Social comparisons, household water use and participation in utility conservation programs: Evidence from three randomized trials

Daniel A. Brent*[1], Joseph H. Cook[2], and Skylar Olsen[3]

[1]Department of Economics, Louisiana State University
[2]Daniel J. Evans School of Public Affairs, University of Washington
[3]Zillow Real Estate Research

July 2015

Abstract

Regulation and political opposition often force water utilities to rely on non-price approaches to manage water demand. Using randomized field experiments in three different water utilities, we assess the effectiveness of social comparisons to reduce demand, and analyze their interaction with existing conservation programs. In two utilities the program decreases consumption by 5%, with significant heterogeneity across the distribution of baseline water use. We do not detect a statistically significant average treatment effect in the third utility. Social norms do not appear to crowd out existing conservation programs: treated households are more likely to participate in additional programs. Of the two utilities with significant treatment effects, higher participation rates in conservation programs account for a very small fraction of water savings (3%) in one utility and a modest fraction (9-25%) in the second. We discuss evidence that social norms may induce participation among the specific type of consumers that utilities wish to target.

Keywords: Water demand; Social norms; Behavioral economics; Water conservation; Program evaluation
JEL Classification Numbers: Q21, Q25, Q58, H76, L95.

---

*Corresponding author: Department of Economics, Louisiana State University, Baton Rouge, LA 70803; email: *dbrent@lsu.edu*. The authors are grateful to the editor and two anonymous referees for constructive feedback as well as Michael Hanemann, Hendrik Wolff, Kerry Smith, Marc Jeuland, Mark Long, and Michael Price for helpful advice. We also received valuable comments from seminar participants at the World Congress of Environmental and Resource Economics, the IWREC Annual Meeting, ASU Urban Water Demand Workshop, and the Monash Behavourial and Experimental Workshop. Additionally the authors are grateful for assistance from Peter Yolles, Ora Chaiken, Chad Haynes, Will Holleran, and Doug Flanzer from WaterSmart for sharing data and providing insights into their company and pilot utilities. Funding from the Cooperative Research Council for Water Sensitive Cities (CRC grant number 20110044) is acknowledged. Although WaterSmart Software reviewed and commented on the manuscript, the authors retained the right to publish, while maintaining confidentiality of the clients, without approval from the company.

1 Introduction

Public water utilities, particularly those in the Western U.S., face pressure to match rising water demand with diminishing or uncertain supplies. California, for example, has required that utilities reduce demand by 20% by 2020[1] and in 2013 Texas adopted a $2 billion water infrastructure fund.[2] To meet these conservation goals, utilities are increasingly looking beyond traditional demand-management measures such as education, watering restrictions, technology standards, rebates for water-efficient appliances, or raising volumetric prices (for a concise introduction to these measures, see Olmstead (2010)). Although price and mandatory restrictions may be more effective than non-pecuniary programs (Renwick & Archibald, 1998), raising water rates is often difficult due to political opposition and regulatory constraints, namely zero profit constraints. Rebates and incentive programs for converting to water-efficient appliances may not be cost-effective if most customers who use them would have bought a high-efficiency appliance anyway. Bennear *et al.* (2013) find that 47% of households who took advantage of a rebate for high-efficiency toilet were planning to buy one even without the rebate, and only 37% of the total water savings from installing the toilets could be attributed to the rebates. Outdoor watering restrictions produce water savings but at a nontrivial welfare cost: Mansur & Olmstead (2012) and Grafton & Ward (2008) find that restrictions implemented in arid regions or during extreme droughts cost households an average of over $100 per irrigation season.

One newer tool is the use of social comparisons - informing customers of their consumption relative to neighbors - to reduce water and energy use. Several studies analyzed data from Opower, a clean technology firm that provides social comparisons in energy use, find energy savings of approximately 2% with various sources of heterogeneity (Allcott, 2011; Ayres *et al.*, 2012; Costa & Kahn, 2013; Allcott & Rogers, 2014). The evidence for social comparisons in the water sector is limited to a single experiment in a single

[1]See http://www.swrcb.ca.gov/water_issues/hot_topics/20x2020/ for more information.
[2]See http://openstates.org/tx/bills/83/SJR1/ for the full text of the bill.

location. An information campaign in Cobb County, Georgia (the metropolitan Atlanta area) that included social comparisons among letters sent to consumers in the spring of 2007 to reduce water consumption during a severe drought has been analyzed in a series of studies (Ferraro *et al.*, 2011; Ferraro & Price, 2013; Ferraro & Miranda, 2013; Bernedo *et al.*, 2014; Bolsen *et al.*, 2014). We discuss results from both the Opower experiments and the Cobb County experiment in more detail shortly.

We collaborated with WaterSmart Software, a clean technology company, to analyze the effects of providing social comparisons on water use paired with information on utility-sponsored conservation programs, what WaterSmart calls a "home water report" (HWR). The company is currently working with 18 utilities, with projects in different stages of development. Projects in three cities in California randomized households into treatment and are mature enough for analysis. We use data from 289,650 meter readings and 7,361 households (50% of which were treated) to contribute to this literature in two important ways.

First, we provide evidence for whether the OPower results can be replicated in the water sector (beyond the Cobb County experiment) in three locations that vary in hydrology, demographics and ideology. Our average treatment effects imply a 5.1% decrease in water use compared to the control group in one utility, and a 4.9% decrease in the second. In the third utility, however, the point estimate for the ATE is negative but is not statistically different from zero. We discuss two reasons why treatment may have been less effective there – namely differences in email versus print reports and tariff structure – but our experimental design does not allow us to test these explanations definitively. Consistent with earlier studies, we find that treatment is most effective on high water users (Allcott, 2011; Ferraro & Price, 2013; Ferraro & Miranda, 2013). There is also little evidence of "boomerang" increases among the low users who received reports telling them they are already efficient water users. As treatment continues over time, the durability of the effect is similar to the pattern seen in existing OPower studies in one utility: they increase over the first few reports received and then stabilize. Results in a second utility are similar except that they seem to disappear by the 18th month of receiving bimonthly reports (i.e. after the 9th report). Treatment does not have a statistically-significant effect in any period in the third utility.

Second, we examine the interaction between the social comparison treatment and existing utility conservation programs such as free home water audits and rebates for efficient toilets or irrigation controllers. Research into green electricity programs finds that traditional economic incentives increase participation rates (Jacobsen *et al.*, 2013), and that there is significant heterogeneity in participation due to income and environmental preferences (Kotchen & Moore, 2007). Understanding how social comparisons interact with existing conservation programs depends on the both the motivation behind the response to comparisons and the specific features of the conservation program. If comparisons raise the moral cost of water consumption then we may expect people to take what Attari (2014) calls "intent-oriented" actions. These are behavioral changes that directly connect a person with the conservation activity, such as turning off the water when brushing one's teeth. However, as Attari (2014) shows, these actions are not the most effective ways to reduce water use; "impact-oriented" actions that generate the largest water savings are primarily investments in water efficiency. If social comparisons induce consumers to participate in rebate programs that improve water efficiency there may be larger and more durable changes than if consumers merely make behavioral changes.

Although a number of studies have examined participation in utility programs, we are aware of only one study that has examined the link between social comparisons and program participation.[3] Using cross-sectional data on participation among OPower households, Allcott & Rogers (2014) find that receiving a home energy report increased participation from 44 households in 1000 to 48 households in 1000, a 0.4 percentage point increase.[4] We find dramatically higher estimates: in the two utilities where we see signif-

[3]Bennear *et al.* (2013) find low causal water savings from rebates in water utilities, and Jacobsen *et al.* (2012) show that certain households increase energy use after participating in green electricity programs.

[4]In another setting with privately-provided public goods, Shang & Croson (2009) finds that providing

icant average treatment effects, receiving the home water report increases participation from 38 in 1000 to 109 in 1000, an increase of 6 percentage points. Although we cannot rule out a simple advertising effect, administrative data show that over 95% of households had already received information on these programs. Additionally, an analysis of treated vs. untreated program participants' water use before and after they sign up for programs suggests that, for some programs, the social comparisons appear to increase salience (or motivate action) among high-water-use households, precisely the type of consumers that the utility wants to target. Incorporating this information into our preferred difference-in-difference model, we find that increased participation in these programs accounts for 9-25% of the observed treatment effect in one utility and 3% in a second utility. The interactions between the HWRs and existing conservation programs place social norms within a utility's broader water conservation portfolio.

Existing studies on social comparisons in household energy and water use

To help frame our research we briefly review the existing work on the use of social comparisons in energy and water. Opower, a software company, began sending "home energy reports" in 2008 that included information on a customer's relative consumption as well as technical information on reducing energy use. OPower now has over 90 utility clients in six countries, and their programs have been studied extensively. Using data from two of the earliest pilots, Ayres *et al.* (2012) found that Opower's "home energy reports" reduced energy demand by an average of 2%, and subsequent studies from a larger group of OPower experiments have found commensurate average treatment effects (Allcott, 2011; Costa & Kahn, 2013).[5] There is considerable heterogeneity in the treatment effect across distribution of average pre-treatment (baseline) electricity use. As expected, the largest baseline users contribute the largest percentage savings (Ayres *et al.*, 2012; Allcott,

information on the charitable donations of others increases donations in both the current campaign as well as future campaigns.

[5]We focus here primarily on analysis of the OPower experiments because of their size, clean identification, and replication. A number of studies in social psychology predated OPower and established the "proof of concept" for social comparisons in energy use (Hutton & McNeill, 1981; Luyben, 1982; Winett *et al.*, 1982; Hirst & Grady, 1982; Siero *et al.*, 1996; Staats *et al.*, 2004; Kurz *et al.*, 2005; Schultz *et al.*, 2007; Goldstein *et al.*, 2008).

