Comparing Experimental and Non-experimental Evaluation Designs Using a Large-Scale Randomized Experiment in Environmental Policy

Juan Jose Miranda & Paul J. Ferraro
Georgia State University

Central Research Question

- Can non-experimental evaluation designs replicate experimental designs?

- Design-replication study: Use randomized evaluation results as benchmark to examine the performance of non-experimental evaluation designs.

Design Replication Studies

- Lalonde (1986): econometric estimates often differ significantly from experimental results.
- Glazerman et al. (2003): nonexperimental estimators often produce results dramatically different from randomized evaluations and that the bias is often large.
- Much of the focus is on matching and OLS in Labor Economics. Much less work on panel, instrumental variables, or regression discontinuity designs.

Environmental Policies

- Frequently implemented or piloted in administrative units like towns, counties, or states.

- To estimate impacts, economists typically look to neighboring administrative units for comparison groups and apply econometric techniques to control for observable and unobservable sources of bias.

Quick Overview

- Merged experimental data (from Cobb County) with non-experimental control group (from Fulton County).
  - Two treatments: one with significant impact, the other not.
- Applied econometric techniques to estimate treatment effects.
  - Panel Data (also, difference in difference), OLS, Matching.
- Following best practices described in the literature, and using matching as a preprocessing method, panel data can replicate the experimental estimates.

Outline

- Introduction
- Field Experiment & Data
- Literature Review
- Results
- Next Steps

Field Experiment

- Cobb County Water Management and Ferraro implemented targeted, residential information campaign as a randomized experiment.

- 3 Message Treatments:
- ~12,000 households/treatment
- ~71,000 households in control group.
  - Treatments sent via first class mail in May 2007.
  - Use two treatments for design replication.

Technical Advice (Treatment 1)

- Information on behavioral changes and technologies that can reduce water consumption.

- Example:
  - For outdoor watering: Fix leaks, water lawn in the morning.
  - For indoor watering: Take short showers rather than baths, replace old toilet and/or old faucets.
  - Information on how to detect leaks and how to get advice for more efficient water use.

Strong Social Norms (Treatment 3)

- Technical Advice + Civic Duty Language + Social Comparison:

- Example:
  - Your own total consumption June to October 2006: 52,000 gallons
  - Your neighbors' average (median) consumption June to October 2006: 35,000 gallons
  - You consumed more water than 73% of your Cobb County neighbors.

Fulton County

- In metro Atlanta with long river border with Cobb County.
- Outcome measured same (water meters).
- Similar (but not identical) water pricing. Same water sources, weather patterns, state and metro regulatory environments and other regional confounding factors during the pre- and post-experiment periods.
  - Southeastern US drought (2007)

The Good and the Bad…

- The good:
  - Counties not dramatically different. Except for treatments, Fulton and Cobb counties did not do anything different post-experiment (only what was required by law in Metro Atlanta).
  - Households did not choose their exposure to treatment except indirectly through their choice to move to the county.
- The bad:
  - Likely to be time-invariant unobservables that affect decision to move to county and water consumption.
  - Could be time-varying unobservables in water conservation efforts.
    - The key identifying assumption is that expected water consumption trend would be the same in treatment and control units in the absence of treatment.

Data

- Cobb County experimental water data (2006 – 2007)
- Fulton County control water data (2006 – 2007)
- Merged with Tax Assessor (at household level)
  - Characteristics of the dwelling, fair market value, age of the home, ownership status (renter/owner).
- Merged with 2000 US Census (at neighborhood level)
  - Socioeconomic information such as race, income, and education.

Heckman et al. (1997, 1998) & Cook et al. (2008)

- Same data sources for treatment and control (outcome measured similarly).
- Treatment and control share same economic environment.
- Selection into treatment plausibly understood and can be modeled (e.g., rich set of observable confounders).
- Baseline outcomes (levels or trends) similar.

