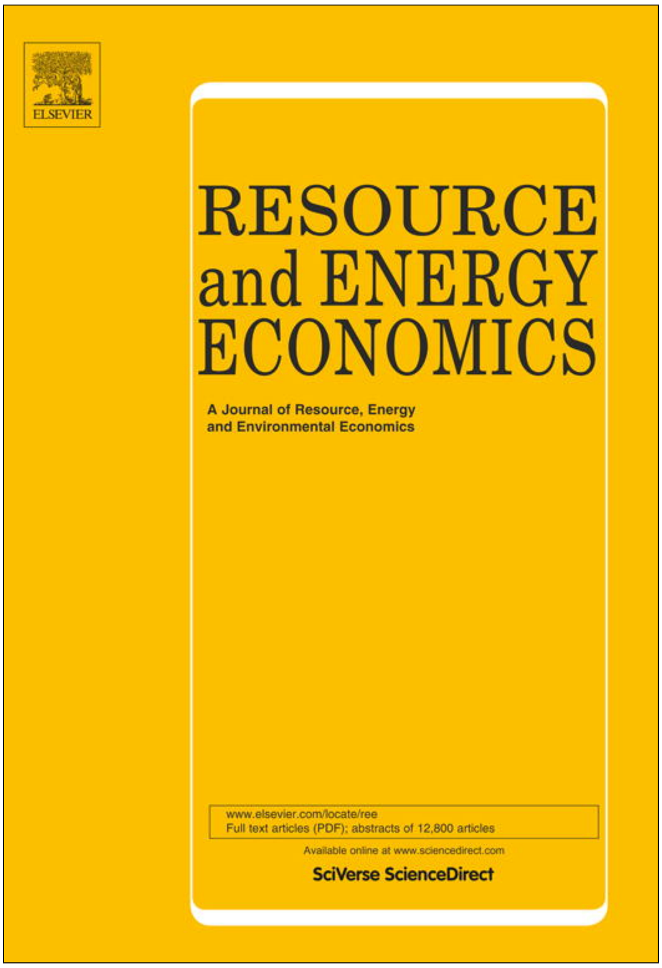
ELSEVIER
RESOURCE
and ENERGY
ECONOMICS
A Journal of Resource, Energy
and Environmental Economics
www.elsevier.com/locate/ree
Full text articles (PDF); abstracts of 12,800 articles
Available online at www.sciencedirect.com
SciVerse ScienceDirect

Resource and Energy Economics 35 (2013) 356–379

Contents lists available at SciVerse ScienceDirect

Resource and Energy Economics

journal homepage: www.elsevier.com/locate/ree

Heterogeneous treatment effects and mechanisms in information-based environmental policies: Evidence from a large-scale field experiment☆

Paul J. Ferraro[a,*], Juan José Miranda[b]

[a] Department of Economics, Andrew Young School of Policy Studies, Georgia State University, PO Box 3992, Atlanta, GA 30302-3992, United States
[b] Inter-American Development Bank, 1300 New York Avenue, N.W., Washington, DC 20577, United States

ARTICLE INFO

Article history:
Received 22 March 2012
Received in revised form 28 March 2013
Accepted 6 April 2013
Available online xxx

JEL classification:
D03
C93
L95
Q25

Keywords:
Program evaluation
Experimental design
Conditional average treatment effects
Quantile average treatment effects
Other-regarding preferences
Social norms

ABSTRACT

Policymakers often rely on non-pecuniary, information-based programs to achieve social objectives. Using data from a water conservation information campaign implemented as a randomized controlled trial, we estimate heterogeneous household responses. Understanding such heterogeneity is important for improving the cost-effectiveness of non-pecuniary programs, extending them to other populations and probing the mechanisms through which the treatment effects arise. We find little evidence of heterogeneous responses to purely technical information or to traditional conservation messages that combine technical information and moral suasion. In contrast, norm-based messages that combine technical information, moral suasion and social comparisons exhibit strong heterogeneity: households that are wealthier, owner-occupied and use more water are more responsive. These subgroups tend to be least responsive to pecuniary incentives. We find no evidence that any subgroup increases their water use in response to the messages. By targeting the messages to subgroups known to be most

☆ This material is based upon work supported by the Cooperative State Research, Education, and Extension Service, U.S. Department of Agriculture, under Agreement No 2003-38869-02007. For providing useful comments that improved the study, the authors thank participants at the workshop in Behavioral Environmental Economics at the Toulouse School of Economics, at the 2011 and 2012 fall research conferences of the Association of Policy Analysis and Management, and the 2010 North Carolina State University Camp Resources workshop, as well as Hunt Allcott, Lori Bennear, Joseph Cook, Michael McKee, Kerry Smith, Laura Taylor and two anonymous reviewers.

* Corresponding author.
E-mail addresses: pferraro@gsu.edu (P.J. Ferraro), juanmi@iadb.org (J.J. Miranda).

http://dx.doi.org/10.1016/j.reseneeco.2013.04.001

responsive, program costs could be reduced by over 50% with only a 20% reduction in the treatment effect. Combining theory and data, we also shed light on the mechanisms through which the treatment effects arise, which has implications for program design and future research on the program's welfare effects.

1. Introduction

Non-pecuniary, information-based environmental policy strategies have long been used to influence individual decision-making (e.g., Smith et al., 1990; Smith and Desvousges, 1990) and are growing in popularity in all social policy fields (Thaler and Sustein, 2008; House of Lords, 2011). Such strategies include norm-based persuasive messages, commitment devices, changes to default options and the provision of technical information to lower transaction costs of information acquisition. Under standard economic assumptions of perfectly informed, rational, self-interested agents, these strategies should be ineffective. However, under behavioral theories that include other-regarding preferences or bounded rationality, they may be effective. A growing empirical literature in economics and psychology suggests that such strategies can indeed affect policy-relevant behaviors (e.g., Bui and Mayer, 2003; Duflo and Saez, 2003; Jin and Leslie, 2003; Bjørner et al., 2004; Schultz et al., 2007; Goldstein et al., 2008; Bennear and Olmstead, 2008; Allcott and Mullainathan, 2010a,b; Habyarimana and Jack, 2011).

In the context of environmental policies and programs, the conceptual and empirical foundations of such strategies remain under-researched (Shogren and Taylor, 2008). A new literature uses randomized controlled trials, which are rare in environmental economics (Greenstone and Gayer, 2009), to test the impacts of non-pecuniary, norm-based messages on environmental outcomes such as energy use (e.g., Ayres et al., 2009; Yoeli, 2009; Allcott, 2012; Costa and Kahn, 2013) and water use (e.g., Ferraro and Price, 2013; Ferraro et al., 2011). These studies find that sending pro-social messages and social comparisons that contrast own consumption to peer-group consumption can reduce, on average, water and energy consumption. Moving beyond estimating average effects, however, is important. Ideally, one would want to also understand the heterogeneity of responses across households (i.e., which subgroups are most responsive) and the mechanisms through which the messages affect behavior.

Understanding the heterogeneous treatment effects of information-based programs yields at least three policy and research-relevant insights. First, policy makers can use this information to more cost-effectively target the treatments to subgroups that are most responsive (Heckman et al., 1997; Djebbari and Smith, 2008). By targeting the subgroups that are most responsive, policy makers avoid wasting money (and political capital) sending information to non-responsive subgroups or subgroups that may react in ways contrary to the policy objective. This tailoring of messages to specific subgroups also helps avoid information overload that can affect decision-making as these information-based approaches grow in scope.[1] Second, understanding heterogeneous responses helps strengthen the generalizability and external validity of randomized controlled trials to different target populations (Angrist, 2004; Manski, 2004; Hotz et al., 2005; Allcott and Mullainathan, 2010a,b; Imai and Ratkovic, 2013). The mean effects of the same experimental design could be different when applied in other populations with different distributions of observable characteristics. Third, by combining theory and information on heterogeneous response, one can explore potential mechanisms through which the causal effects are generated. As Deaton (2010) noted in his critique of the way in which randomized controlled trials are done in economics, we need to move from determining *whether* a treatment is effective to determining

[1] We thank Associate Editor McKee for pointing this out.

why it is effective.[2] With a better understanding of mechanisms, information-based programs can be better designed to improve their efficiency.[3]

We study heterogeneity and mechanisms in the context of a large-scale field experiment that was run in 2007 in partnership with a water utility in metropolitan Atlanta, GA, USA. To induce voluntary reductions in water use during a drought, three types of messages were sent, at random, to households. The three treatments comprised: (i) a tip sheet with information about reducing water consumption (pure information message), (ii) a tip sheet and a personalized letter promoting pro-social behavior (weak social norm message); and (iii) a tip sheet, personalized letter promoting pro-social behavior, and a social comparison of the household's water consumption with the median county consumption (strong social norm message). Each treatment group comprised roughly 11,700 houses and the control group comprised roughly 71,800 houses. Ferraro and Price (2013) report short-term average treatment effects and Ferraro et al. (2011) report longer-term average treatment effects.

When estimating heterogeneous treatment effects from experimental or non-experimental data, there is a substantial risk of labeling spurious correlations as conditional treatment effects. We mitigate this risk through our experimental design and a multi-step framework that trades increasingly stringent assumptions for increasingly precise characterizations of the heterogeneity. We find little evidence of heterogeneous responses to the pure information and weak social norm messages, but strong evidence of heterogeneous responses to the strong social norm message: households that are wealthier, owner-occupied and use more water are more responsive. These households are identified in the literature to be *least* responsive to pecuniary incentives. In contrast to studies from the psychology and behavioral economics literature that predict some users may respond to social comparisons by increasing their use (e.g., Schultz et al., 2007), we find no evidence that low users or any other subgroups increase their water use in response to the messages.

We demonstrate how information on heterogeneous responses can be used to improve the program's cost-effectiveness through more precise targeting of messages using publicly available data. With regard to mechanisms, the evidence suggests that norm-based messages induce behavioral (variable-cost), rather than technological (fixed-cost), changes in outdoor water use and work through social preferences, rather than by serving as signals of privately efficient behavior to boundedly rational agents. The next section reviews the most relevant literature. Section 3 describes our methodology. Section 4 describes the experimental design and data. Section 5 provides results.