2011), but there is little evidence of a "boomerang effect" where below-average users increase their use after receiving the home energy report. This may be because the reports include a normative injunction via emoticons (i.e. a smiley face) or labels that below-average users are "doing great"; in a smaller experiment pre-dating Opower, Schultz *et al.* (2007) found that providing emoticons eliminated an observed boomerang effect among below-average households.

The experiment using social comparisons for water conservation in Cobb County was large, with 71,000 control households and roughly 35,000 households receiving one of three treatments: 1) a "tip sheet" with technical information on how a household could conserve water, 2) the technical information plus a letter from the utility encouraging customers to conserve water and "do their part" (what the authors label a "weak" social norm), and 3) technical information plus a comparison of the households' June to October 2006 water use with their "neighbor's average (median) consumption" during the same period ("strong" social norm). The last treatment also included a statement: "You used more water than XX% of your Cobb County neighbors". The social norm treatment mailings were sent only once (in May 2007), although the research team has tracked water use for control and treated households through summer 2013 (Bernedo *et al.*, 2014). Ferraro *et al.* (2011) find that the "weak" and "strong" social norm treatments reduced consumption relative to the control group by 2.7% and 4.8% during the summer of 2007, the latter is considerably larger than the effects observed in the energy sector. The treatment that only provided technical information did not have a statistically-significant impact (Ferraro *et al.*, 2011), though subsequent work showed a statistically, but not economically, significant effects in some models (Ferraro & Price, 2013). Both Ferraro *et al.* (2011) and Ferraro & Price (2013) find that the effects quickly wane in the months after treatment, but then remained stable. Bernedo *et al.* (2014) find that treatment effects are not detectable in the one-quarter of treated homes where the person living there in 2007 has moved, but that after excluding these houses the treatment effect of receiving one comparison in the summer of 2007 are still statistically-significant in the summer of 2013 and imply a reduction of 1.4%. As in the energy sector, treatment effects were larger in households with higher baseline water use (Ferraro & Price, 2013), as well as in households that were wealthier and owner-occupied (Ferraro & Miranda, 2013).

The next section describes the WaterSmart "Home Water Report" treatment, the three participating utilities, our data sources, and the experimental design. Section 3 describes the estimation strategy and presents empirical results on average treatment effects as well as their heterogeneity and durability. Section 4 explores the connection between the Home Water Report treatment and participation in a utility's existing conservation programs. The final section discusses reasons why the social comparison may not have been effective in the third utility, calculates the cost-effectiveness of HWRs and approximate increases in water tariffs needed to achieve the same water use reduction, and discusses the implications of interactions between social comparisons and other water conservation programs.

2 Background & Data

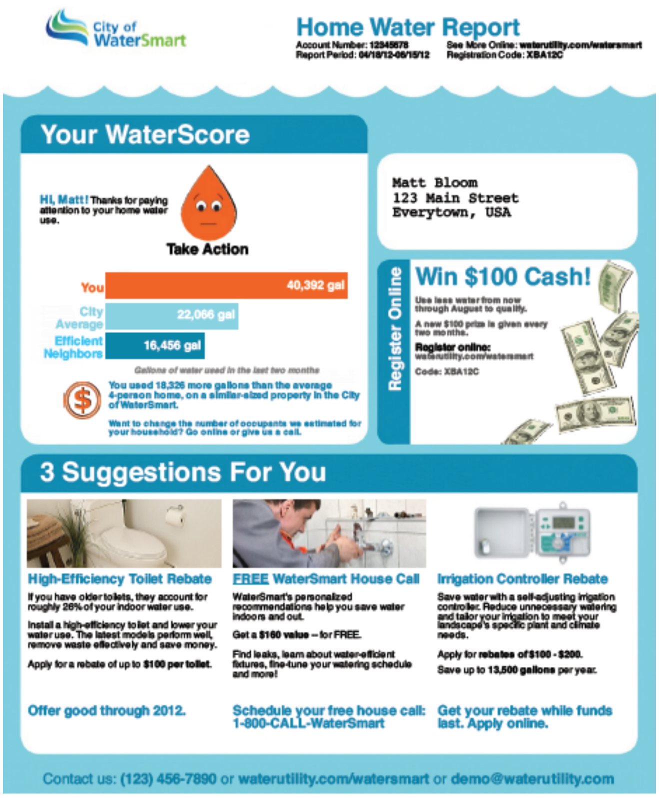
2.1 WaterSmart Software

WaterSmart Software is a clean technology company that contracts with water utilities to help them manage demand. In addition to assistance with analyzing and interpreting meter reading data, WaterSmart primarily focuses on helping utilities reduce water consumption by providing consumers with information through customized Home Water Reports (HWRs) (Figure 1) and an online customer account portal. Since customers opt-in to view their online account, we focus here on the treatment effect for households receiving a HWR.

The one-page Home Water Report as tested in our three locations has three components. The main component (in the upper left of the figure) is a social comparison. WaterSmart estimated the household's total water consumption over the prior two months from utility billing records and compared that to the consumption of "average neighbors" and "efficient neighbors". "Neighbors" are defined as households that have the same number of occupants and similar irrigable area across the utility, such that the general water requirements within a peer group are comparable. “Efficient neighbors” were peers with consumption in the bottom 20%. Households with consumption above the median of their peer group receive a “Take Action” normative message (shown in Figure 1), those with consumption between the median and 20^{th} percentile receive a “Good” message, and those below the 20^{th} percentile receive a “Wise” message.[6] Utility Cs normative message also depends on their water use relative to their water allocation from their water rate, which we describe in more detail below.

The second component (across the bottom of Figure 1) is a list of three personalized recommendations for strategies to save water. Two types of recommendations involved providing technical advice, either through a community class or a free home water audit where a professional would help the household identify leaks and ways to conserve water. Other recommendations highlighted rebates available from the utility for installing higher-efficiency toilets or clothes washers; converting lawns to xeriscaping; installing higher-efficiency sprinkler systems; or purchasing an automatic irrigation controller. Based on data available from the utility (described more below), and on results from a baseline household survey with limited responses, WaterSmart personalized these recommendations to the greatest extent

Figure 1: Home Water Report



Notes: Home Water Reports have WaterScores of “Wise”, “Good”, or “Take Action”. The bottom panels contain suggestions that are customized based on household data. The right “Win $100” offers different incentives or messages

[6] Home Water Reports showing the latter two categories are provided in the Appendix.

possible. For example, if a household had no outdoor area it was not given a recommendation regarding irrigation, and if the household reported having a low-flow toilet in the survey it was not shown an incentive for toilets. Our context differs in this way from the "strong" social norm treatment in the Atlanta experiments where households were given generic "tips" in addition to the comparison; our households received more tailored information. Unfortunately we do not observe exactly what each household saw; our data does not contain the personalized tips at the level of individual households.

The third component (in the upper right of Figure 1) offered households an incentive to sign up for an online account for more detailed data on water use and additional water conservation tips. A screenshot of the online web portal is provided in the Appendix. In Utility A and Utility B the HWRs were mailed to customers in print form, while in Utility C they were delivered via email, a potentially important modal difference that we can speculate about, but not reliably test in our data. Dolan & Metcalfe (2013) report that print copies of social norms for electricity conservation are more effective than digital versions delivered via email. The content, however, does not differ substantially between email and print HWRs.

2.2 Utilities

We use data from randomized trials of HWRs in three utilities in California that vary in hydrological and socioeconomic conditions (Table 1). We have been asked to conceal their identities and will refer to them as Utility A, B and C. Utility A is relatively small, serving approximately 10,000 customers in a town north of the Bay Area. Utility B serves over 1 million people in the Bay Area. Due the size of its service area there is also more variation in both climatic zones and demographic characteristics than the other two utilities. Utility C serves approximately 300,000 customers in a drier and less temperate part of California. This is apparent in Table 1 by observing Utility C's lower rainfall and higher average temperature. Because of the variation in climate, we expect that seasonal outdoor water use such as landscaping and refilling pools will also vary across the utilities, although we do not observe outdoor water use, only total water use. Median household income is higher in Utility B and Utility C than the statewide average of $61,400.[7]

Table 1: Summary Statistics By Pilot

Pilot	Water Use	Temp (F)	Rain (in)	Income	House Value	Ideology Index
Utility A	204	70	2.1	60,056	360,332	68
Utility B	291	67	1.6	99,522	720,944	57
Utility C	360	72	0.8	123,240	973,335	34

Notes: Water Use is average pre-treatment water consumption in gallons per day; Temp is the average monthly maximum temperature in degrees Fahrenheit; Rain is the monthly average in inches; Income is the average median income at the census block group from 2012 estimates from the US Census's American Community Survey; House Value comes from Zillow data; Ideology Index (described in the Appendix) ranges from 0 to 100, with higher values indicating more pro-environment votes at the census-block level.

The three utilities operated a number of conservation programs in addition to the WaterSmart program during the study period (Table 2). All the utilities offered rebate programs, as well as other programs tailored to the community's needs. Utility C offered programs to address outdoor water use, while Utility A focused on community engagement. From administrative data we observe whether and when households participate in these programs for all three utilities.

Table 2: Utility Conservation Programs

	Utility A	**Utility B**	**Utility C**
Rebates			
Toilets	X	X	X
Clothes Washer	X	X	X
Lawn Conversion	X	X	X
Sprinklers			X
Irrigation Controller		X	X
Technical Advice			
Home Water Audits	X	X	
Community Classes	X		

Notes: Only programs that are available from the utility are included. Some programs are administered through regional bodies, and additional resources are available from state and regional agencies.

