogenous
  - Match communities on ex ante demand for loans

- Even after propensity score matching, payday location may still be endogenous
  - Use communities that are not hit by disaster as a benchmark for payday and non-payday economies diverging
  - Making the assumption that changes are comparable, not that disaster and non-disaster communities are comparable
  - Using matched set:

$$Payday\ Effect = \left(\Delta \overline{\omega}_{Disaster}^{Payday} - \Delta \overline{\omega}_{Disaster}^{NoPayday} \right) - \left(\Delta \overline{\omega}_{NoDisaster}^{Payday} - \Delta \overline{\omega}_{NoDisaster}^{NoPayday} \right)$$

Regression Discontinuity Methods

- Assumptions
  - Analyze outcome around a rule determining who gets treatment
  - Rule threshold is unknown or placement in threshold rule evaluation is not controllable

- Examples :
  - (original) Thistlewaite and Cambell (1960) evaluate the effects of national merit awards on career aspirations
    - Those around making national merit don't know what cutoff is

  - Seru, Keys, Vig (2010) evaluate effect of securitization. Loans above credit score 620 easier to securitize
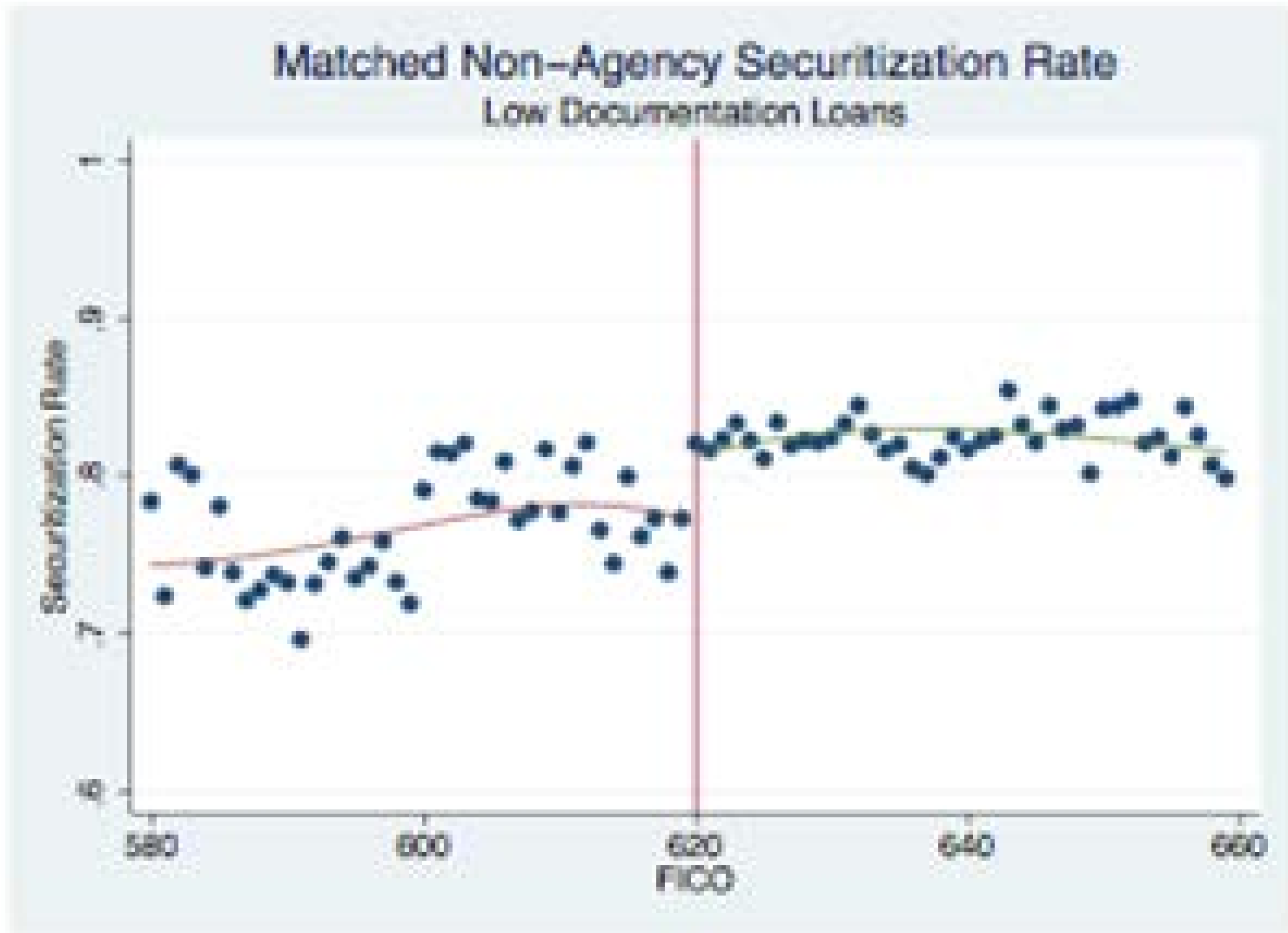   - Credit score observable, but not controllable

Figure showing discrete jump in securitization rate at fico score of 620

Wrapping up

- Field experiments allow for <u>randomized control trials</u>
  - In my opinion, policy implementation is too important not to run trials before changes
    - We are in a second-best, behavioral world, hard for economists to guess outcomes
    - Cheap relative to mis-policy
  - But get econometrics right ex ante… take the time
  - **Make sure implementers have no agenda**

- But don't exclude possibility of <u>natural experiments</u> and <u>regression discontinuity</u> to speak to policies
  - Sometimes setting can be exploited with just observational data, especially because we are a federalist system with lots of laws