2. Experiments with information-based environmental programs

A small number of experimental studies estimate the effects of pro-social and social comparison messages on energy and water use. Schultz et al. (2007), Ayres et al. (2009), Allcott (2012), Costa and Kahn (2013) and Yoeli (2009) focus on energy consumption, while Ferraro and Price (2013) and Ferraro et al. (2011) focus on water consumption. Although the studies differ in terms of location and the content and framing of the messages, the authors find that pro-social messages with social comparisons reduce consumption.[4]

Only three studies examine heterogeneous treatment responses. Allcott (2012) runs a quantile regression and Ferraro and Price (2013) examine how treatment responses vary as function of being below or above the median use. Both studies report that larger users appear to respond more, on average, to social comparison messages. Costa and Kahn (2013) use data from the California

[2] Pawson and Tilley (1997) argue that 30 years of program evaluation in sociology, education and criminology has been largely uninformative because it has focused on whether programs work, instead of on why they work.

[3] Evidence of heterogeneity also informs future non-experimental studies that use instrumental variable (IV) or panel data estimators to estimate treatment effects from information-based strategies. For IV estimators, heterogeneity implies that the estimates should be interpreted as local average treatment effects (LATE) rather than population average treatment effects (ATE). For panel data estimators, which typically assume homogenous treatment effects and use conditional variance weighting on observables, heterogeneity implies that the coefficient on the treatment effect is likely under- or over-estimated, depending on the nature of the heterogeneity.

[4] Studies on social comparisons without pro-social messages, however, have found no effect (see, for example, the review by Fischer, 2008), suggesting that mixing both types of messages may be necessary.

OPower experiment to test whether responses vary with political affiliation (obtained for a subsample from public records). A particular subgroup of Democratic households, on average, reduces their consumption by 3.9%, whereas a particular subgroup of Republican households, on average, reduces by 1.1%. These three studies do not examine the likely validity of assumptions required to interpret the correlations as conditional causal treatment effects and do not examine heterogeneity further. Moreover, these studies do not probe the mechanisms through which treatment effects are generated.

3. Methodology

When exploring treatment effect heterogeneity, the subgroups are formed after the experiment is implemented. Thus there is a substantial risk of finding statistically significant differences among subgroups when no true treatment effect heterogeneity exists (Imai and Strauss, 2011). To mitigate this potential bias, we use five complementary steps to estimate heterogeneous treatment effects. First, prior to the analysis, we select only a few subgroups based on theory, field experience and policy relevance. Second, we demonstrate that, although randomization was not conducted within these subgroups, our large sample size, combined with randomization within 390 small neighborhood strata, generated within-subgroup balance in pre-treatment water use among treated and control households. This balance suggests no systematic bias when drawing inferences from treatment-control outcome contrasts within subgroups. Third, we begin testing for heterogeneity with a nonparametric approach developed by Crump et al. (2008), which tests for the presence of heterogeneity without attempting to characterize the nature of the heterogeneity. Then we impose additional assumptions and estimate quantile treatment effects (Firpo, 2007; Bitler et al., 2005, 2006, 2008; Djebbari and Smith, 2008; Heckman et al., 2002). Finally, we isolate systematic variations among subgroups through interactions terms between the treatment variable and other covariates.

3.1. Heterogeneity in treatment responses

To estimate heterogeneous responses, we select covariates (subgroups) that are observable to policymakers and that theory or empirical studies suggest could be important modifiers of the treatment effects. We wish to keep the number of covariates small to avoid charges of data mining. We select (i) two measures of previous water use, (ii) three household characteristics, and (iii) two neighborhood characteristics. In our final analysis that uses multiple hypothesis tests, we adjust the Type I error rate for sequential tests. We present the covariates in the order we believe reflects their policy-relevance, and thus the order of the sequential tests.

Ferraro and Price's analysis shows that previous water use predicts future water use. Moreover, for utilities, previous water use is the easiest characteristic on which to target future messages. We use the two variables used by Ferraro and Price (2013) in their treatment effect regressions: June–November 2006 billed use (corresponds to May–October use, which is the main water use season) and April–May 2007 billed use (to reflect changes in landscaping prior to treatment assignment in May 2007).

Mansur and Olmstead (2012) find that, as theory would predict, high-income households in urban areas are less price-sensitive to changes in residential water prices. Whether such households are more or less responsive to *non-pecuniary* approaches is an open empirical question. We cannot observe household income, but we can observe the fair market value of the home in the year in which the treatment was assigned. Based on the high correlation between housing value and income, we use fair market value as a proxy for income (and wealth).[5]

Davis (2012) shows that renters are significantly less likely to have energy efficient appliances, like clothes washers and dishwashers. These results are consistent with the hypothesis that when tenants pay the utility bills, landlords may buy cheap inefficient appliances. In our sample, almost all renters

[5] The 2007 American Housing Survey shows a high correlation (>0.90) between housing values and incomes, and thus we believe it is reasonable to assume a similar or higher correlation between housing values and wealth.

are directly billed (multi-dwelling structures, like apartment buildings, are not in our sample). With regard to water conservation, owner-occupants have a greater incentive to invest in cost-saving water conservation innovations that are capitalized into the value of the home. Owner-occupants may also have greater social connections to their neighbors and thus be more responsive to pro-social messages. Rohe et al. (2001) posit that homeowners are more likely to participate in community activities and might be more civically active because they have higher location-based investments (homes) than renters, higher transaction costs associated with moving, and stronger expectations of staying in their homes longer. DiPasquale and Glaeser (1999) offer evidence that homeowners are more likely to have greater social capital than renters (e.g., homeowners are more likely to participate in solving local problems, and more likely to be members of non-professional organizations).

Owner-occupants, however, may have weaker incentives than renters to reduce water consumption. For example, DiPasquale and Glaeser (1999) also found that homeowners are 12 percent more likely to garden than non-homeowners. Furthermore, landscaping may be detrimentally affected by water conservation, which could affect home values (landlords may be unable to shift this risk to tenants). Ownership status is revealed by the owner's homestead exemption status (only owner-occupiers receive a homestead exemption).

The age of the home also reflects the scope and incentives for water conservation. Older homes, on average, have older water-intensive capital (e.g., toilets), which are more cost-effective to replace, and they are more likely to have repairable leaks.[6]

Environmental preferences of household occupants may also affect treatment responses. We cannot observe environmental preferences, but survey evidence suggests that environmental preferences vary with education levels and race (e.g., Greenberg, 2005). Because we (and water utilities) cannot observe education and race at the household-level, we use measures of education (percent with bachelor's degree or higher) and race (percent white) at the census block group. The average number of households per block group in our sample is about 425 households.

Of course, other covariates may moderate treatment effects in our experiment, such as risk or time preferences. While they may be relevant for theory, this kind of heterogeneity is less relevant for policy makers because it is not easily observed. Recall, we are not making claims that the observable characteristics themselves cause the observed differences in treatment responses. Instead, we wish to measure heterogeneous treatment responses conditional on observable characteristics, with which policymakers can improve program targeting, gain insights into the external validity of experimental results, and better understand the mechanisms through which the treatments operate.

3.2. Nonparametric tests

Crump et al. (2008) propose tests of two null hypotheses: the conditional average treatment effect is equal to zero (Zero CATE) and the conditional average treatment effect is constant (Constant CATE). We conduct both tests using all the subgroup covariates described in Section 3.1 and we reject the null for *p*-values less than 0.05. The Zero CATE null hypothesis states that the impact of a program is zero on average for all subgroups. Testing this hypothesis is relevant for Treatment 1 (pure information), which did not have a mean treatment effect different from zero, but which may have had an impact for some subgroups. For each of the other two treatments, which generated nonzero mean impacts, the natural question is whether the treatment effects are constant across subgroup; i.e., the Constant CATE null hypothesis. Crump et al. (2008) prove that both tests can be implemented through regression analysis using an ordinary least squares (OLS) estimator.

The null hypothesis for the Zero CATE test is that the average effect for the subpopulation with covariate values X is equal to zero for all values of X. The alternative hypothesis is that the average effect for the subpopulation with covariate values X is different from zero for some values of X. To test this hypothesis, we run an OLS regression for treated and control group separately controlling for X. After obtaining the quadratic form of the difference of estimated coefficients vector ($e(\beta_1 - \beta_0)$),

[6] We do not explore heterogeneity conditional on lot size because lot size is highly correlated with fair market value.

we divide it by the variance–covariance matrix ($e(V_1 + V_0)$). This test statistic follows a chi-square distribution with k degrees of freedom:

$$e(\beta_1 - \beta_0)[e(V_1 - V_0)]^{-1}e(\beta_1 - \beta_0) \sim \chi^2_{(k)}$$

The null hypothesis for the Constant CATE test is that the average treatment effect (ATE) for the subpopulation with covariates values X is equal to the ATE for all values of X. The alternative hypothesis is that the average effect for subpopulation with covariates value X is different from the ATE for some values of X. To test this hypothesis, we run an OLS regression with treated and control groups controlling for X. After obtaining the quadratic form of the estimated coefficients vector excluding the constant term ($e_{k-1}(\beta_1 - \beta_0)$), we divide it by the variance–covariance matrix excluding the constant term ($e_{k-1}(V_1 + V_0)$). The constant term is excluded because it represents the average effect for everybody. This test statistic follows a chi-square distribution with $k-1$ degrees of freedom (because the constant term is excluded):

$$e_{k-1}(\beta_1 - \beta_0)[e_{k-1}(V_1 - V_0)]^{-1}e_{k-1}(\beta_1 - \beta_0) \sim \chi^2_{(k-1)}$$

Following Crump et al. (2008, p. 397), we select the model specification in three ways: (i) include all covariates; (ii) 'top down' selection of covariates, where one starts with the full set of covariates and sequentially (one by one) drops the covariate with the smallest t-statistic until all remaining covariates have a t-statistic larger than or equal to 2 in absolute value; and (iii) 'bottom up' selection of covariates, where, for each covariate, one runs K regressions with just an intercept and the covariate (K = number of covariates), and then selects from this set the covariate with the highest t-statistic, after which one runs, for each of the remaining covariates, $K-1$ similar regressions, choosing the one with the highest t-statistic, and continuing the process until no potential covariate has a t-statistic equal to or above 2 in absolute value. We present results using the most flexible specification with higher degree order terms of continuous variables (Crump et al., 2008).[7]

3.3. *Quantile treatment effects*

The nonparametric tests described in Section 3.2 provide evidence of whether heterogeneous treatment effects exist, but do not characterize this heterogeneity. The next step is to impose some parametric assumptions to begin characterizing this heterogeneity. Within a quantile regression framework, we use estimates of the effects of the treatment on the outcome distribution, or quantile treatment effects (QTE), to infer the presence of heterogeneous treatment effects. QTE show the difference of two marginal distributions at different quantiles (Imbens and Wooldridge, 2009), τ_q,

$$\tau_q = F^{-1}_{Y(1)}(q) - F^{-1}_{Y(0)}(q) \tag{1}$$

rather than the quantile of treatment effect, $\tilde{\tau}_q$,

$$\tilde{\tau}_q = F^{-1}_{Y(1)-Y(0)}(q). \tag{2}$$

In other words, quantile regressions tell us about the effects on the outcome distribution, rather than on households, which is sufficient to inform us about the presence of heterogeneous treatment effects across quantiles of water use.