Because the HWR may signal to households that their neighbors are able to use water

[7] Median household income 2008-2012, http://quickfacts.census.gov/qfd/states/06000.html

more efficiently and thus lower their water bills (Ayres *et al.*, 2012; Ferraro & Miranda, 2013), the structure of water rates and any rate changes during treatment may affect our results. Utility A has the simplest tariff structure, with a fixed cost and single volumetric charge. (Appendix Figure A.8 provides a visual description of rate structures in the three utilities). Utility B has an increasing block rate structure with three tiers. The rate structure is relatively flat, with a 51% increase in marginal price from the lowest to the highest tier. Utility C has a somewhat unique, "budget-based" tariff structure. At the *end* of each billing cycle each household is assigned a water budget or "allocation" based on occupancy, irrigable area and weather conditions during the billing period. The tiers of Utility C's increasing block tariff are related to the percentage of the household's actual water use compared to its allocation. A household that used 25% more water than its allocation, for example, would fall in the third rate tier and pay $2.76 per ccf (1 ccf = 748 gallons) in 2012, in addition to a $9.85 monthly fixed charge. The rate for the highest tier, reserved for consumption that is more than 200% above allocation, is over 600% higher than the base rate. The allocation is also linked to the normative message in the HWR. A household that consumes above their allocation receives the "Take Action" message on their HWR. Households receive the "Good" message if they are below their allocation but above the median for their peer group, and the "Wise" message if consumption is both below the median and below their allocation.

The rate structure in Utility A did not change during the treatment period, though it did in the other two utilities. Rates for all three tiers increased by 10% in Utility B during the treatment period, and Utility C's rates minimally increased in the 2nd, 3rd, and 4th tier by 2-4 %. Although we perform a robustness check for Utility B we do not explicitly incorporate water rates into the analysis presented here because the use of time fixed effects controls for all time-varying price effects at the utility level.[8] Time fixed effects do not account for the impact of changing prices if households move *across* pricing

[8]The coefficient on an interaction of the treatment effect with an indicator for the period after the price increase is roughly 1.5 percentage points and statistically insignificant. The main treatment effect declines by less than 0.5 percentage points and is still significant at the 1% level.

tiers, although none of the tariff revisions altered the width of consumption blocks, so the revisions would be unlikely to alter any "bunching" observed in the tariff structure. Although even small changes in water rates can impact water demand (Klaiber *et al.*, 2014), randomization should minimize their effect on the main treatment effect of interest.

2.3 Data

The utilities (via WaterSmart) provided water metering records at the household level. Meters were read every two months in Utility A and Utility B and every month in Utility C. We remove observations of zero consumption and outliers that are great than eight times the interquartile range that are likely data errors. The utilities also provided information on the structural features of the house such as lot size, square footage, number of bedrooms and bathrooms, year built, and irrigable area size.[9]

We use daily temperature and precipitation data from the Global Historical Climatology Network-Daily (GHCN-D) dataset maintained by NOAA's National Climatic Data Center. We match households to the nearest weather station with available temperature and precipitation data and aggregate daily values within the irregular read periods of household water read observations. This allows us to capture the exact temporal weather conditions for each reading period that are specific to individual households, as well as account for microclimates within the larger service areas. We rely on cooling degree-days and the number of days with precipitation as weather controls, both of which are divided by the total number of days in the billing period.[10] Our results are robust to other weather controls (results available from the authors on request), including evapotranspiration.[11]

Rather than using coarser-scale household income data from the US Census, we proxy income using block level median housing values from Zillow, an on-line real estate com-

---

[9]Irrigable area is either observed through survey responses or estimated by the size of outdoor area that is determined from data on lot size and the footprint of the home.

[10]Cooling degree-days measure the cumulative sum of average daily temperature minus 65 degrees Fahrenheit.

[11]Evapotranspiration (ET) data measures the consumptive needs of turf grass and is an important determinant of residential water demand (Olmstead *et al.*, 2007; Mansur & Olmstead, 2012), but we have a non-trivial number of missing observations using spatially-matched zip-code-level daily ET data from the California Irrigation Management Information System (CIMIS).

pany with a database of millions of homes and sales transactions across the US. We also construct an index at the census-block level to proxy for a household's environmental ideology, which we label the "Green Ideology Index (GII)", based on six votes during the 2008 and 2010 elections. We provide more detail on the index in the Appendix.

2.4 Experimental Design

WaterSmart identified a random sample of customers in all three utilities; in (smaller) Utility A the sample comprised 95% of all households. In Utility A they excluded households with a zero meter read for the current period. In the other two utilities, they excluded households with less than two years of historical data, households with zero reads, non-resident accounts, and extreme outliers (3 standard deviations over their historical mean usage). Remaining households in the sample were randomly selected to receive Home Water Reports by WaterSmart.[12] Sample sizes, start dates, and the number of households treated in these pilots are shown in Table 3. Each observation is expressed as the average gallons of water consumed per day during the past meter reading period, calculated by dividing total gallons by the number of days in the period. Households received a HWR in the middle of each meter reading period. This is because the HWR requires consumption data from the previous period to generate the metric on relative consumption. A portion of consumption in the first treated meter read period is therefore untreated, attenuating the treatment effect. Rather than include these half-treated periods, we focus our analysis on meter reads that occurred completely after receiving the first HWR.[13] Treated households in Utility A received 8 HWRs in total during the period analyzed; households in Utility B and Utility C received 9 and 13 HWRs, respectively.

The control and treatment groups are well balanced on observables (Table 4). Most importantly, they are balanced on "baseline" water use, which we define as household water use averaged over the entire pre-treatment period. There are small differences

[12] In Utility A, WaterSmart sorted households in the sample on current period consumption and selected every other household for treatment. We believe the exclusion criteria were applied after randomization into treatment in Utility A, but before randomization in Utilities B and C. In Utility B, the overall sample was stratified based on three geographic zones and seven parcel size classifications

[13] Results including the first period are similar and are available from the authors on request.

Table 3: Sample Sizes

Pilot	Start Date	End Date	N: Obs	N: Post-Treat	HHs	Treated HHs
Utility A	2011-09-20	2013-01-01	39,975	13,882	1,889	992
Utility B	2012-06-28	2013-08-23	98,929	23,692	3,092	1,545
Utility C	2012-07-26	on going	150,746	28,620	2,380	1,180

between control and treatment in Utility C on our ideology index and the age of the housing stock and the treatment group in Utility C has 0.14 fewer occupants than the control group. Although these differences would bias our average treatment effect when using simple difference in means, a difference-in-difference specification controls for these with household fixed effects.

Table 4: Balance of Observables Across Treatment

(a) Utility A

	Control	N_C	Treatment	N_T	Difference	p-value
Baseline Water	204.9	897	200.6	992	4.31	0.510
Assess Value	358,750	851	358,597	941	154	0.978
Ideology	67.5	654	68.0	712	-0.43	0.352
Occupants	2.65	897	2.61	992	0.033	0.559
Lot Size	9,165	821	8,188	903	977	0.212
Year Built	1983.8	895	1983.9	992	-0.079	0.916
Single Family Home	0.70	897	0.68	992	0.015	0.495

(b) Utility B

	Control	N_C	Treatment	N_T	Difference	p-value
Baseline Water	279.5	1,547	278.8	1,545	0.72	0.928
Assess Value	714,723	1,508	713,211	1,502	1,511	0.923
Ideology	57.1	979	57.6	978	-0.53	0.488
Occupants	2.93	1,547	3.01	1,545	-0.082	0.065
Lot Size	8,703	1,546	8,177	1,542	526	0.110
Year Built	1951.7	1,547	1951.0	1,545	0.72	0.481
Single Family Home	0.98	1,547	0.97	1,545	0.0084	0.132