Design Replication Studies

- Outside labor market programs:
  - Educational programs (Agodini and Dynarski, 2004; Hill et al., 2004; Wilde and Hollister, 2007)
  - Poverty reduction programs (Diaz and Handa, 2006; Handa and Maluccio, 2010)
  - Migration (McKenzie et al., 2006)
  - Elections (Arceneaux et al., 2005)

- Methods:
  - Propensity Score Matching (Agodini and Dynarski, 2004; Hill et al., 2004; Diaz and Handa, 2006; Handa and Maluccio, 2010; Wilde and Hollister, 2007)
  - Instrumental Variable (McKenzie et al., 2006, Arceneaux et al., 2005)
  - Panel Data (Smith and Todd, 2005; Heckman et al., 1997; Heckman et al., 1998)
  - Regression Discontinuity (Black, Galdo and Smith, 2007; Buddelmeyer and Skoufias, 2004; Lamadrid-Figueroa et al., 2008)

Baseline Evolution of Water Consumption

* In thousands of gallons

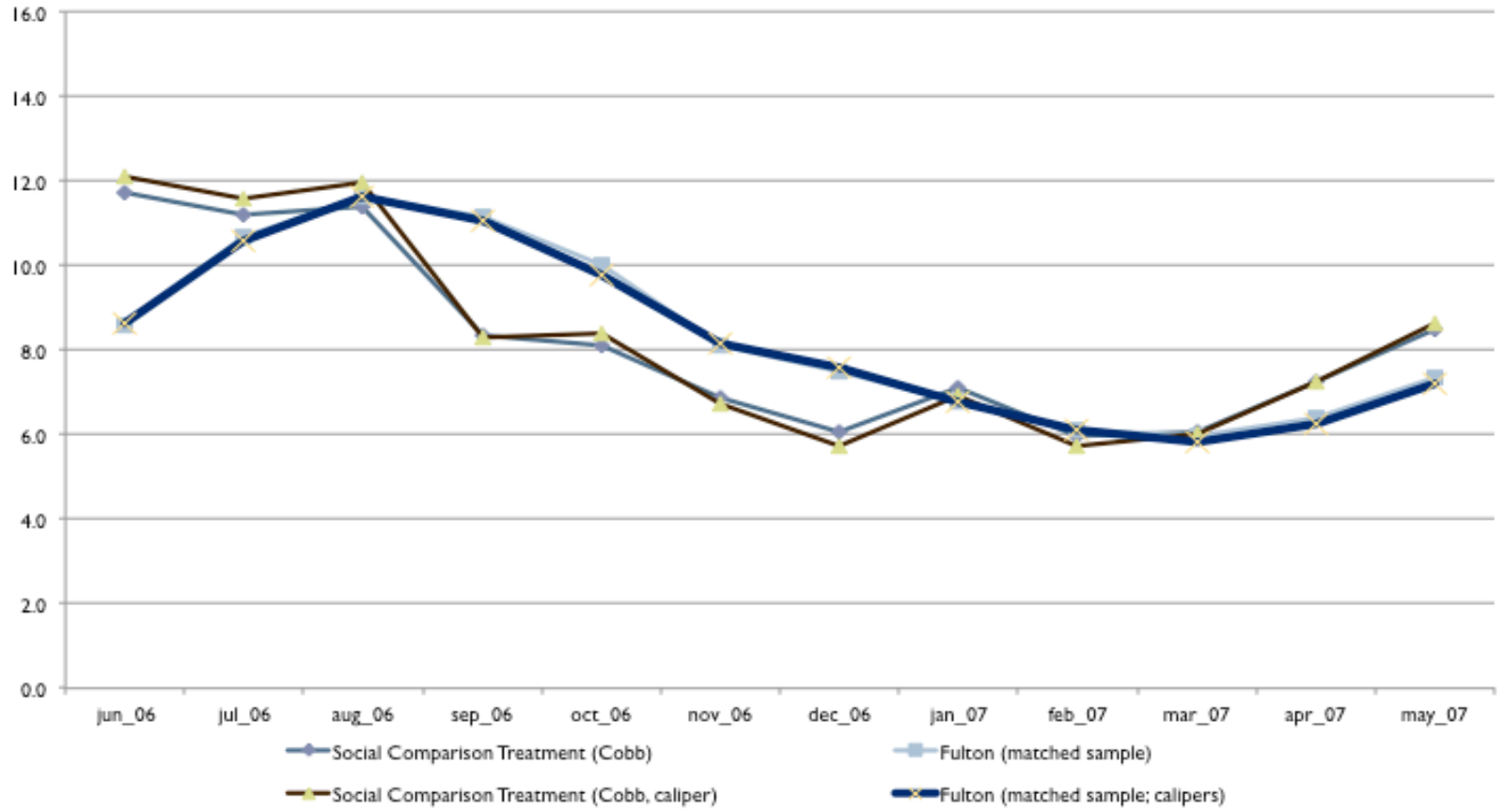
Results: Panel Data

Dependent Variable: Monthly Water Consumption (in thousands of gallons)