Although we do not need an estimate of the effects on households – i.e., the distribution of treatment effects – to infer heterogeneity of causal effects, an estimate of this distribution could be useful to policy makers. For example, one could use the distribution of treatment effects to infer the fraction of the sample for which the treatments increased water use. To estimate this distribution, one needs the joint conditional distribution of treated and untreated states (Heckman et al., 1997; Djebbari and Smith, 2008). Randomized experiments, however, only provide the marginal distribution of treated

[7] The tests were conducted using the command *test_condate* for STATA, which is available at Oscar Mitnik's website: http://moya.bus.miami.edu/∼omitnik/.

outcomes and the marginal distribution of untreated outcomes (which permit the estimation of the ATE).

Nevertheless, under a rank preservation assumption, QTE estimate the distribution of treatment effects (Firpo, 2007; Bitler et al., 2005, 2006, 2008). Rank preservation implies that household's ranks in the outcome distribution are the same regardless of whether they are assigned to treatment or control groups (Bitler et al., 2008). If household ranks do not change under exposure to the treatment, the ranks in the two marginal distributions from the experiment correspond. For example, the median outcome in the treated distribution has as its counterfactual the median outcome in the untreated distribution (and so on for all quantiles). The impact of the treatment on the distribution would thus be equivalent to the distribution of treatment effects. Paraphrasing Angrist and Pischke (2009), assume, for example, that a message lowers the bottom decile of the water use distribution. We would not necessarily know if someone who would have been a low user without the message is now using less water. We know only that those who use less water with the message are using less water than bottom-decile users would have used without the message. They may not be the same users. In contrast, if the rank preservation assumption holds, the same discovery would mean that the message reduces use among users at the quantile being examined (e.g., the message reduced use among users in the bottom decile).

To test the rank preservation assumption, we follow Bitler et al. (2005) and Djebbari and Smith (2008) by using observable covariates of treated and control households. If these covariates vary significantly between treated and control groups in a given quantile, the variation provides evidence against the rank preservation assumption. Dividing our sample into quartiles, we find that 25 percent of the 84 possible combinations (7 covariates, 4 quartiles and 3 treatments) show statistically significant differences (without adjusting for the multiple sequential tests, which would reduce the number of null hypotheses rejected).[8] These results suggest that some rank reversal may be present based on the covariates selected. In particular, households with high fair market value and high previous water consumption may have migrated down the outcome distribution when treated. We conclude there is support for viewing our QTE results as a useful approximation to the distribution of treatment effects.

3.4. *Subgroup analysis*

The nonparametric and quantile regression approaches do not specify which subgroups are most responsive to treatment. Using the covariates in Section 3.1, a household is labeled as "high value" if its value is above the median. For ownership status, the subgroups are owners and renters. We explore subgroup variation through interactions terms between the treatment variables and the high-value (or owner) dummy variables in a regression framework (Heckman et al., 1997, 2002; Djebbari and Smith, 2008). To guard against spurious findings, we first conduct an *F*-test to test the null hypothesis that overall that there are no subgroup differences (Type I error rate = 0.05). If this hypothesis is rejected, we look at each subgroup in turn, after adjusting the Type I error rate for sequential hypothesis testing with a conservative Bonferroni adjustment (i.e., we take our pre-determined Type I error rate 0.05 and divide it by the number of tests; the null of no difference is rejected if $p < 0.0075$). Recall that we are estimating causal effects conditional on observable characteristics, not making claims that the observable characteristics themselves cause the observed differences in treatment responses.

4. Experimental design and data

The Cobb County Water System (CCWS) experiment comprised three treatment groups and one control group (see Ferraro and Price, 2013 for the original treatment letters):

- Pure Information (Treatment 1): A 'tip sheet' listing different ways to most effectively reduce water use.

[8] Using the same test and data from PROGRESA, the Mexican conditional cash transfer program, Djebbari and Smith (2008) rejected 30% and Lehmann (2010) rejected 31% of the possible combinations.

Table 1
Pre-treatment descriptive statistics.

Variable	(1) Technical advice (T1)	(2) Weak social norm (T2)	(3) Strong social norm (T3)	(4) Control	(5) *F*-Statistic	(6) *p*-Value
Water consumption June–November 2006[a]	58.286	58.012	58.381	58.142	0.20	0.90
Water consumption April–May 2007[a]	15.952	15.841	15.957	15.867	0.38	0.77
House's fair market value ($)	257,824	260,984	260,888	258,647	0.95	0.42
Age of house	20.753	20.830	20.710	20.723	0.23	0.88
% owner occupiers	0.844	0.836	0.835	0.844	3.19	0.02
% population 25 years ≥bachelor[b]	0.728	0.727	0.728	0.728	0.02	0.99
% of households white[b]	0.842	0.842	0.843	0.842	0.16	0.92

Sources: Experimental Data, 2007 Cobb County Tax Assessor Database, 2000 US Census.
[a] In thousands of gallons.
[b] At census block group level.

- Weak Social Norm (Treatment 2): The 'tip sheet' and a personally addressed letter from CCWS officials encouraging water conservation.
- Strong Social Norm (Treatment 3): The 'tip sheet', the letter from CCWS officials encouraging water conservation, and a social comparison that compared the household's 2006 summer water use to the median County household us. Summer season is from June to September, which is reflected in July to October monthly bills.

In May 2007, the three treatments and control were randomly assigned (mailed) to all residential customers who lived in their homes from May 2006 to April 2007 and used at least 20,000 gallons during the 2006 summer watering season (about 80 percent of the population). Each treatment was sent to roughly 11,700 houses, with roughly 71,800 houses serving as controls. For details about the treatments and experimental design, see Ferraro and Price (2013) and Ferraro et al. (2011). Ferraro and Price (2013) find that pure information (Treatment 1) had a small and statistically insignificant effect, while the weak social norm message (Treatment 2) reduced water use by about 2.5 percent ($p < 0.01$). In contrast, the strong social norm message (Treatment 3) reduced water use by almost 5 percent ($p < 0.01$). Ferraro et al. (2011) find that only the strong social norm message significantly affects water in the following years; up to three watering seasons after treatment assignment, albeit with smaller effects.

We merged the experimental data with the 2007 County Tax Assessor Database and the 2000 US Census (census block group) using home addresses as the merger link. Tax Assessor data provide relevant information about fair market value, ownership status and the age of the home. The Census provides data on race and education levels. We matched 97 percent of the experimental sample to the tax assessor data and 89 percent to the census data.

Table 1 presents descriptive statistics by treatment and control group for the pre-treatment period. Columns (1)–(3) display mean values for households assigned to treatments 1–3. Column (4) displays means for the control group. Columns (5) and (6) show the *F*-statistic and *p*-value from a test of the null hypothesis that mean values are equal across treatment and control groups. With the exception of ownership status, for which the mean differences are less than one percentage point, there are

Table 2
Mean on summer 2006 and *F*-test within subgroups (in thousand of gallons).

Subgroup		(1) Treat 1	(2) Treat 2	(3) Treat 3	(4) Control	(5) *F*-statistic	(6) *p*-Value
Previous water use (June–November 2006)	Below Median	22.46	22.37	22.45	22.42	0.35	0.79
	Above Median	57.48	57.60	57.36	57.40	0.06	0.98
Previous water use (April–May 2007)	Below Median	27.91	27.24	27.32	27.51	2.04	0.11
	Above Median	53.52	53.79	53.55	53.50	0.09	0.96
Fair market value	Below Median	31.04	30.15	30.37	30.36	2.70	0.04
	Above Median	48.14	48.59	48.92	48.54	0.46	0.71
Ownership status	Renter	38.75	38.65	38.98	38.66	0.06	0.98
	Owner	39.75	39.42	39.71	39.61	0.24	0.87
Age of home	Below Median	41.79	42.05	42.35	41.86	0.42	0.73
	Above Median	37.35	36.42	36.83	36.99	1.30	0.27
% with higher degree[a]	Below Median	33.54	33.59	33.66	33.73	0.14	0.93
	Above Median	44.56	44.03	44.51	44.35	0.27	0.85
% white[a]	Below Median	33.22	33.25	33.19	33.45	0.35	0.79
	Above Median	44.88	44.36	44.84	44.60	0.29	0.83

Source: Experimental Data, 2007 Cobb County Tax Assessor Database, 2000 US Census.
[a] At block group level.

no statistically significant differences in pre-treatment variables (recall our sample is over 100,000 observations). These results support the claim that randomization was effective.

The treatments were not, however, randomized within the subgroups. Nevertheless, our sample size is large, our randomization was done within small neighborhood groups and our subgroup set is small. Thus we would expect that observable and unobservable characteristics that affect water use would be well balanced between treatment and control groups within subgroups. To provide evidence of this balance, we examine pre-treatment water use across the treatment and control groups within each subgroup (see Table 2). For example, we test (*F*-test) whether pre-treatment mean water uses across treatment and control groups are statistically indistinguishable from each other within the group of renter-occupied households, then within the group of owner-occupied households, then within the group of above-median fair market value households, then within the group of below-median fair market value households, etc. With sixteen sequential tests and Type I error rate set to 0.05, we would expect approximately one of them to reject the null hypothesis of no difference through chance alone at the $p < 0.05$ level. In no test is the null hypothesis rejected.

Table 3
Post-treatment water use descriptive statistics (in thousands of gallons).