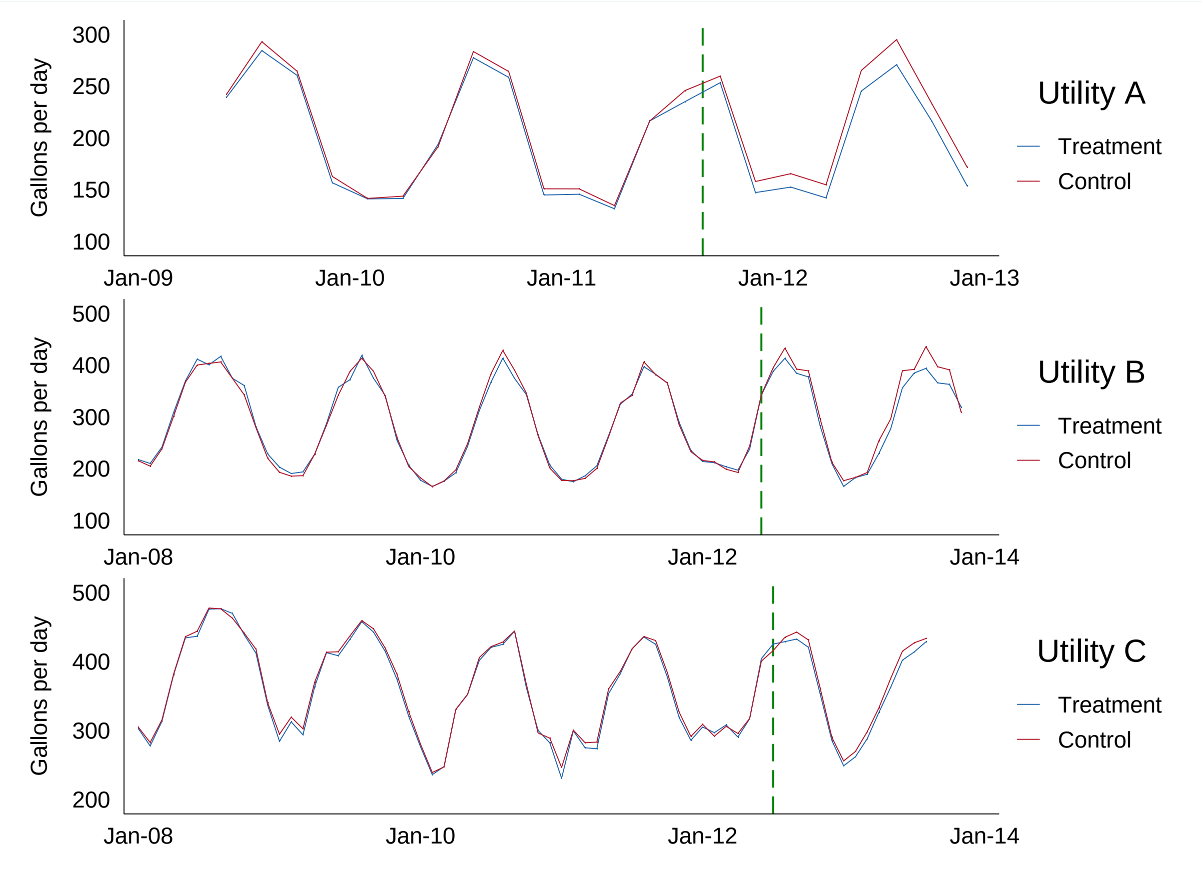
(c) Utility C

	Control	N_C	Treatment	N_T	Difference	p-value
Baseline Water	359.1	1,200	355.4	1,180	3.72	0.632
Assess Value	974,357	1,068	965,938	1,061	8,419	0.751
Ideology	33.4	921	34.3	945	-0.89	0.061
Occupants	4.03	1,200	3.88	1,180	0.14	0.000
Lot Size	5,620	1,053	5,432	1,041	188	0.148
Year Built	1990.0	1,200	1991.0	1,180	-1.01	0.028
Single Family Home	0.89	1,200	0.89	1,180	-0.0074	0.566

Notes: p-value refers to a two-sided t-test. Assess Value is the most recent property assessment, in dollars; Ideology (described in the text) ranges from 0 to 100, with higher numbers indicating more pro-environment votes at the census-block level; Lot Size is measured in square feet.

Figure 2 plots the time series of average water use (in gallons per day) across the treatment and control group for each of the utilities. The importance of outdoor water use is clear in all three utilities because of the strong seasonality of demand. Prior to the intervention the treatment and control groups have very similar trends over time, which we feel is sufficient support for the common trends assumption in a difference-in-difference model. After the intervention (dashed vertical line), water use in the treatment group begins to diverge visibly in Utilities A and B but not in Utility C. In the next section we use a regression framework to identify the main average treatment effects, heterogeneity in effects by baseline water consumption, and the durability of the effect of receiving the HWRs.

Figure 2: Average water use (gallons per day) during study period, by treatment group and utility



Notes: The vertical dashed line indicates the start of the program for each pilot.

3 Identification of Treatment Effects

3.1 Average Treatment Effects

We estimate three specifications for the average treatment effect (ATE):

$$w_{it} = \alpha + \gamma T_i + \xi_{it};\ \forall\ P_t = 1 \tag{1.1}$$

$$w_{it} = \alpha + \gamma T_i + \beta X'_{it} + \tau_t + \xi_{it};\ \forall\ P_t = 1 \tag{1.2}$$

$$w_{it} = \alpha_i + \gamma T_i \times P_t + \beta X'_{it} + \tau_t + \xi_{it};\ \forall\ P_t = \{0, 1\} \tag{1.3}$$

The dependent variable w_{it} is average daily water consumption in gallons for household i at time t divided by average control group consumption in the treatment period and multiplied by 100. This is the specification used in Allcott (2011) and has the advantage of maintaining the interpretation of the treatment effect as a percentage change in water use without dampening the impact of large users who are expected to experience the largest reductions in water use. T_i is an indicator variable for the treatment group, P_t is an indicator for the treatment period, X_{it} is a vector of control variables, τ_t is a metering period-by-year (time) fixed effect, and ξ_{it} is the idiosyncratic error term. Since households were randomized into treatment a simple t-test for difference in average, post-treatment water use between the treatment group and the control group provides a valid statistical tool for the average treatment effect (equation 1.1, and column 1 in Table 5). Equation 1.2 augments the model with covariates to improve the precision of the ATE and correct for any remaining differences across treatment groups after randomization. We present results for three specifications of X'_{it}: baseline consumption, again defined as each household's water use averaged over the pre-treatment period (column 2 of Table 5); "seasonal" baseline consumption, defined as household water use averaged over the summer, winter, and spring/fall shoulder seasons (column 3); and "seasonal" consumption plus structural features of the home (column 4). All three of these specifications include cooling degree days, the number of days with precipitation, and time fixed effects (τ_t) as controls.

Equation 1.3 (column 5 in Table 5) is a difference-in-difference (DiD) model that includes household fixed effects, time fixed effects, and weather variables.[14] The time fixed effects capture most of the effect of seasonality while the weather variables, matched using daily weather data to the precise metering period for each household, represent the deviation from average weather conditions for a specific household in a given period. Huber-White standard errors (Woolridge, 2002) are clustered at the household level to account for serial correlation within a household over time.

The average treatment effect is negative and highly statistically-significant in all specifications for both Utilities A and B; receiving a HWRs reduces household water consumption by 5.1-7.7% in Utility A and 4.9-5.2% in Utility B.[15] Though the point estimates are negative in Utility C, the magnitudes are smaller and not statistically significant in any of the specifications.[16]

Most of the efficiency gains in estimators, relative to the difference in means, are achieved by including baseline water consumption, which decreases the standard errors by more than 50% in all three utilities. The standard errors are smallest in Utility C, partly due to monthly as opposed to bimonthly metering periods, indicating that the failure to reject the null hypothesis is not primarily due to statistical noise.

Our preferred model is the difference-in-difference specification (column 5) because it allows us to flexibly control for household level unobservables with fixed effects and because we feel randomization (and Figure 2) provides sufficient support for the common trends assumption. Allcott & Rogers (2014) restrict the sample to post-treatment data and include seasonal pre-treatment consumption covariates (corresponding to columns (3) and (4) Table 5) to more flexibly control for water use patterns than a single house-

[14]The normal post-treatment indicator P_t is not included because it is captured by the reading period fixed effects.

[15]Utility B has the largest treatment effect in gallons per day, saving approximately 16 gallons per day for the specification in column 5; the ATE corresponds to 10 gallons per day in Utility A.

[16]We run a pooled regression with pilot-level interactions for all variables to test whether the differences in average treatment effects among the pilots are statistically-significant. An F-test rejects the null for a constant ATE across pilots at the 1% level. In grouped F-tests between each pair of pilots we cannot reject the null for equal ATEs for any of the pilot groups.

Table 5: Specifications for the Average Treatment Effect

(a) Utility A

	(1)	(2)	(3)	(4)	(5)
Treatment Effect	-7.70**	-5.34***	-5.22***	-5.42***	-5.11***
	(3.31)	(1.50)	(1.52)	(1.49)	(1.53)
Baseline Consumption	No	Yes	No	No	No
Baseline Seasonal Consumption	No	No	Yes	Yes	No
Household Controls	No	No	No	Yes	No
DiD	No	No	No	No	Yes
Household FEs	No	No	No	No	Yes
Year-Period FEs	No	Yes	Yes	Yes	No
R^2	0.002	0.568	0.570	0.579	0.082
Households	1,825	1,825	1,818	1,727	1,889
Observations	12,034	12,034	11,998	11,418	38,099

(b) Utility B

	(1)	(2)	(3)	(4)	(5)
Treatment Effect	-5.14*	-5.23***	-4.93***	-5.08***	-4.90***
	(3.10)	(1.32)	(1.23)	(1.24)	(1.33)
Baseline Consumption	No	Yes	No	No	No
Baseline Seasonal Consumption	No	No	Yes	Yes	No
Household Controls	No	No	No	Yes	No
DiD	No	No	No	No	Yes
Household FEs	No	No	No	No	Yes
Year-Period FEs	No	Yes	Yes	Yes	Yes
R^2	0.001	0.631	0.643	0.645	0.071
Households	2,958	2,668	2,612	2,538	3,091
Observations	20,134	18,141	17,776	17,292	85,217

(c) Utility C

	(1)	(2)	(3)	(4)	(5)
Treatment Effect	-2.25	-1.25	-1.49	-1.04	-1.33
	(2.48)	(0.98)	(0.97)	(1.01)	(0.98)
Baseline Consumption	No	Yes	No	No	No
Baseline Seasonal Consumption	No	No	Yes	Yes	No
Household Controls	No	No	No	Yes	No
DiD	No	No	No	No	Yes
Household FEs	No	No	No	No	Yes
Year-Period FEs	No	Yes	Yes	No	Yes
R^2	0.000	0.645	0.649	0.651	0.081
Households	2,300	2,299	2,299	2,045	2,379
Observations	26,533	26,530	26,530	23,684	148,517

Notes: The dependent variable is average daily water consumption in gallons normalized by average consumption in the control group during the treatment period. Columns (1) - (4) use only post-treatment data, whereas column (5) uses observations both pre and post-treatment. Robust Hubert-White standard are clustered at the household level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

hold fixed effect in a DiD model. Although this specification produces a higher powered estimate in their data, in our setting incorporating seasonal consumption only produces efficiency gains for Utility B; Utility A and C have similar standard errors in columns (2) - (5). Furthermore, when analyzing the interaction between the treatment and participation in existing utility programs below we find it useful to examine water use prior to treatment.