	Experimental Estimates	Social Comparison Treatment (Full Sample)	Technical Information Treatment (Full Sample)
Technical Information Treatment*Post Treatment	-0.006		-0.618***
	(0.054)		(-0.091)
Social Comparison Treatment*Post Treatment	-0.353***	-0.967***	
	(0.046)	(0.087)	
Observations	1394455	697316	698450
Number of id	82027	41023	41093

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Baseline Evolution of Water Consumption



* In thousands of gallons

Results: Panel Data (Matched samples on pre-treatment water trend)

Dependent Variable: Monthly Water Consumption (in thousands of gallons)

	Experimental Estimates	Social Comparison Treatment (Matched Sample)	Technical Information Treatment (Matched Sample)
Technical Information*Post Treatment	-0.006		-0.109*
	(0.054)		(-0.065)
Social Comparison*Post Treatment	-0.353***	-0.510***	
	(0.046)	(0.061)	
Observations	1394455	339626	341529
Number of id	82027	19978	20090

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Results: Panel Data (Calipers sample)

Dependent Variable: Monthly Water Consumption (in thousands of gallons)

	Experimental Estimates	Social Comparison Treatment (Caliper Sample)	Technical Information Treatment (Caliper Sample)
Technical Information*Post Treatment	-0.006 (0.054)		-0.001 (-0.070)
Social Comparison*Post Treatment	-0.353*** (0.046)	-0.345*** (0.067)	
Observations	1394455	239530	243983
Number of id	82027	14090	14352

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Results: Simple Matching Estimator

Dependent Variable: Summer 2007 (in thousands of gallons)

	(1)	(2)	(3)	(4)
	Social Comparison		**Technical Information**	
Experimental	-1.727 (0.180)		-0.129 (0.206)	
Matching	No Caliper & No Bias Adjustment	Caliper & Bias Adjustment	No Caliper & No Bias Adjustment	Caliper & Bias Adjustment
Estimate	-1.854	-1.119	-0.061	0.558
SE	0.256	0.684	0.275	0.508
T-stat	-7.257	-1.636	-0.221	1.097
P-value	0.000	0.102	0.825	0.273