Variable	(1) Technical advice (T1)	(2) Weak social norm (T2)	(3) Strong social norm (T3)	(4) Control	(5) Diff (1)–(4)	(6) Diff (2)–(4)	(7) Diff (3)–(4)
Post-treatment data							
Summer 2007[a]	36.35	35.39	34.87	36.40	−0.05	−1.00**	−1.53**
Summer 2008[a]	25.51	25.33	24.99	25.49	0.02	−0.17	−0.50*
Summer 2009[a]	27.78	27.38	27.18	27.42	0.36	−0.04	−0.24
Winter 07/08[b]	21.63	21.58	21.43	21.71	−0.08	−0.13	−0.28*
Winter 08/09[b]	21.83	21.57	21.63	21.79	0.04	−0.22	−0.16

Source: Experimental data.
[a] Summer season comprises July to October use.
[b] Winter season comprises December to March use.
* $p < 0.05$.
** $p < 0.01$.

Table 4
Test of Zero CATE and Constant CATE (covariates with higher order terms).

	Treatment 1		
	2007 (Zero CATE)	2008 (Zero CATE)	2009 (Zero CATE)
Top down selection of covariates	2	2	2
Bottom up selection of covariates	2	2	2
All covariates	2	2	2
% rejections of the null ($p < 0.05$)	100.0%	100.0%	100.0%
	Treatment 2		
	2007 (Constant CATE)	2008 (Zero CATE)	2009 (Zero CATE)
Top down selection of covariates	1	0	2
Bottom up selection of covariates	0	0	2
All covariates	0	0	2
% rejections of the null ($p < 0.05$)	16.7%	0.0%	100.0%
	Treatment 3		
	2007 (Constant CATE)	2008 (Constant CATE)	2009 (Constant CATE)
Top down selection of covariates	2	2	2
Bottom up selection of covariates	2	2	2
All covariates	2	2	2
% rejections of the null ($p < 0.05$)	100.0%	100.0%	100.0%

Note: These results are summarized from details in online Appendix 1. They represent the number of times that the null hypothesis is rejected. For Treatment 1, we evaluate Zero CATE (2007, 2008, 2009). For Treatment 2 we evaluate Constant CATE (2007) and Zero CATE (2008, 2009). For Treatment 3, we evaluate Constant CATE (2007, 2008, 2009). Online Appendices 2 and 3 show the results for the other two specifications (flexible dummy and continuous variables without higher order terms, respectively). See text for details.

Table 3 presents descriptive statistics by treatment and control group for post-treatment water use. This analysis replicates and complements the results reported in Ferraro and Price (2013) and Ferraro et al. (2011) with our slightly smaller sample. Columns (1)–(4) display mean values for treatment and control groups. Columns (5)–(7) display differences with respect to the control group and the statistical significance of these differences. In summer 2007, treated households consumed less water than the control group, with the difference statistically different from zero for Treatments 2 and 3.[9] In summer 2008, that difference remains significant only for Treatment 3. In summer 2009, none of the treatment effects is significantly different from zero.[10] In winter months, when most water use is indoor water use, only the effect of Treatment 3 in 2007/2008 is statistically significant.

5. Results

5.1. Nonparametric tests

Given the mean effects of Treatment 1 in all years and of Treatment 2 beyond 2007 were statistically indistinguishable from zero, we test whether the mean effects of these treatments are zero for all subgroups in these years (Zero CATE Test; H0: CATE = 0). For Treatment 2 in summer 2007 and Treatment 3 in all years, we test whether the treatment effect is constant (Constant CATE Test; H0: CATE = constant). Table 4 summarizes the results for the summer water seasons using the three

[9] Ferraro and Price also show the differences among treatments are statistically different, as is the trend when they are ordered in terms of water use as predicted by their theory (T3 < T2 < T1 < Control).

[10] Ferraro, Miranda and Price increase the statistical precision of these 2008–2009 estimates by estimating a regression model that includes controls for other covariates that contribute to the variability of water use and the randomization strata, and find an effect for treatment 3 in 2009 at the 5% level.

methods of covariate choice. Each panel shows the results for a specific treatment and each column shows the number of times that the null hypothesis is rejected for a given year ($p < 0.05$).

For Treatment 1, the null hypothesis of zero CATE for all subgroups is always rejected. These results imply that some subgroups may have an average effect different from zero, corroborating Ferraro and Price (2013) results: when excluding the top and bottom 0.25 percentile of the distribution, and controlling for previous water use in a regression framework, Treatment 1 had a small (370 gallons) yet statistically significant ($p < 0.05$) reduction in water use. However, when using other specifications (e.g., without higher-order terms or with flexible coding of the covariates as dummy variables), we never reject the null hypothesis. Thus we conclude that there is evidence that some subgroups of the experimental sample respond to Treatment 1, but we want to see whether the quantile regression analyses corroborate this conclusion before pursuing a parametric analysis with interaction terms.

For Treatment 2, the null hypothesis of Zero CATE in 2008 is never rejected, but it is always rejected for 2009 (yet like the results for Treatment 1, alternative specifications yield less consistent results). With respect to the null of Constant CATE in 2007, the null hypothesis is only rejected one time out of six tests, which we consider very weak evidence of heterogeneous treatment effects.

For Treatment 3, the null hypothesis of Constant CATE is always rejected (including with other specifications). Thus we conclude there is strong evidence of heterogeneous treatments effects for Treatment 3.[11]

5.2. *Quantile regressions*

The quantile graphs for each treatment over the three summer periods are depicted in Figs. 1–3. Each graph plots the average treatment effect (dashed line), the QTE (solid line), and the respective confidence intervals of these point estimates. For Treatment 1 (Fig. 1), most of the distribution lies near the zero effect line for all three years without substantial heterogeneity. For Treatment 2 (Fig. 2), heterogeneity is detected only in 2007 and only clearly apparent in the upper half of the distribution. For Treatment 3 (Fig. 3), there is clear evidence of substantial heterogeneity in 2007, particularly for the lower and upper parts of the distribution. Summer 2008 also shows heterogeneity in the upper part of the distribution, but not as much as in the previous year. The impacts in 2009 are more homogenous.

Thus the results of the quantile regressions are consistent with the nonparametric tests: strong evidence of heterogeneity in responses to Treatment 3, and weak or no evidence of such heterogeneity for the other two treatments. Together, the nonparametric tests (Section 5.1) and the quantile regressions imply the following result regarding heterogeneous treatment responses:

Result 1. There is strong evidence of heterogeneous responses for Treatment 3 (strong social norm) in 2007 and 2008. The evidence for heterogeneous responses for Treatment 3 in 2009 and for Treatments 1 and 2 in all years is weak or non-existent.

If one assumes rank preservation (see Section 3.3 for evidence) and interprets Figs. 1–3 as approximations of the distribution of treatment effects, they imply another finding:

Result 2. Assuming rank preservation, there is no evidence of an increase in water use as a result of receiving a treatment message anywhere along the distribution of users. Treatment messages either reduce water use or have no effect.

5.3. *Subgroup analysis*

Given the strong evidence in the previous two sections of heterogeneous treatment effects for Treatment 3 (strong social norm), and the weak evidence for heterogeneous treatment effects for the other two treatments, we focus the subgroup analysis on Treatment 3. Recall that in Section 4, we demonstrated that pre-treatment mean water use across treatment and control groups are

[11] We also ran all the tests using education and race measured at the census tract instead of the census block. For all tests and all treatments, the frequency of rejections does not change (results available upon request).

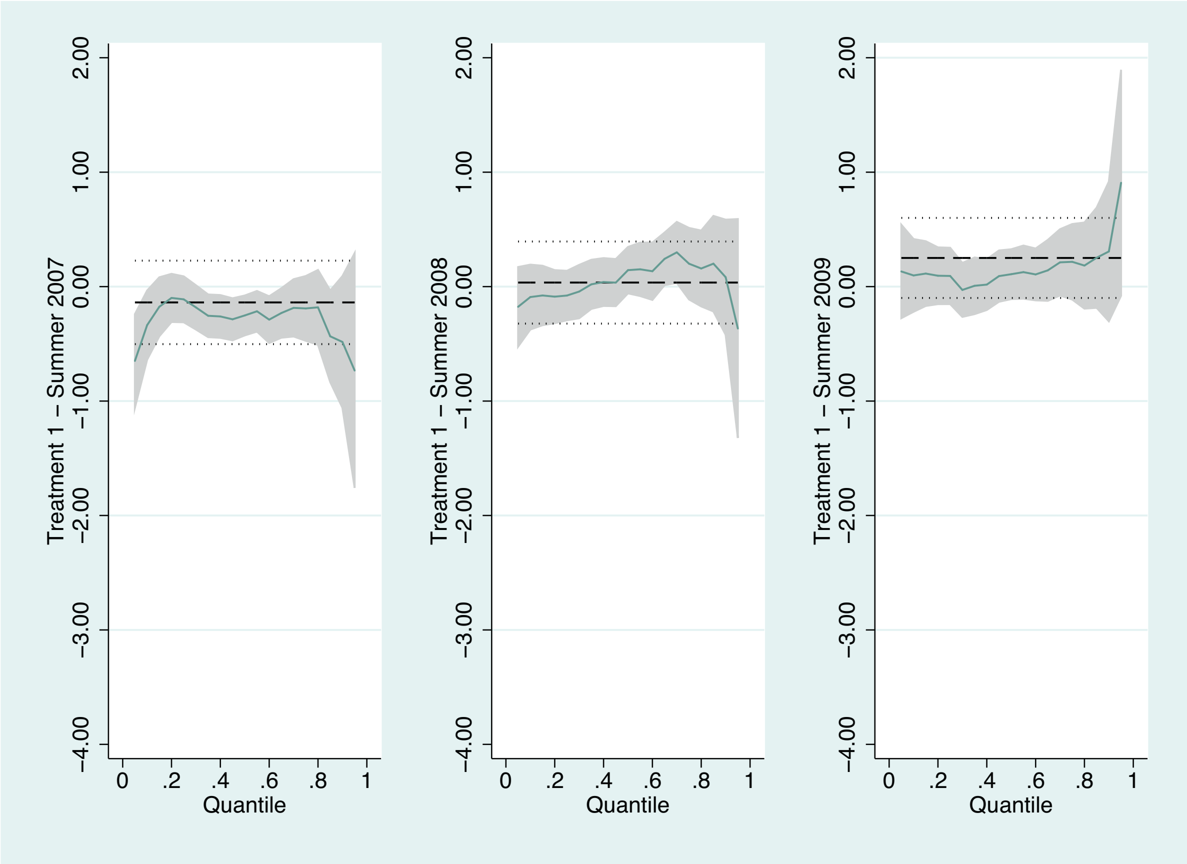

Fig. 1. Quantile treatment effects for Treatment 1: pure information (summer 2007, summer 2008, summer 2009). *Note*: Graphs plot quantile estimates and mean treatment effects across water use distribution. Dashed line depicts the average treatment effect in a linear regression (OLS) framework, while dotted line represents its 95% confidence interval. Solid line depicts the quantile treatment effect and the shadowed area represents its 95% confidence interval. Online Appendix 10 shows rank preservation test results.

statistically indistinguishable from each other within each subgroup. This result provides evidence that randomization was effective at balancing household characteristics that affect post-treatment water use across the treatment and control groups, permitting the analysis of subgroup treatment effect heterogeneity.