3.2 Treatment Effect Heterogeneity

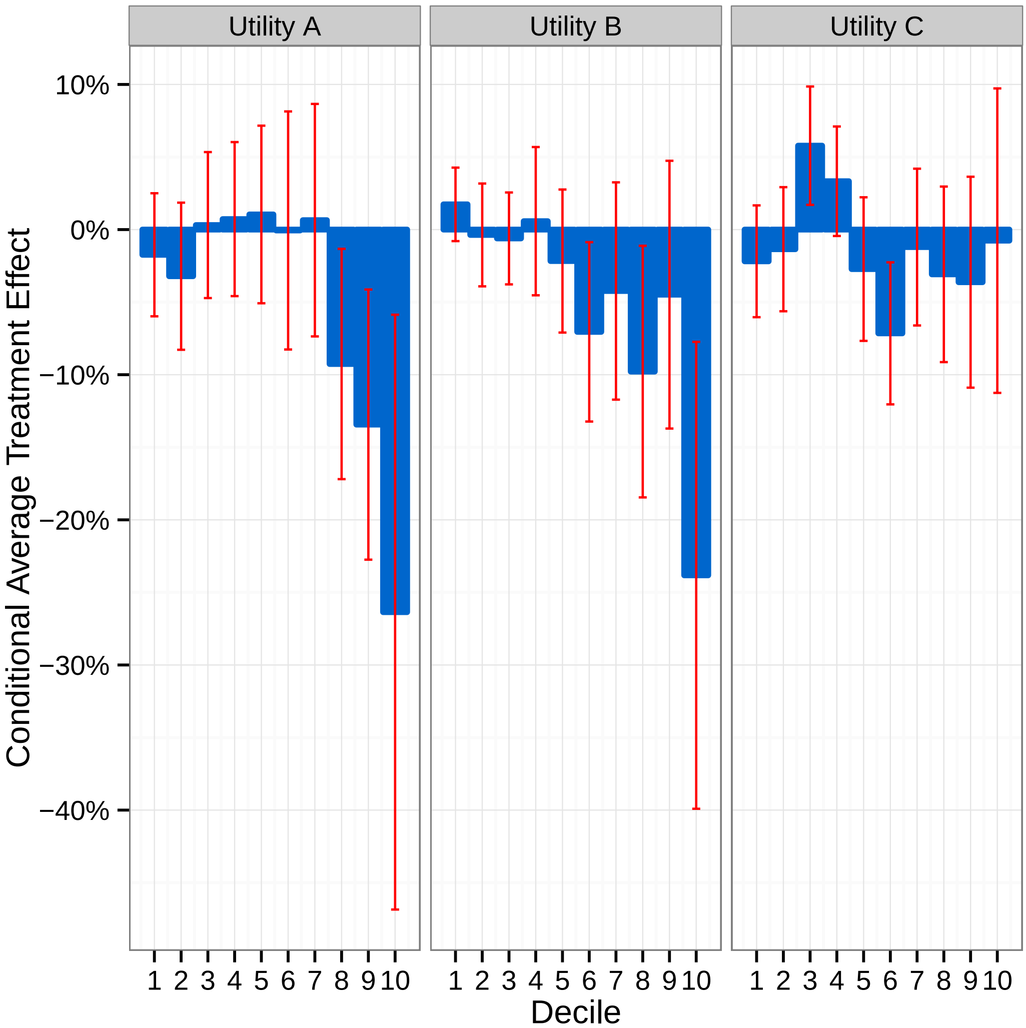
We examine heterogeneity in treatment responses by interacting the treatment effect with deciles of baseline water consumption for each utility. These regressions interact the treatment effect with subpopulations of interest based on specific variables and therefore are interpreted as conditional average treatment effects (CATE). In this specification the treatment effect for the 10^{th} decile of baseline consumption focuses on households in the sample that populate the high end of the water use distribution. Baseline deciles are calculated by taking the average pre-treatment household usage across each utility and ordering it into 10 equally-sized groups. We perform a similar technique to create deciles of the ideology index and housing values for each utility. The CATEs, unlike the ATEs, cannot be interpreted as a purely experimental result because treatment was not randomized within each of the subgroups that we use in the analysis. However, all of our interaction variables are still orthogonal to treatment due to randomized assignment.[17] The regressions include all the same variables as presented in equation 1.3 in addition to interactions between the treatment effect and deciles of water, ideology, and housing values. We also include an interaction of the deciles with a post-treatment dummy to control for time varying factors during the treatment period that are specific to each subgroup. The formal specification is available in the Appendix.

In Utilities A and B the CATEs are statistically significant in the high deciles of baseline consumption (Figure 3), which is consistent with evidence from both electricity and water: high users are the most responsive to social norms (Allcott, 2011; Ferraro & Price, 2013; Ferraro & Miranda, 2013). Utility C has statistically-significant coefficients

[17] All the three utilities are reasonably balanced on water use within each baseline consumption decile as seen in Table A.2 in the Appendix.

for the sixth decile, but not for the highest deciles. There is not a clear and consistent pattern across the three utilities in heterogeneity in the treatment effect by housing values or the average pro-environment voting record of their census block (see Appendix Figures A.7 and A.8 for these results). One important caveat in interpreting the CATE results is that the normalization was not performed over each of the baseline consumption deciles. Rather, we maintain the specification of the dependent variable as defined above, and present results using unnormalized gallons per day in the Appendix.

Figure 3: Heterogeneity: Baseline Water Use



Notes: The graphs are created from regressions of normalized daily average water consumption on interactions of the treatment effect with deciles of baseline water consumption using the DiD model. Interactions of deciles of baseline consumption and a post-treatment indicator are included as additional controls. The blue vertical bars are the point estimates and the red error bars represent 95% confidence intervals based on cluster robust standard errors. A separate regression is estimated for each utility.

3.3 Durability

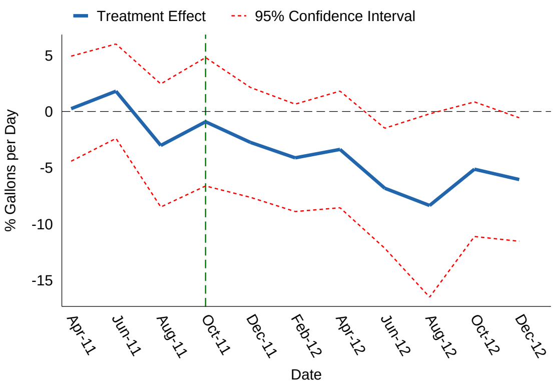
We next interact the treatment effect with dummies for each treatment period to examine the durability of the effect of receiving a HWR while the HWRs are still being received.[18] These temporal dynamics depend on two channels: the seasonality of water use and the permanence or impermanence of conservation behaviors discussed in the introduction. Since we do not observe treatment over multiple years we cannot separately identify these two channels.

The results from interactions of the treatment effect with individual reading periods for each pilot are shown graphically in Figure 4. The left vertical axis represents the coefficient estimate as percent change in normalized water consumption with 95% confidence intervals. We also include three periods prior to treatment (dashed vertical line) to test for balance prior to treatment; as expected the ATE is not significant prior to treatment. The pattern of treatment effects for Utility A shows that consumers gradually take water conservation actions after continued exposure to the treatment. The largest effect is during the summer months when most outdoor water use occurs, but there are not statistically significant differences in treatment effects during summer months relative to winter months. The pattern in Utility B is very similar except that the point estimate of the treatment effect completely attenuates in the final period in our data (December 2013). Utility B also shows an increase in the magnitude of the effect during the summer months, as the statistically significant periods correspond to months with outdoor water use. The treatment effect in Utility C is not statistically different from zero in any of the treatment periods.

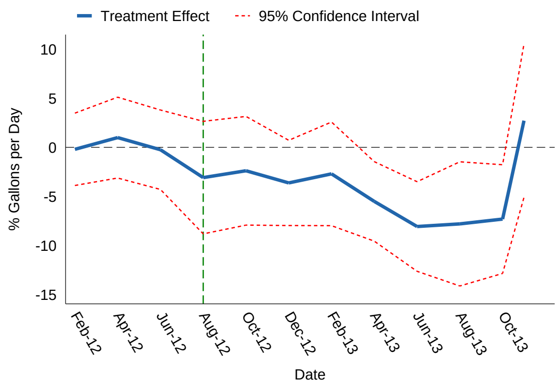
[18]Because the control group in our pilots also began receiving the HWR treatment at the end of the experiment, our research design allows us to examine durability but not persistence (as defined by Allcott & Rogers (2014)) of the reports after they are no longer received.

Figure 4: Durability of Treatment Effects - Percentage

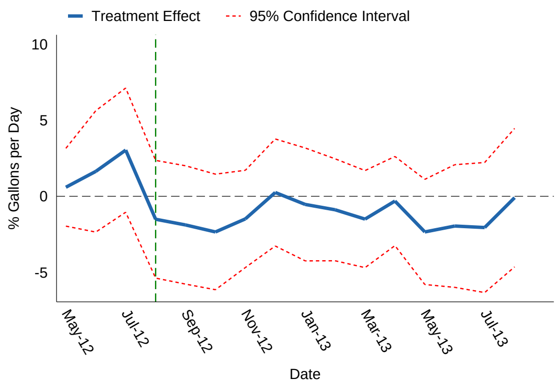
Utility A



Utility B



Utility C



Notes: The solid line is comprised of point estimates from a regression on an individual treatment periods. The dashed lines represents the 95% confidence interval constructed from robust standard errors clustered at the household level. The vertical dashed line indicates the start of the treatment. All regressions contain household and year-period fixed effects as well as weather controls.

4 Interactions with Existing Conservation Programs

Social comparisons may impact existing conservation programs through two main channels, first by increasing program participation rates and second by attracting "better" customers to the programs. Throughout this section we will continue to refer to the provision of the HWRs as the "treatment", while "participation", "enrollment", and "programs" refer to the uptake of the existing utility conservation programs shown in Table 2.