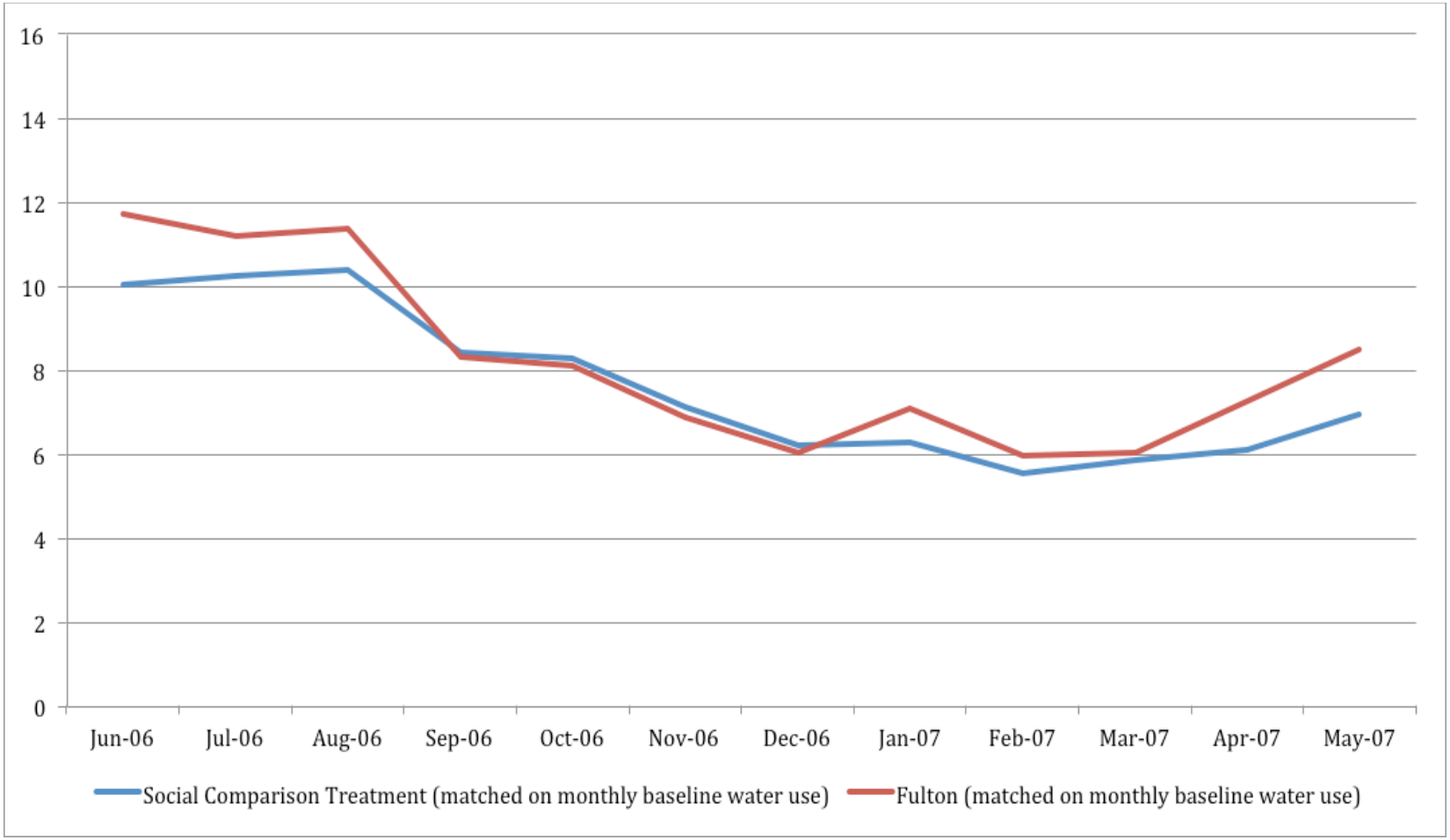
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Other Observations

- Tax and census data are critical. Cannot get good estimates in any design without them.

- Matching on monthly water use rather than seasonal can be too much of a good thing.

Baseline Evolution of Water Consumption



* In thousands of gallons

Results: OLS

Full Sample
Dependent Variable: Water Consumption Summer 2007 (in thousands of gallons)

	Experimental Estimates			Social Comparison Treatment			Technical Advice Treatment		
Social Comparison Treatment	-1.646***	-1.723***	-1.725***	-2.736***	-1.802***	-0.916			
	(0.179)	(0.180)	(0.180)	(0.470)	(0.545)	(0.598)			
Technical Advice Treatment	-0.163	-0.128	-0.127				-1.057**	0.373	1.095*
	(0.203)	(0.206)	(0.206)				(0.497)	(0.590)	(0.621)
Other Controls									
Previous water use	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household var.		Yes	Yes		Yes	Yes		Yes	Yes
Socioeconomic var.			Yes			Yes			Yes
Observations	85209	81585	81585	42795	42484	42484	42857	42543	42543
R-squared	0.618	0.626	0.626	0.295	0.296	0.297	0.298	0.299	0.300

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Next Steps

- Identifying important source of bias
- Testing the performance of tests (placebo, hidden bias)

Preliminary Conclusions

- If one follows the prescriptions of Heckman et al. (1997, 1998) and Cook et al. (2008), non-experimental design can perform as well as experimental design in our case.

- But not straightforward to know ex ante if one is following the prescriptions in many cases.

- Use matching as a preprocessing method to make these assumptions more plausible. Then, non-experimental designs perform as well as experimental designs in our case.