In regression models that include the treatments, the subgroups, and all interactions of the treatments and the subgroups, we reject the null hypothesis that the effect of Treatment 3 is the same in all subgroups for 2007 ($p < 0.001$). For 2008 and 2009, we cannot reject this null hypothesis ($p \approx 0.15$). As expected, we cannot reject the null hypothesis for any year for treatments 1 and 2. Thus we focus on estimating individual subgroup effects for Treatment 3 in 2007. For completeness, we present subgroup analyses for all treatments, but caution the reader against interpreting any point estimates as evidence of heterogeneous treatment effects for treatments 1 or 2.

Following Heckman et al. (2002), we simplify the presentation by running independent regressions for each subgroup covariate. Table 5 presents the results. In the lower panel are the p-values for a hypothesis test of no difference across subgroups within a treatment, unadjusted for repeated hypothesis testing. If we adjust each p-value using the very conservative multiple-hypothesis testing adjustment described in Section 3.4, we draw the same inferences. We also draw the same inferences from the full regression with interaction terms.

Result 3. For Treatment 3 (strong social norm), greater responses are observed in households that used more water in the past, live in more expensive homes and are occupied by owners rather than renters.

Table 5
Subgroup analysis for water consumption (summer 2007).

	Dependent variable: Summer 2007						
	(1) Previous water use (June–November 2006)	(2) Previous water use (April–May 2007)	(3) Fair market value	(4) Owners/renters[a]	(5) Age of home	(6) % White	(7) % with higher degree
Treatment 1 (technical advice)	−0.320 (0.171)	−0.136 (0.196)	0.171 (0.264)	0.399 (0.976)	0.121 (0.472)	−0.550 (0.342)	−0.551 (0.342)
Treatment 2 (weak social norm)	−0.444* (0.185)	−0.559** (0.197)	−0.732** (0.268)	−0.310 (0.777)	−0.651 (0.442)	−0.678 (0.352)	−0.925* (0.371)
Treatment 3 (strong social norm)	−0.653** (0.159)	−0.722** (0.198)	−0.772** (0.251)	0.248 (0.754)	−1.533** (0.405)	−1.421** (0.322)	−1.100** (0.340)
Subgroup var. (high = 1)	27.45** (0.200)	27.22** (0.208)	16.54** (0.212)	2.634** (0.341)	−4.275** (0.221)	9.694** (0.227)	8.642** (0.228)
Treat 1 × high subgroup var.	0.445 (0.559)	0.420 (0.589)	−0.415 (0.595)	−0.530 (1.027)	−0.294 (0.616)	0.894 (0.628)	0.876 (0.630)
Treat 2 × high subgroup var.	−0.774 (0.519)	−0.614 (0.541)	−0.383 (0.552)	−0.802 (0.835)	−0.773 (0.570)	−0.775 (0.587)	−0.296 (0.588)
Treat 3 × high subgroup var.	−1.994** (0.485)	−2.176** (0.503)	−1.406** (0.523)	−2.101** (0.808)	0.0419 (0.543)	−0.334 (0.555)	−0.887 (0.560)
Constant	23.019** (0.064)	23.874** (0.075)	28.110** (0.097)	34.170** (0.320)	38.506** (0.165)	31.282** (0.130)	31.816** (0.134)
Observations	102,887	102,887	102,871	102,869	102,461	94,833	94,833
R^2	0.22	0.25	0.08	0.01	0.01	0.03	0.02
p-Value equal impact T1	0.43	0.48	0.49	0.61	0.63	0.15	0.16
p-Value equal impact T2	0.14	0.26	0.49	0.34	0.18	0.19	0.61
p-Value equal impact T3	0.00	0.00	0.01	0.01	0.94	0.55	0.11

Note: All water consumption variables are in thousands of gallons.
Online Appendix 4 shows subgroup analysis results including all covariates together. Online Appendices 5 and 6 shows details for summer 2008 and summer 2009, for which the null hypotheses that the treatment effect is the same in all subgroups cannot be rejected.
Robust standard errors in parentheses.

[a] In the case of owners (=1)/renter (=0), interaction terms are for owner group rather than high group.

* $p < 0.05$.

** $p < 0.01$.

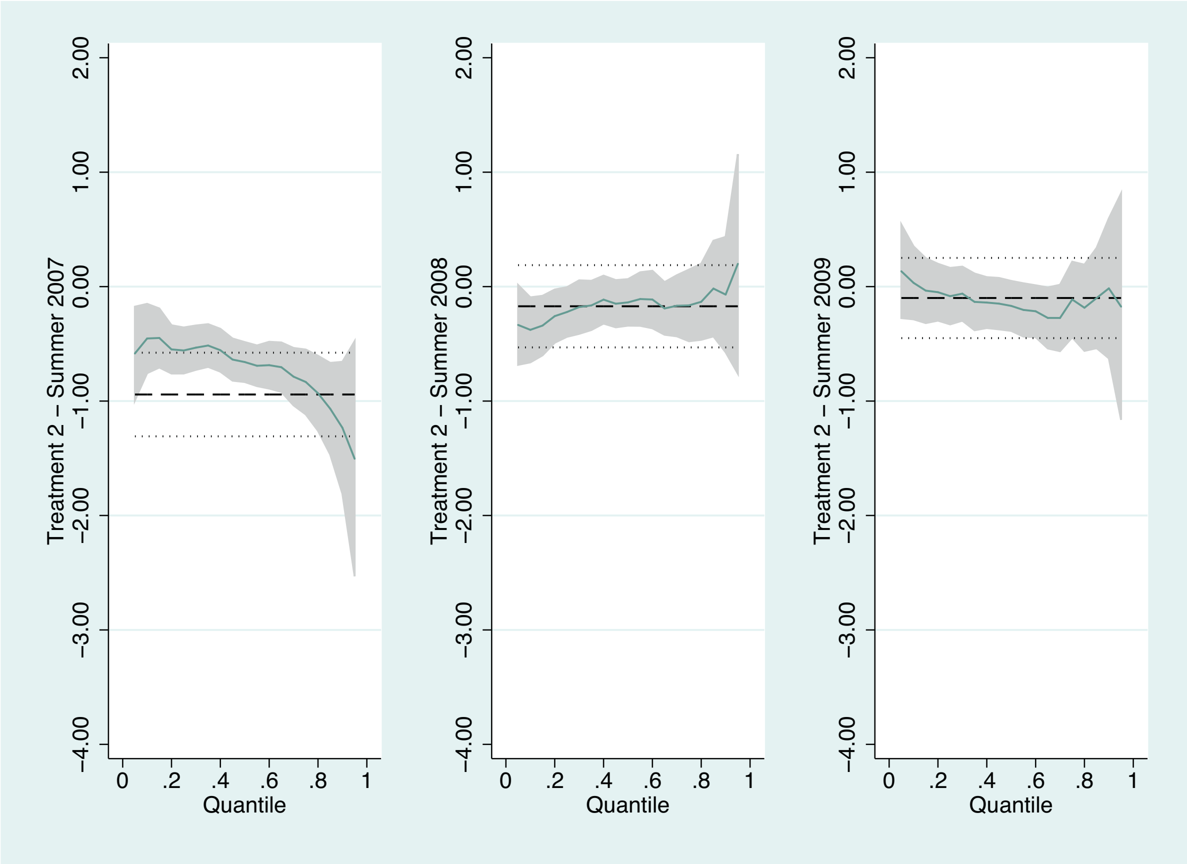

Fig. 2. Quantile treatment effects for Treatment 2: weak social norm (summer 2007, summer 2008, summer 2009). *Note*: Graphs plot quantile estimates and mean treatment effects across water use distribution. Dashed line depicts the average treatment effect in a linear regression (OLS) framework, while dotted line represents its 95% confidence interval. Solid line depicts the quantile treatment effect and the shadowed area represents its 95% confidence interval. Online Appendix 10 shows rank preservation test results.

We remind readers that the lack of strong evidence for heterogeneous responses to Treatments 1 and 2 (or to Treatment 3 in 2008 and 2009) from the combined analyses in the last three sections could arise from a lack of statistical power. The average treatment effects are small for these two treatments. Thus there may be heterogeneous responses, but, despite our large sample size, we cannot detect them. Whether such responses are likely to be policy-relevant is an open question.

5.4. *Targeting information campaigns*

Ferraro and Price (2013) show that Treatment 3 is the most cost-effective treatment at a cost of $0.58 per thousand gallons reduced in 2007. They then demonstrate that by targeting only households at or above the county median use for the previous summer, CCWS could obtain 88 percent of the original reduction for 65 percent of the original cost.[12] Could the information from Sections 5.1 to 5.3 be used to further improve targeting?

First, Ferraro and Price's analysis should be updated to reflect the treatment effects of the full 3-year period: 36% of the total treatment effect occurs 2008–2009. Thus Treatment 3 is actually far more cost-effective than their analysis implies: it costs $0.37 per thousand gallons reduced. If instead of targeting households based on their use during the previous year's summer, as in Ferraro and Price (2013), the utility were to target households based on their use in the two months before the campaign, it could

[12] Recall that approximately the bottom quintile of water consumers is not part of the experiment.