Receiving a HWR may operate through the first channel – participation – in at least three ways. First, it is possible that the HWRs simply make people aware of the existence of the conservation programs. We term this the "pure advertising effect". Second, customers may have been aware of the programs but were not motivated to capitalize on them; the HWR highlighting their relative consumption provides this motivation, driving up participation rates. Third, the social comparison may overload consumers with information, crowding out existing conservation efforts. These effects may all be occurring at once in subsets of our data, and we have no experimental approach to tease them apart. In the next section we estimate the aggregate impact of providing social norms on the probability of engaging in an additional utility program. If the aggregate impact is positive, we can conclude that some combination of the first two channels is, on average, dominating the third in our data and setting.

There may also be differences in the water savings associated with conservation programs across treatment status if the social comparisons affect how consumers sort into conservation programs. Bennear *et al.* (2013) show that it is important to account for the types of households that sign up for toilet rebates when evaluating the water savings and cost effectiveness of the program. Consider three types of households that sign up for a toilet rebate. The first household was planning to replace their toilet anyway, perhaps because of a bathroom remodel, and the utility gets no additional water savings from the money spent on that rebate. The second household is extremely motivated to conserve water, perhaps for ideological reasons, and is eager to take advantage of any programs available to lower their consumption. They may, however, already have relatively low water consumption. The third household pays less attention to their water consumption, but can be motivated to conserve water in order to lower their water bill. If social comparisons induce more of the latter two types of consumers, and especially the third, to take advantage of toilet rebates then the program will generate greater water savings per participant and the cost-effectiveness of the program will improve. For example, households that sign up for a toilet rebate in order to save water are more likely to have particularly inefficient toilets, or average more flushes per toilet, perhaps due to a large number of occupants per bathroom. These are precisely the type of consumers that a utility wants to target in their conservation programs.

4.1 Do HWRs Increase Participation Rates?

To address how treatment affects participation rates we estimate a DiD random effects panel data logit regression where the dependent variable is a dummy equal to one if a household participates in a program during a given period and zero otherwise. We pool all the programs into two time periods - before and after treatment. Table A.4 in the Appendix shows that prior to the treatment period there are no significant differences in the participation rates in the treatment and control groups. We run two different specifications of the dependent variable. The first specification pools all programs together to create a single indicator for participation, defined as "Any Program", and the second, "Any Rebate", only includes rebates for water efficient appliances. The key distinction between "Any Program" and "Any Rebate" is the inclusion of water audits in "Any Program". Water audits consist of a home visit by a water professional who provides tailored suggestions for saving water. We make this distinction because water audits are an information program that may overlap with HWRs. Additionally, rebates involve an investment in water durables that will have a persistent impact on water use, whereas water audits may suggest behavioral changes that generate water savings that wane over time. While we do have data on individual programs, because of the low number of observations these models are under-powered; raw participation data are presented in Table A.3 in the Appendix. We control for household characteristics that may influence participation, including an indicator for single family detached home, the number of bathrooms, the size of the home in square feet, and the year built. Our specification is:

$$\Pr(c_i^l = 1|X_i) = \frac{\exp(\gamma T_i \times P_t + \theta P_i + \lambda T_i + \beta X_i)}{1 + \exp(\gamma T_i \times P_t + \theta P_i + \lambda T_i + \beta X_i)} \tag{2}$$

where all the variables are the same as those defined in equation 1. Since some consumers participate in multiple programs we also run a panel data negative binomial regression treating participation data as counts.[19] The results from both models are presented as marginal effects in Table 6, with standard errors calculated by the delta method.[20] The interpretation for panel (a) is the change in the participation rate due to treatment while the parameters in panel (b) represent the expected change in the number of programs for a given household due to receiving the HWRs.

The treatment increases the probability of participation in conservation programs in the pooled specification for both Utility A and Utility B, but not in Utility C. The effects in Utility A and B, at 7.9 and 6.6 percentage points (pp) respectively, are much larger than the results found in Allcott & Rogers (2014) that range from 0.36 to 0.42 pp. The increase in participation rates in Utility A is predominantly due to the uptake of water audits; treatment has no effect on rebate participation. In contrast, treatment led to statistically and economically significant increases in rebate uptake in Utility B. The results from the count model in panel (b) of Table 6 are similar in sign and significance, but the magnitudes increase for Utilities A and B indicating that treated consumers are more likely to participate in multiple programs. In sum, these results do not suggest that HWRs are crowding out other conservation efforts.

We believe it is unlikely this is a pure advertising effect. We do not have the exact content of each HWR for all utilities or for most of our treatment period, so analysis of this data is inconclusive. A subset of customers were, however, surveyed by Utilities A and B prior to treatment about their knowledge of the utility's conservation programs. In Utility A, over 95% of households reported that the utility had contacted them about conservation programs (Appendix Table A.5). In Utility B, 96% of households reported that they had learned about programs through contact with the utility or through outreach programs such as at gardening or appliance stores (75% said the utility had contacted them directly). These proportions are balanced across treatment and control groups (Ta-

[19] A likelihood ratio test for over-dispersion rejects the null that the over-dispersion parameter is equal to zero, indicating that the negative binomial model is preferable to the Poisson model.

[20] The marginal effects are computed at the observed values of covariates with the random effect set to zero. Marginal effects computed at the sample means are all statistically significant and of similar, though slightly lower, magnitudes.

Table 6: Effect of Treatment on Enrollment

(a) Random Effects Logit

	Any Program			Any Rebate		
	(1) Utility A	(2) Utility B	(3) Utility C	(4) Utility A	(5) Utility B	(6) Utility C
Treatment Effect	0.0789***	0.0659***	-0.0058	0.0147	0.0516**	-0.0021
	(0.0207)	(0.0231)	(0.0385)	(0.0208)	(0.0231)	(0.0469)
DiD Controls	Yes	Yes	Yes	Yes	Yes	Yes
Household Controls	Yes	Yes	Yes	Yes	Yes	Yes
Households	1,884	3,092	1,924	1,884	3,092	1,924
Observations	3,729	6,100	3,511	3,729	6,100	3,511

(b) Negative Binomial

	Any Program			Any Rebate		
	(1) Utility A	(2) Utility B	(3) Utility C	(4) Utility A	(5) Utility B	(6) Utility C
Treatment Effect	0.1936***	0.1042***	0.0299	0.0395	0.0821**	0.0625
	(0.0384)	(0.0324)	(0.0607)	(0.0286)	(0.0322)	(0.0641)
DiD Controls	Yes	Yes	Yes	Yes	Yes	Yes
Household Controls	Yes	Yes	Yes	Yes	Yes	Yes
Households	1,884	3,092	1,924	1,884	3,092	1,924
Observations	3,729	6,100	3,511	3,729	6,100	3,511

Notes: The dependent variable in panel (a) is a dummy for participation in a given water utility conservation program, and in panel (b) is the count of programs in a given period. The data are pooled temporally to focus on the pre-intervention and post-intervention periods. The marginal effects, computed at the observed values with the random effect set to zero, are reported along with standard errors obtained from the delta method. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

ble A.5). While the survey may not be a representative sample, the results in Table A.5 do suggest that control group was also informed about utility programs, and that pure advertising effect is a less compelling explanation for the increase in participation rates among consumers that receive the social comparison.

4.2 Social Norms, Conservation Programs, and Water Use

We now turn to the second channel: do HWRs help attract "better" customers to conservation programs? To investigate this we jointly estimate the impact of conservation programs and social norms on water demand. This identifies the contribution of increased participation rates on the magnitude of the average treatment effect for social comparisons.[21] We attempt to identify the change in water demand for three distinct subgroups: (1) treated households that did not sign up for programs, (2) comparison households that do sign up for programs, and (3) treated households that sign up for programs. The average impact of conservation programs may not be causal because selection into the programs is endogenous and we cannot identify an appropriate instrument.

To identify the effect for the third subgroup we estimate the marginal effect of treatment on water savings in conservation programs, which is essentially an interaction term that answers the question: are average water savings from conservation programs larger in the treated population? The regressions add two terms to our preferred DiD specification in equation 1.3:

$$w_{it} = \alpha_i + \gamma T_i \times P_t + \pi_1 C^l_{it} + \pi_2 \tilde{C}^l_{it} + \theta P_t + \beta X'_{it} + \tau_t + \xi_{it} \tag{3}$$

C^l_{it} is the cumulative sum of programs that household i has participated in at time t, for l defined by the categories all programs, all rebates, and audits. We include audits separately for Utility A, but not Utility B, because audits are the most commonly adopted program among treated households, comprising over 80% of enrollments. Additionally, as stated above we expect different interactions between social norms and information programs relative to rebate programs. $\tilde{C}^l_{it}$ is the sum of programs initiated in the post-treatment period for households in the treatment group. This is not a pure interaction term ($\tilde{C}^l_{it} \neq C^l_{it} \times T_i \times P_t$) because we want to isolate programs that occur *after* the treatment begins, and C^l_{it} includes enrollment in programs before treatment began. For example, we do not want $\tilde{C}^l_{it}$ to pick up the effect of treatment on a toilet rebate that a treated household received before they actually received a HWR. Since households can participate in multiple programs, C^l_{it} and $\tilde{C}^l_{it}$ not only capture the presence but also

[21] Assessing the aggregate savings from all utility demand side management programs is beyond our scope. Rather we decompose how much of the treatment effect of HWRs is due to increased participation, and determine if the HWRs increase the effectiveness of existing conservation programs. However, when calculating the total savings from all programs it is important not to estimate the effect of programs in isolation from the HWRs in order to avoid double counting savings in the HWR that are due to increased participation (or symmetrically, savings in participation that are due to increased rates from HWRs).

the intensity of program participation. Now γ estimates the ATE while controlling for conservation programs, π_1 is the effect of the conservation program, and π_2 is the change in the effectiveness of conservation programs within the treatment group after receiving the HWRs.

Since participation in programs is endogenous, we focus on the differential effect (π_2) of the programs across treatment status for which identification relies on the random assignment into treatment. Recent work in a cross sectional setting shows in a cross sectional setting an interaction term between an endogenous and an exogenous regressor can be consistently estimated with OLS (Bun & Harrison, 2014; Nizalova & Murtazashvili, Forthcoming). Though we are exploiting non-experimental variation to identify the effect of treatment on water savings from we consider this quasi-interaction term plausibly exogenous. We cannot identify the causal full marginal effect of treated households that sign up for programs due to the endogeneity of C_{it}^l. To summarize, if π_1 is negative then a utility's conservation programs "work" by saving water. If π_2 is negative, then conservation programs save more water among treated households who participate than comparison households who participate, most likely because they nudge households who were not already planning to take advantage of a rebate or who had higher baseline water use.

One observable signature for this kind of sorting behavior in program enrollment is household water use before enrolling in a program. Therefore we examine the behavior of households *prior* to signing up for a program by running regressions similar to equation 3 that focus on the periods prior to program participation. Specifically we replace C_{it}^l with BC_{it}^l, where $BC_{i,h}^l$ is equal to one if $C_{it}^l \geq 1 \, \forall \, h < t$ and otherwise is equal to zero. Correspondingly, we replace $\tilde{C}_{it}^l$ with $\tilde{BC}_{it}^l$. Essentially BC_{it}^l is a dummy for whether a household will participate in a program at a later date, and $\tilde{BC}_{it}^l$ is a dummy for households that will participate in a program after receiving a HWR.

The results for both sets of regressions are presented in Table 7.[22] We only present the

[22]The full set of parameter estimates for these regressions are available upon request.

regressions for Utilities A and B because we found no effect of social norms on program participation in Utility C.

Table 7: Water Demand and Interactions between Social Norms and Conservation Programs

(a) After enrollment in a program

	Utiliy A			Utility B	
	(1) Any Program	(2) Any Rebate	(3) Audit	(4) Any Program	(5) Any Rebate
Treatment Effect	-4.64***	-5.05***	-4.69***	-4.77***	-4.83***
	(1.54)	(1.53)	(1.54)	(1.35)	(1.35)
Program (C_{it}^l)	-7.57***	-8.15***	-17.05***	-1.45	-2.01*
	(1.99)	(2.73)	(4.96)	(1.21)	(1.21)
Treated Programs ($\tilde{C}_{it}^l$)	0.91	-3.74	10.05*	-2.46	-1.44
	(3.01)	(8.46)	(5.29)	(5.31)	(4.81)
Household FEs	Yes	Yes	Yes	Yes	Yes
Year-Period FEs	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.183	0.183	0.183	0.181	0.181
Households	1,889	1,889	1,889	3,091	3,091
Observations	38,099	38,099	38,099	85,217	85,217

(b) Before enrollment in a program

	Utiliy A			Utility B	
	(1) Any Program	(2) Any Rebate	(3) Audit	(4) Any Program	(5) Any Rebate
Treatment Effect	-4.80***	-5.09***	-4.88***	-4.80***	-4.80***
	(1.54)	(1.53)	(1.53)	(1.34)	(1.34)
Before Program (BC_{it}^l)	8.88***	9.26***	16.21***	0.89	1.41
	(2.79)	(2.90)	(4.61)	(1.39)	(1.34)
Before Treated Program ($\tilde{BC}_{it}^l$)	4.89	3.95	-3.56	3.41	4.01
	(4.40)	(6.64)	(5.93)	(6.98)	(5.62)
Household FEs	Yes	Yes	Yes	Yes	Yes
Year-Period FEs	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.183	0.183	0.183	0.181	0.181
Households	1,889	1,889	1,889	3,091	3,091
Observations	38,099	38,099	38,099	85,217	85,217

Notes: The dependent variable is water consumption in gallons per day normalized by average post-treatment consumption in the control group. All regressions have household and reading period fixed effects as well as weather controls. Robust standard errors clustered at the household level are reported in parentheses. Panel (a) examines water use for households that have participated in a program by time t, and Panel (b) focuses on water use for households prior to their participation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Consider the results for "Any Program" in Utility A (column (1) of Table 7). All comparisons are relative to an untreated household that never signs up for any conservation.

An average treated household that does not enroll in a conservation program reduces water consumption by 4.64%. An average control household that enrolls in a conservation program saves 7.57% (π_1). An average treated household that signs up for a program saves 11.3% in total (-4.64-7.57+0.91).[23] The point estimate of π_2 for "Any Program" is positive and insignificant for Utility A, indicating that in aggregate the treatment does not increase the effectiveness of conservation programs. However, this result masks heterogeneity between different types of programs. Restricting attention to participation in rebates (column 2), the point estimate of π_2 becomes negative but not statistically-significant. π_2 is positive and marginally significant for audits (column 3): audits are *less* effective among the treated population. An audit will generate additional savings of 7% (-17.05 + 10.05) for a treated household compared to 17% for an untreated household. One explanation is that HWRs increase uptake of audits mostly among our second type of household that is already strongly motivated to save water and may be approaching diminishing marginal returns on additional conservation. For example, they may have already installed a new toilet and an irrigation controller, but when they received a HWR with audit information they thought participating in the audit might identify some additional savings. This can be seen empirically in panel (b) of Table 7 where, based on the point estimate of $\tilde{BC}_{it}^{l}$ in column (3), treated households that eventually signed up for an audit were already using less water before the audit occurred. Though this result is not statistically significant it does stand out from all the other coefficients (though also insignificant) on $\tilde{BC}_{it}^{l}$. These parameters indicate that treated households that eventually sign up for (non-audit) programs use *more* water prior to program enrollment than untreated households prior to enrollment.

Participation in any conservation program does not produce statistically significant water savings in Utility B, though the estimate for rebates is of the expected sign and significant at the 10% level. One explanation for the relative ineffectiveness of conservation programs in Utility B is that they are not attracting high water users to their programs.

[23] Recall that these last two estimates that include enrollments are not interpreted as causal results.

Columns (4) and (5) in panel (b) of Table 7 show that prior to signing up for a program, households in Utility B are roughly average water users: the coefficient on BC_{it}^{l} is small and insignificant. In contrast, households in Utility A that sign up for programs use 9-16% *more* water than the utility average prior to enrollment. This suggests that the lack of additionality (Bennear *et al.*, 2013) is of greater concern in Utility B. The HWRs may mitigate this somewhat in Utility B: the positive (but insignificant) sign on $\tilde{BC}_{it}^{l}$ suggests that treated households that will eventually sign up for a program use more water prior to enrollment than comparison households that will eventually sign up, but neither of the results are statistically significant.

Recall that the ATEs in Utilities A and B in our preferred specification (column 5 of Table 5) were reductions of 5.11% and 4.90%. When we control for participation in existing conservation programs, the ATE in Utility A declines by 0.5 percentage points (to 4.64%) and 0.1 percentage points in Utility B, corresponding to a 9.2% and 2.7% drop in the ATE respectively. An alternative approach to estimate the impact of conservation programs on the ATE is to calculate the predicted additional participants due to treatment from the negative binomial regressions and multiply that by the average savings for treated participants. Using this method we find that increased participation produced savings of 1.3 percentage points in Utility A, comprising 25% of the treatment effect.[24] We do not use this approach for Utility B since the water savings from programs is not statistically significant. Even though treatment induces dramatic increases in enrollment in utility conservation programs, these enrollments explain less than 25% of savings from providing the social comparison.

5 Discussion

In this paper we analyzed data from three randomized field experiments that harness social comparisons as a water conservation tool. The treatment caused water savings of roughly 5% in two out of the three utilities (Utilities A and B). Interacting the treat-

[24]The expected change in the number of programs due (0.1936) multiplied by the effect of programs among the treated population (-7.57=0.91) equals 1.3, which is 25% of the base effect (5.11).

ment with deciles of the baseline consumption corroborates the general results of other social norm campaigns (Allcott, 2011; Ferraro & Miranda, 2013) that treatment effects are largest for high-use households. Higher users are more likely to be more responsive because they have more flexibility to reduce consumption and have the potential to achieve greater financial gains from conservation. Similar to existing studies, the savings are durable throughout the treatment in Utility A, but seem to attenuate completely at the end of our data in Utility B.

In the third utility (Utility C), however, we do not find a statistically significant treatment effect, which is to our knowledge the first such finding in the water and energy literature, though publication bias may be a factor here. There are at least two potential reasons for this. First, as described in Section 2.3, Utility C's individualized, "allocation-based" rate structure already provides very strong incentives for conservation, particularly for households with large water use that are most responsive to social norms. Households above 150% of their allocation pay punitive prices that are much higher than any price faced by households in Utility A and Utility B. A significant portion of households are paying water rates in excess of $4.80/ccf and over 10% pay prices close to $10/ccf (see Appendix Figure A.5 for prices by decile of baseline consumption in Utility C). Furthermore, the three upper tiers are explicitly labeled in the rate structure as "inefficient", "excessive" and "wasteful", which may impart a normative message very similar to the HWR to households in the control group. In fact the normative message on the HWR is linked to household water use relative to their allocation in addition to a peer group.[25] Second, reports were emailed in Utility C rather than mailed in the other utilities. Dolan & Metcalfe (2013) similarly find that emailed Home Energy Reports do not have a statistically-significant impact. Further research explicitly comparing delivery modes within an experimental design would be useful.