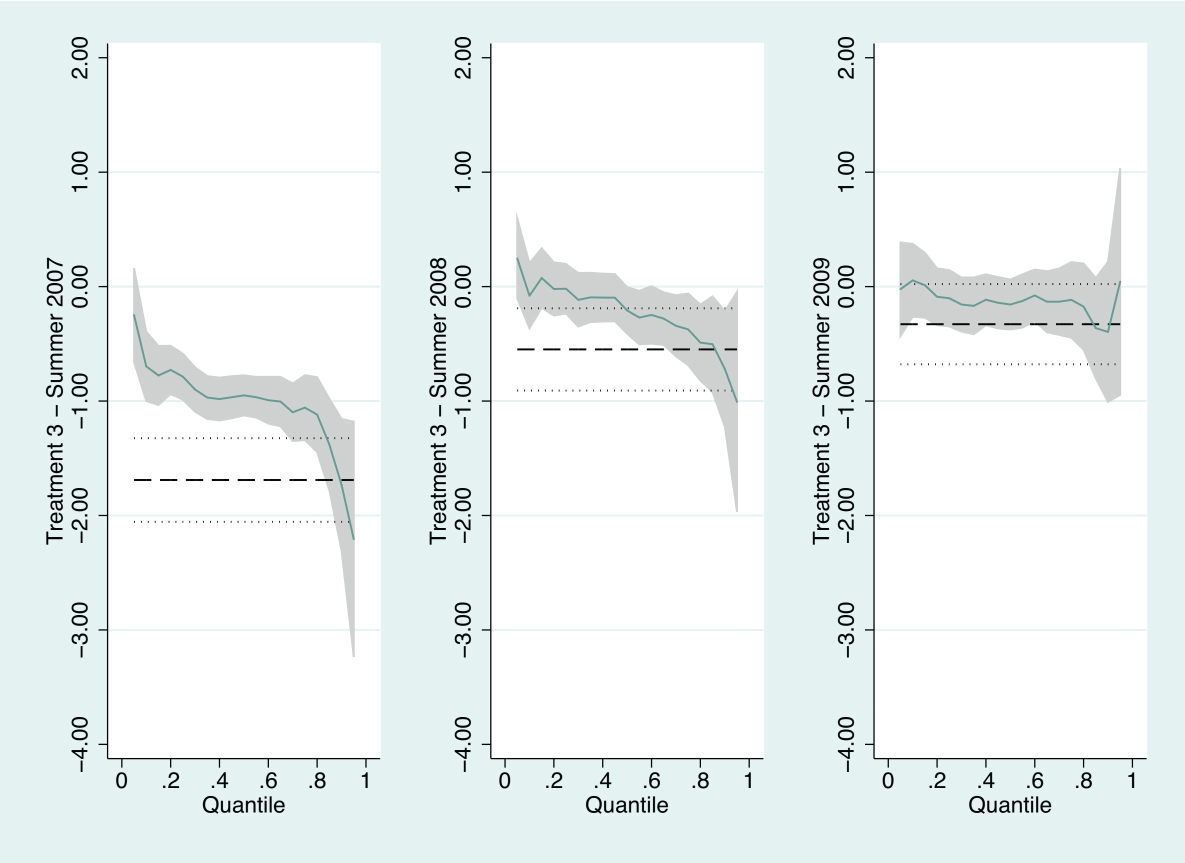

Fig. 3. Quantile treatment effects for Treatment 3: strong social norm (summer 2007, summer 2008, summer 2009). *Note*: Graphs plot quantile estimates and mean treatment effects across water use distribution. Dashed line depicts the average treatment effect in a linear regression (OLS) framework, while dotted line represents its 95% confidence interval. Solid line depicts the quantile treatment effect and the shadowed area represents its 95% confidence interval. Online Appendix 10 shows rank preservation test results.

obtain 80 percent of the original reduction for 48 percent of the original cost ($0.23 per thousand gallons reduced). If the utility were willing to sacrifice further reductions in use to achieve greater cost-effectiveness, combinations of targeting could further increase cost-effectiveness to $0.20 per thousand gallons reduced (e.g., target large water users who own their home). These possibilities suggest another result:

Result 4. By targeting on households identified as being more responsive to treatment, the water utility can reduce the overall program cost by over 50 percent with less than a 20 percent decline in the total number of gallons reduced.

Environmental programs and policies often cause controversy because they are seen as regressive: the relative burden of change is higher on poorer households. This problem is particularly acute for pecuniary approaches because the poor typically have more elastic demand. The heterogeneous response analysis suggests a potential solution in the context of the non-pecuniary Treatment 3 intervention. By targeting only wealthier households (i.e., high fair market value of the home), the utility could still achieve 60% of original reduction for about 48% of the original cost ($0.27 per thousand gallons reduced), and simultaneously avoid the politically charged controversy over equity.

Other targeting strategies are also possible (e.g., determining the conservation reductions possible if avoiding owner-occupied homes were deemed politically desirable). Combined with information on the benefits from reduced water use, heterogeneous treatment effect estimates could be used to determine an optimal targeting strategy.

5.5. *Mechanisms of the strong social norm message (Treatment 3)*

Given that Treatment 3 had the largest, most persistent effect and the strongest evidence of heterogeneous treatment effects, we focus the mechanism analysis on it. Ideally we would have designed the experiment explicitly to test mechanism hypotheses (Ludwig et al., 2011) or would have direct observations of the plausible mechanisms (e.g., the specific ways in which residents use water or technologies installed). Like most randomized controlled trials, however, we do not have these things. Thus we rely on a combination of theory (i.e., substantive assumptions about behavior) and data to eliminate rival mechanism explanations of the observed empirical patterns. Using this approach, we present evidence with regard to three mechanism hypotheses:

(i) The treatment effects are driven mainly by continuous, behavioral changes with recurring costs (e.g., watering outdoors less frequently or washing full loads of laundry or dishes) rather than one-time, behavioral or technological investments (e.g., fixing leaks, buying new appliances)
(ii) The treatment effects are driven mainly by changes in outdoor water use rather than changes in indoor water use; and,
(iii) The social comparison in the message affects behavior by highlighting social norms rather than by sending signals about privately efficient behavior (i.e., highlighting cost-savings opportunities).

We explain the motivation and policy relevance of each mechanism hypothesis below.

5.5.1. *Recurring behavioral changes versus one-shot technological investments*

The first hypothesis is relevant for understanding the long-term effects of the program. This hypothesis can be viewed as asking whether the home is treated or the home dwellers are treated. If the home were treated (e.g., a leak fixed; an efficient irrigation system installed), one would expect on-site treatment effects to persist after the current inhabitants depart. To examine this hypothesis, we use three pieces of information.

First, if households reduce water use mainly through one-shot investments (e.g., fix leaks, install low-flow toilets), one would predict relatively constant treatment effects across years within seasons. Yet, as indicated in Table 6, the effects wane over time within season (the same waning occurs also in

Table 6
Linear regressions of water seasons (with meter route fixed effects).

	(1) Summer 2007	(2) Winter 07/08	(3) Summer 2008	(4) Winter 08/09	(5) Summer 2009
Treatment 1 (technical advice)	−0.237 (0.189)	−0.121 (0.121)	−0.0702 (0.166)	0.0335 (0.189)	0.241 (0.169)
Treatment 2 (weak social norm)	−0.991** (0.171)	−0.108 (0.122)	−0.190 (0.184)	−0.239 (0.181)	−0.0604 (0.167)
Treatment 3 (strong social norm)	−1.741** (0.166)	−0.359** (0.129)	−0.637** (0.161)	−0.223 (0.181)	−0.344* (0.162)
Water use from June–November 2006	0.347** (0.013)	0.0258** (0.005)	0.127** (0.010)	0.038** (0.0091)	0.170** (0.011)
Water Use in Apr and May 2007	0.829** (0.045)	0.335** (0.017)	0.414** (0.025)	0.237** (0.016)	0.427** (0.0251)
Constant	1.874 (1.595)	16.207** (0.736)	7.086** (0.841)	15.788** (0.720)	14.140** (1.156)
Observations	106,669	106,669	106,669	106,669	106,669
R^2	0.63	0.13	0.25	0.02	0.33

Note: All water consumption variables are in thousands of gallons. Winter season runs from December to March billing. Summer season runs from July to October billing.
Robust standard errors in parentheses.
* $p < 0.05$.
** $p < 0.01$.

Table 7
Linear regressions: movers and non-movers (with meter route fixed effects).

	(1)	(2)	(3)	(4)
	Summer 2007		Summer 2008	
	Mover[a]	Non-mover	Mover[a]	Non-mover
Treatment 1 (pure information)	−1.938 (1.058)	−0.127 (0.192)	−1.367 (0.968)	−0.008 (0.169)
Treatment 2 (weak social norm)	−1.329 (1.051)	−0.970** (0.174)	−0.357 (1.104)	−0.175 (0.187)
Treatment 3 (strong social norm)	−1.931* (0.959)	−1.695** (0.169)	0.826 (0.999)	−0.671** (0.163)
Water use from June to November 2006	0.226** (0.0293)	0.352** (0.0133)	0.124** (0.0322)	0.126** (0.0102)
Water use in April and May 2007	0.985** (0.0872)	0.817** (0.0462)	0.133* (0.0587)	0.423** (0.0259)
Constant	−12.935* (5.905)	2.679 (1.622)	6.177* (2.424)	7.292** (0.861)
Observations	3667	102,811	3667	102,811
R^2	0.53	0.64	0.22	0.25

Note: All water consumption variables are in thousands of gallons.
Robust standard errors in parentheses

[a] New residents between December 2007–September 2008.
* $p < 0.05$.
** $p < 0.01$.

spring months). Nevertheless, inter-year variations in other factors could also explain this observed pattern.

Second, if households were to reduce water use mainly through one-shot, fixed-cost investments, one would expect such investments to be more likely in older houses where such investments are more cost-effective (e.g., they are more likely to have leaking pipes and older appliances; see Section 3.1). Table 5, however, shows that there is no difference in the responses between older and newer houses.

The third and final test exploits the re-framing of the first hypothesis in terms of asking whether the home or home dweller is treated. If one-shot, fixed-cost investments were driving reductions, the treatment effects should not disappear when the message recipients move out of their homes. We define "movers" as those households where the customer identification number changed between December 2007 and September 2008. Table 7 shows that, in summer 2007, movers (who had not yet moved) and non-movers reacted similarly to Treatment 3. In fact movers reduced a bit more than non-movers: 1900 versus 1700 gallons (the difference is not statistically different from zero in a pooled regression model). In summer 2008, however, Treatment 3 had a *positive*, but statistically insignificant, effect on households in which the message recipients had moved out (the difference between movers and nonmovers in a pooled regression model is statistically significant at $p = 0.06$).