Utilities have several options for water conservation at their disposal; so how cost-effective are Home Water Reports in Utilities A and B? WaterSmart charges utilities an

[25] While this linkage could impact the treatment effect Allcott (2011) finds no significant changes in consumption due to the specific normative message a household receives.

average of $10 per customer per year for a print HWR program. (It is important to note that WaterSmart also performs data analysis, conducts surveys on water use and attitudes, and acts as an active partner in water conservation so the water savings from HWRs are only one component of the benefits to the utility.) The cost per thousand gallons saved is $2.61 in Utility A and $1.73 in Utility B. We estimate that the marginal price increase required to achieve the same reduction in water use using a short-run demand elasticity of -0.38 (Olmstead, 2010) would be 13.6% and 14.5% in Utilities A and B respectively. This would increase an average household's annual water bill by approximately $26 dollars in Utility A and $67 in Utility B. These relatively large price changes to reduce water consumption stem from inelastic water demand, which is one reason why non-pecuniary programs such as social norms are attractive to utilities. The cost effectiveness of HWRs compare favorably with other conservation programs, such as the estimate of $11-15 per thousand gallons from high efficiency toilet rebates estimated in Bennear *et al.* (2013). Even if toilet rebates produce 100% additive savings (which they do not find) the cost of conservation would still be $4-5 per thousand gallons. Our estimates are higher, however, than the estimate of $0.37 per thousand gallons for the Cobb County program (Ferraro & Miranda, 2013). We find roughly the same average treatment effects, but their better cost-effectiveness is driven largely by remarkably "persistent" treatment effects: the social comparison was only sent one time yet effects are still detectable (and countable, in a cost-effectiveness calculation) six years later. We examine a shorter treatment period, and it is possible that if HWRs were ended in the two utilities those areas too might continue to see treatment effects.

Finally, we contribute to an important gap in the literature by finding that sending Home Water Reports in these two utilities increased uptake of existing conservation programs substantially. This contributes up to 25% of the ATE and demonstrates that social norms do not crowd out existing conservation programs. The effect on participation rates are much larger than in studies using social norms for energy conservation (Allcott & Rogers, 2014). Additionally, we find weak evidence that social norms may actually improve the effectiveness of rebate programs by promoting inefficient households to replace appliances, however there is considerable noise in the empirical estimates. Therefore combining social norms with rebate programs may alleviate concerns of additionality raised by Bennear *et al.* (2013). Improvements in conservation programs do not extend to information campaigns such as home audits; rather increased uptake in these programs within the treatment group are likely a signal of households with a proclivity towards water conservation.

Observing concrete water conservation actions also improves our knowledge about the psychological drivers through which social norms cause consumers to reduce demand. Treated households that sign up for audits already used less water before the audit; thus the audit serves as a signal of pro-conservation attitudes. The social comparison likely motivates these households through moral suasion. In contrast, treated households that eventually sign up for rebates are high water users that reap large financial rewards from rebates, indicating that the information in the HWR catalyzes them to consider cost-effective ways to reduce their bill. These two patterns in the data provide some evidence that social norms are triggering both intrinsic and extrinsic motivations . However, these conclusions are based on observing relatively few households that were both treated and enrolled in existing conservation programs, and future work should test these hypotheses explicitly in an experimental setting. Our findings suggest that utilities could strategically integrate social norms into their set of existing conservation programs by targeting consumers who are large users and have not yet signed up for water efficiency rebates. Utilities can leverage social comparisons to improve the effectiveness of existing programs. While we identified one of the mechanisms that translate social norms to water savings, it is important to continue searching for other actions that consumers take in response to social comparisons in order to better understand how they will work in different settings.

References

Allcott, Hunt. 2011. Social norms and energy conservation. *Journal of Public Economics*, 95(9-10):1082–1095.

Allcott, Hunt, and Rogers, Todd. 2014. The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, 104(10):3003–3037.

Attari, Shahzeen Z. 2014. Perceptions of water use. *Proceedings of the National Academy of Sciences*, 111(14):5129–5134.

Ayres, Ian, Raseman, Sophie, and Shih, Alice. 2012. Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *Journal of Law, Economics, and Organization*, 29(5):992–1022.

Bennear, Lori S, Lee, Jonathan M, and Taylor, Laura O. 2013. Municipal rebate programs for environmental retrofits: An evaluation of additionality and cost-effectiveness. *Journal of Policy Analysis and Management*, 32(2):350–372.

Bernedo, Maria, Ferraro, Paul J, and Price, Michael. 2014. The persistent impacts of norm-based messaging and their implications for water conservation. *Journal of Consumer Policy*, 37(3):437–452.

Bolsen, Toby, Ferraro, Paul J., and Miranda, Juan Jose. 2014. Are voters more likely to contribute to other public goods? Evidence from a large-scale randomized policy experiment. *American Journal of Political Science*, 58(1):17–30.

Bun, Maurice JG, and Harrison, Teresa D. 2014. OLS and IV estimation of regression models including endogenous interaction terms. UVA Econometrics Discussion Paper 2014/02, University of Amsterdam, Amsterdam, NL.

Costa, Dora L., and Kahn, Matthew E. 2013. Energy conservation "nudges" and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association*, 11(3):680–702.

Dolan, Paul, and Metcalfe, Robert. 2013. Neighbors, knowledge, and nuggets: Two natural field experiments on the role of incentives on energy conservation. CEP Discussion Paper No. 1022, Centre for Economic Performance, Oxford, UK.

Ferraro, Paul J., and Miranda, Juan Jose. 2013. Heterogeneous treatment effects and mechanisms in information-based environmental policies: Evidence from a large-scale gield experiment. *Resource and Energy Economics*, 35(Apr.).

Ferraro, Paul J., and Price, Michael K. 2013. Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. *Review of Economics and Statistics*, 95(1):64–73.

Ferraro, Paul J., Miranda, Juan José, and Price, Michael K. 2011. The persistence of treatment effects with norm-based policy instruments: Evidence from a randomized environmental policy experiment. *American Economic Review: Papers and Proceedings*, 101(3):318–322.

Goldstein, Noah J, Cialdini, Robert B, and Griskevicius, Vladas. 2008. A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, 35(3):472–482.

Grafton, R. Quentin, and Ward, Michael B. 2008. Prices versus rationing: Marshallian surplus and mandatory water restrictions. *Economic Record*, 84(September 2008):S57–S65.

Hirst, Eric, and Grady, Stephen. 1982. Evaluation of a Wisconsin utility home energy audit program. *Journal of Environmental Systems*, 12(4):303–320.

Hutton, R Bruce, and McNeill, Dennis L. 1981. The value of incentives in stimulating energy conservation. *Journal of Consumer Research*, 291–298.

Jacobsen, Grant D., Kotchen, Matthew J., and Vandenbergh, Michael P. 2012. The behavioral response to voluntary provision of an environmental public good: Evidence from residential electricity demand. *European Economic Review*, 56(5):946–960.

Jacobsen, Grant D, Kotchen, Matthew J, and Clendenning, Greg. 2013. Community-based incentives for environmental protection: The case of green electricity. *Journal of Regulatory Economics*, 44(1):30–52.

Klaiber, H Allen, Smith, V Kerry, Kaminsky, Michael, and Strong, Aaron. 2014. Measuring price elasticities for residential water demand with limited information. *Land Economics*, 90(1):100–113.

Kotchen, Matthew J., and Moore, Michael R. 2007. Private provision of environmental public goods: Household participation in green-electricity programs. *Journal of Environmental Economics and Management*, 53(1):1–16.

Kurz, Tim, Donaghue, Ngaire, and Walker, Iain. 2005. Utilizing a social-ecological framework to promote water and energy conservation: A field experiment. *Journal of Applied Social Psychology*, 35(6):1281–1300.

Luyben, Paul D. 1982. Prompting thermostat setting behavior: Public response to a presidential appeal for conservation. *Environment and Behavior*, 14(1):113–128.

Mansur, Erin T., and Olmstead, Sheila M. 2012. The value of scarce water: Measuring the inefficiency of municipal regulations. *Journal of Urban Economics*, 71(3):332–346.

Nizalova, Olena Y, and Murtazashvili, Irina. Forthcoming. Exogenous treatment and endogenous factors: Vanishing of omitted variable bias on the interaction term. *Journal of Econometric Methods*.

Olmstead, Sheila M. 2010. The economics of managing scarce water resources. *Review of Environmental Economics and Policy*, 4(2):179–198.

Olmstead, Sheila M., Hanemann, W. Michael, and Stavins, Robert N. 2007. Water demand under alternative price structures. *Journal of Environmental Economics and Management*, 54(2):181–198.

Renwick, Mary E, and Archibald, Sandra O. 1998. Demand side management policies for residential water use: Who bears the conservation burden? *Land Economics*, 74(3):343–359.

Schultz, P Wesley, Nolan, Jessica M, Cialdini, Robert B, Goldstein, Noah J, and Griskevicius, Vladas. 2007. The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18(5):429.

Shang, Jen, and Croson, Rachel. 2009. A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *The Economic Journal*, 119:1422–1439.

Siero, Frans W, Bakker, Arnold B, Dekker, Gerda B, and Van Den Burg, Marcel TC. 1996. Changing organizational energy consumption behaviour through comparative feedback. *Journal of Environmental Psychology*, 16(3):235–246.

Staats, Henk, Harland, Paul, and Wilke, Henk AM. 2004. Effecting durable change: A team approach to improve environmental behavior in the household. *Environment and Behavior*, 36(3):341–367.

Winett, Richard A, Love, Susie Q, and Kidd, Charlotte. 1982. The effectiveness of an energy specialist and extension agents in promoting summer energy conservation by home visits. *Journal of Environmental Systems*, 12(1):61–70.

Woolridge, Jefferey M. 2002. Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: MIT Press.