Although none of these three observations by itself necessarily excludes one mechanism or the other, when combined they suggest a fifth result:

Result 5. The evidence is consistent with the hypothesis that the effects of the strong social norm message (Treatment 3) are driven mainly by behavioral changes with recurring costs rather than fixed-cost investments in technology.

5.5.2. Outdoor versus indoor water use changes

The second hypothesis is relevant to understanding the environmental effects of the treatment. Most of the indoor water used in Cobb County returns to the surface water system from which it was drawn. Because of processes like evapotranspiration and infiltration, most of the outdoor water used does not return on a time scale relevant for stream flow. Thus the environmental benefits from conservation are much larger for outdoor water use. To examine this hypothesis, one would ideally be able to observe outdoor and indoor water use separately, but Cobb County does not measure these uses separately. So we must depend on two observations.

First, theory and previous empirical evidence (e.g., Mansur and Olmstead, 2012) show that outdoor water demand is more price elastic than indoor demand. A plausible prediction is that households responding to the strong social norm message would likely first look to reduce water from outdoor

Table 8
Linear regressions of July 2007, December 2007 and July 2008 (with meter route fixed effects).

	(1) July 2007	(2) December 2007	(3) July 2008
Treatment 1 (technical advice)	−0.054 (0.0618)	−0.050 (0.045)	−0.067 (0.062)
Treatment 2 (weak social norm)	−0.334** (0.0618)	−0.057 (0.0453)	−0.054 (0.062)
Treatment 3 (strong social norm)	−0.548** (0.0618)	−0.087 (0.0453)	−0.220** (0.0623)
Water use from June to November 2006	0.103** (0.001)	0.008** (0.001)	0.036** (0.001)
Water use in April and May 2007	0.228** (0.002)	0.092** (0.001)	0.119** (0.002)
Constant	−0.899* (0.386)	−1.215** (0.283)	3.481** (0.389)
Observations	106,669	106,669	106,669
R^2	0.56	0.15	0.19

Standard errors in parentheses.
* $p < 0.05$.
** $p < 0.01$.

use, just as they would respond to a price increase. Moreover, empirical evidence from psychology and behavioral economics suggests that norm-based interventions are more effective when the behavioral responses are visible to others (e.g., Yoeli, 2009). This evidence would imply that households would be more likely to reduce outdoor water use because such responses are more visible to their neighbors.[13]

Second, treatment effects are largest in the months in which outdoor watering is typically observed. The lowest consumption in Cobb County occurs in December, during the winter when, water utility employees say, most use is indoor use. The highest consumption occurs in July, during the summer when most outdoor watering occurs. In 2006, before the experiment was implemented, the average December consumption in our sample was 6007 gallons and the average July consumption was 11,470, a 90 percent increase. We thus believe a contrast of treatment effects in December and July captures differences in indoor versus outdoor use.

Table 8 presents estimates of the average treatment effects for July 2007, December 2007, and July 2008 from regressions that include the strata in which the randomization was conducted (meter routes) and pre-treatment water use variables. In July 2007 and July 2008, Treatment 3 has large and statistically significant effects on water use, and the effect in July 2008 is half of the effect in July 2007. In December 2007, however, the effect of Treatment 3 is small and statistically insignificant. Comparing coefficients across regressions, the treatment effects in July 2007 or July 2008 versus December 2007, the differences are statistically significant ($p < 0.01$ and $p < 0.04$, respectively).

To argue that these results do not imply most of the treatment effect comes through outdoor watering, one would have to argue that only the effect on indoor water use wanes (thus the July 2008 effect is the same as the outdoor effect in July 2007). Such an argument contradicts the data. First, it is hard to imagine how behavioral changes indoors (rather than technological changes) could have accounted for half the water reduction observed in July 2007 when about 60 percent of indoor water use is for toilets, washing and bathing.[14] Second, the treatment effect from July 2009 is small and insignificantly different from zero (−0.081, $p = 0.23$) and significantly different from the treatment effect in July 2008 ($p = 0.06$). Thus one would have to argue that waning in the outdoor treatment effects only started after July 2008, which seems implausible. We therefore believe that the evidence supports a sixth result:

Result 6. The evidence is consistent with the hypothesis that the effects of the strong social norm message (Treatment 3) are driven mainly by changes in outdoor water use.

5.5.3. Social versus private preferences

The third hypothesis has not, to our knowledge, been raised in the literature on social comparisons. Rather than working through social preferences, the social comparison may work simply by conveying

[13] We thank a reviewer for pointing this out.
[14] http://www.epa.gov/WaterSense/pubs/indoor.html.

costly information about private costs and benefits. In an incomplete-information world with costly information acquisition or boundedly rational agents, households may not be optimizing their water use. The lack of a treatment effect from the information-only treatment (Treatment 1) suggests that households already know how they can reduce water use. They may assume, however, that adopting (or disadopting) these practices would not be utility-maximizing given their beliefs about costs and benefits. Yet when confronted with information about others' water use, they may update their beliefs (e.g., "I didn't know there could be gains from adopting these tips until I saw how my use compared to others' use.").

Thus the "social" comparison may actually be a "private" signal. Rather than harnessing pro-social preferences, the comparison helps self-interested, utility-maximizing agents get closer to the privately optimal water use pattern under complete information. We focus on the social comparison, rather than Treatment 3 in its entirety, because (1) Ferraro and Price (2013) showed that the tip sheet had no detectable effect and that there were statistically significant differences across treatments, and (2) Treatment 2 could only have affected behavior through social preferences. Thus we can conclude that *some* of the effect from Treatment 3 arose from social preferences. The question that remains is "Through what mechanism did the addition of the social comparison reduce water use further?" Answering this question is relevant to unexplored topic of the welfare effects of social comparison-based interventions: if they operate through self-interested preferences rather than pro-social preferences, they are likely welfare-enhancing. If they operate through pro-social preferences, they could be welfare-reducing (e.g., if social comparisons instill guilt, which lowers utility).

To examine this hypothesis, we use three pieces of information. First, we take advantage of Cobb County's increasing block price structure and test for the presence of private preference-based mechanisms. The price per additional gallon of water use increased at two thresholds: 9000 gallons/month (from $2.21 to $2.55/1000 gallons) and 16,000 gallons/month (to $2.88/1000 gallons). If private preferences motivate treatment responses, households that were using just above these threshold limits in the pre-treatment summer period should, on average, be more likely to reduce their water use post-treatment than households just below the threshold limits (because the expected cost savings are higher for households above the threshold).

Using bandwidths of 500, 1000, 2000, 3000 and 4000 gallons around the threshold, we test whether households just above the threshold respond more to the message than households just below the threshold within the bandwidth. We define "above the threshold" in two ways: as a dummy variable and as a continuous variable (the difference between water use in summer 2006 and the threshold). We estimate models both with and without the other subgroup-defining covariates. We also re-estimate all the models using only households who were consistently above or below the threshold every month in the previous summer (rather than above or below based on average monthly consumption during the summer). Thus we estimate 52 regression models and summarize the results in Table 9. In only three of them could we reject the null hypothesis of no difference between those above and below the threshold within the bandwidth, and in these cases the sign of the estimated coefficient was inconsistent with the hypothesis (in fact, for 34 of the 52 cases, the estimated coefficient was inconsistent with the hypothesis).

Second, we combine the results from Sections 5.3, 5.5.1 and 5.5.2 with assumptions about heterogeneous responses when private preferences motivate water use reductions. In Section 5.5.1, we presented evidence consistent with the hypothesis that households respond to the treatment with behavioral changes that have recurring costs. If this hypothesis were true and water reductions were driven by private preferences, renters and owners should be equally likely to decrease water use to save money because savings are immediate and not capitalized into the value of the house, as they might be with one-shot, technology investments. In Section 5.5.2, we presented evidence consistent with the hypothesis that households responded to the treatment with reductions mainly in outdoor water use. Such reductions could damage landscaping investments (and perhaps property values) and in a private preference framework, one would thus predict owners to respond less to treatment. In other words, if private preferences are driving decisions, we predict that owners should not respond more to treatment. In contrast to this prediction, however, we observed in Section 5.3 that owners are more responsive to the treatment (even after conditioning on all the other covariates).

Table 9
Block price threshold regressions by treatment.

	(1) 9000	(2) 16,000	(3) 9000	(4) 16,000	(5) 9000	(6) 16,000	(7) 9000	(8) 16,000	(9) 9000	(10) 16,000
	bw = 500		bw = 1000		bw = 2000		bw = 3000		bw = 4000	
Measure 1: dummy above threshold										
Treat 1 × above bandwidth	0.878	−1.566	1.030	3.413	−0.0295	2.457	0.177	0.709	0.094	−0.095
	(1.182)	(3.691)	(0.835)	(3.523)	(0.597)	(2.200)	(0.511)	(1.786)	(0.460)	(1.333)
Treat 2 × above bandwidth	1.849	−4.285	1.215	−2.140	−0.227	−0.930	−0.283	−0.819	−0.523	−0.631
	(1.368)	(3.440)	(0.905)	(2.445)	(0.605)	(1.827)	(0.521)	(1.472)	(0.474)	(1.264)
Treat 3 × above bandwidth	2.409*	−4.336	1.345	−0.843	0.206	0.182	−0.276	1.582	−0.387	1.841
	(1.084)	(3.320)	(0.841)	(2.437)	(0.634)	(1.779)	(0.538)	(1.529)	(0.486)	(1.195)
Other covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Measure 2: continuous variable, difference with respect to threshold										
Treat 1 × dummy above BW × Diff.	−0.356	7.621	−0.281	3.612	−0.595*	−1.968*	−0.230	−0.413	−0.161	0.108
	(2.149)	(6.848)	(0.761)	(3.425)	(0.258)	(0.815)	(0.149)	(0.476)	(0.096)	(0.150)
Treat 2 × dummy above BW × Diff.	−0.302	−1.061	−0.840	2.567	−0.242	0.0794	−0.193	0.172	−0.157	0.0897
	(2.695)	(5.581)	(0.846)	(1.936)	(0.274)	(0.845)	(0.150)	(0.401)	(0.103)	(0.144)
Treat 3 × dummy above BW × Diff.	−0.633	0.305	−0.430	1.499	−0.587*	0.307	−0.234	1.261**	−0.143	−0.107
	(2.046)	(6.214)	(0.767)	(2.184)	(0.256)	(0.778)	(0.150)	(0.446)	(0.098)	(0.139)
Other covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Measure 3: dummy above threshold for each month (rather than aggregated)										
Treat 1 × above bandwidth	n.a.	n.a.	n.a.	n.a.	5.120	–	−0.589	−2.066	1.905	−2.031
	n.a.	n.a.	n.a.	n.a.	(3.164)	–	(1.567)	(6.462)	(1.438)	(6.307)
Treat 2 × above bandwidth	n.a.	n.a.	n.a.	n.a.	5.809	−3.616	1.307	−1.315	1.255	9.506
	n.a.	n.a.	n.a.	n.a.	(4.157)	(12.50)	(1.606)	(6.180)	(1.206)	(11.39)
Treat 3 × above bandwidth	n.a.	n.a.	n.a.	n.a.	2.483	–	2.134	−2.212	0.870	−7.241
	n.a.	n.a.	n.a.	n.a.	(3.216)	–	(1.676)	(11.00)	(1.271)	(6.219)
Other covariates					Yes	Yes	Yes	Yes	Yes	Yes

Note: These results are summarized from details in online Appendices 7–9. Other covariates: fair market value, age of home, ownership status, % white, % with higher degree and each term of the interaction covariates. Results for Measure 3 with bandwidth of 500 and 1000 gallons are not available due to small samples.
Robust standard errors in parentheses.

* $p < 0.05$.
** $p < 0.01$.

Table 10
Social norms or signal of privately optimal behavior (with meter route fixed effects).

	Summer 2007
Treatment 1 (pure information)	−0.0282 (0.320)
Treatment 2 (weak social norm)	−0.898** (0.265)
Treatment 3 (strong social norm)	−2.127** (0.264)
High proportion renters (=1 if >median)	0.113 (0.282)
High proportion renters × Treat 1	−0.219 (0.396)
High proportion renters × Treat 2	−0.144 (0.356)
High proportion renters × Treat 3	0.888* (0.351)
Water use from June to November 2006	0.335** (0.014)
Water use in April and May 2007	0.823** (0.052)
Ownership status	0.701** (0.189)
Fair market value	1.79e−05** (3.12e−06)
Age of home	0.027* (0.011)
Constant	−2.422 (1.799)
Observations	95,233
R^2	0.64

Note: All water consumption variables are in thousands of gallons.
Robust standard errors in parentheses
* $p < 0.05$.
** $p < 0.01$.

Third, we assert that if private preferences for cost savings were driving reductions in water use, then after holding owner/renter status constant, a change in the percentage of renters in the neighborhood (census block group) should not influence the response of a household to Treatment 3. The political science and economics literature cited in Section 3.1, however, suggests this percentage may affect a household for whom the social comparison is working through social preferences because neighborhoods with greater ownership rates have greater social connections. Thus the norm created by the social comparison to neighbors is more relevant or salient for them. Moreover, we would not expect any such interaction between the proportion of households renting in a neighborhood and Treatment 1 or Treatment 2 because there is no social comparison in these treatments. Thus keeping these treatments in the model helps protect against spurious rejections of the null hypothesis because of other factors that may be correlated with an increase in renters in a neighborhood. We further protect against such rejections by adding previous water use, home and homeowner characteristics and 390 neighborhood fixed effects (meter route strata) to the model.

We define high-rental neighborhoods as those who have a percentage of renters greater than the median. We then create interaction terms with a dummy variable for above-median proportion of renters and the treatment dummy variables. Results are shown in Table 10. The interaction term is positive and significantly different from zero for only Treatment 3. Holding ownership status, home characteristics, previous water use and other neighborhood characteristics constant, households who receive Treatment 3 are more responsive in census block groups with low percentages of renters.

Although none of these three observations by itself necessarily excludes one mechanism or the other, when combined they suggest a final result:[15]

Result 7. The evidence is consistent with the hypothesis that the social comparison induces greater water use reductions by highlighting social norms rather than by sending signals about privately efficient behavior.

[15] Additional evidence comes from a recent working paper in political science that uses the same experimental sample merged with voting behavior and presents results consistent with the hypothesis that internalized pro-social preferences promote action for the public good across behavioral contexts (Bolsen et al., 2012).

6. Conclusions

Non-pecuniary, information-based strategies are increasingly being used to influence individual decision-making to achieve policy objectives. Despite their increasing popularity, their conceptual and empirical foundations remain under-researched. Although a growing number of scholars are conducting randomized experiments to test information-based strategies, many only report average treatment effects, thereby ignoring variation in the treatment effects. Moreover, none have attempted to elucidate the mechanisms through which these strategies operate.

In experimental studies in which treatments are not randomized within subgroups, or in any observational study, one must be cautious when estimating heterogeneous treatment effects. Unlike many studies of heterogeneous treatment effects in social experiments, we reduce the risk of mislabeling spurious correlations as heterogeneous treatment effects by combining complementary empirical approaches in an experiment with a large sample size and randomization conducted within small neighborhood strata. These attributes afford us better statistical power and experimental control in estimating heterogeneous treatment effects across observable subgroups in the population.

In a study of an information-based environmental program to induce voluntary reductions in the use of a common pool resource, we find strong evidence of heterogeneous treatment effects for a message that augments pure technical information with pro-social language and social comparisons (strong social norm message). In contrast, we were unable to detect strong evidence of heterogeneous treatment effects from pure information alone or pure information with pro-social language but no social comparison (i.e., traditional conservation messages).

Furthermore, the results from a subgroup analysis and quantile regressions imply (assuming rank preservation is a good approximation; see Section 3.3) that the information treatments either reduce or have no effect on water use; they never increase it for any subgroups.[16] This result has interesting implications for the "boomerang" effect, a heterogeneous response predicted by social psychologists whereby low resource users discover through the social comparison that they are low users and, in response, increase their use (e.g., Schultz et al., 2007). To avoid this predicted response, OPower's norm-based, energy conservation program (Section 2) supplements its social comparisons with emoticons: if a household's energy use is below the average of its comparison group, it receives a green "smiley face" that is assumed to prevent the boomerang effect. Despite the absence of emoticons in our experiment, we find no evidence of a boomerang effect (perhaps because the negative framing of the household's percentile in this treatment acts as an injunctive norm).

With regard to mechanisms, the evidence suggests that the strong social norm message operates through behavioral changes with recurring variable costs rather than one-shot, fixed-cost investments (which makes the message's effects more likely to wane over time), and through changes in outdoor rather than indoor watering (which increases the environmental benefits of the messages). We also explore a third mechanism hypothesis that posits a rival explanation of how social comparisons affect behavior: rather than operating through social preferences, they may convey costly information about privately efficient behavior to households with incomplete information. The evidence, however, is inconsistent with this rival explanation. Social comparisons do seem to work through social preferences, which would have important implications for future attempts to estimate the full welfare implications of information-based policies and programs (e.g., if the messages work by inducing guilt).

Our study has at least three policy implications. The first relates to the external validity of the CCWS experiment: sites with poor households, many renters, or low water use may not see as large of an impact as Cobb County did from an information campaign that augments information and pro-social language with social comparisons.

Second, making information about heterogeneous treatment effects available to decision makers can greatly improve program cost-effectiveness. We demonstrated that with improved targeting based on observable household characteristics, the overall costs of the program could be reduced by over 50

[16] Earlier versions of Costa and Kahn, 2013, for example, claimed Republicans increased their electricity consumption after receiving a norm-based message.

percent with less than a 20 percent decline in the aggregate impact. Targeting can also improve the equity of outcomes or achieve other political objectives.

Third, as suggested by Ferraro and Price (2013), pecuniary and norm-based, non-pecuniary policies may be complementary. The strong social norm message had an immediate effect on water use in the month after the message was sent and high-income households are most responsive to the message. Thus in contrast to water conservation programs that use pecuniary incentives, for which average responses are slow and high-income households are least responsive (e.g., Mansur and Olmstead, 2012), programs based on norm-based incentives work quickly and are most effective among high-income households. Moreover, price changes are typically expected to lead to a persistent change in the quantity demanded, whereas the evidence from this experiment suggests the effects of norm-based approaches wane over time. Thus the two approaches may be preferred in different contexts (e.g., a need to change short-term demand rather than long-term demand) and may be complementary when combined. Future experiments should directly test the hypothesis of complementarity between the two approaches by randomly assigning pecuniary and non-pecuniary incentives, in isolation and in combination (and, if possible, randomizing within subgroups of interest).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.reseneeco.2013.04.001.

References

Angrist, J., Pischke, J.-S., 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, Princeton, NJ.

Allcott, H., 2012. Social norms and energy conservation. Journal of Public Economics 95 (October (9–10)), 1982–1995.

Allcott, H., Mullainathan, S., 2010a. Behavioral and energy policy. Science 327 (March (5970)), 1204–1205.

Allcott, H., Mullainathan, S., 2010. External Validity and Partner Selection Bias, NBER Working Paper 18373.

Angrist, J., 2004. Treatment effect heterogeneity in theory and practice. Economic Journal 114 (March), C52–C83.

Ayres, I., Raseman, S., Shih, A., 2009. Evidence from Two Large Field Experiments that Peer Comparison Feedback can Reduce Residential Energy Usage, NBER Working Paper #15386.

Bennear, L., Olmstead, S., 2008. The impacts of the "right to know": information disclosure and the violation of drinking water standards. Journal of Environmental Economics and Management 56, 117–130.

Bitler, M., Gelbach, J., Hoynes, H., 2005. Distributional Impacts of the Self-sufficiency Project, NBER Working Paper #1626.

Bitler, M., Gelbach, J., Hoynes, H., 2006. What mean impacts miss: distributional effects of welfare reform experiments. American Economic Review 96 (4), 988–1012.

Bitler, M., Gelbach, J., Hoynes, H., 2008. Distributional impacts of the self-sufficiency project. Journal of Public Economics 92, 748–765.

Bolsen, T., Ferraro, P., Miranda, J.J., 2012. Are Voters More Likely to Contribute to Other Public Goods? Evidence from a large-scale randomized policy experiment, Working Paper.

Bjørner, T.B., Hansen, L.G., Russell, C.S., 2004. Environmental labeling and consumers' choice: an empirical analysis of the effect of the Nordic Swan. Journal of Environmental Economics and Management 47 (3), 411–434.

Bui, L., Mayer, C., 2003. Regulation and capitalization of environmental amenities: evidence from the toxic release inventory in Massachusetts. The Review of Economics and Statistics 85 (3), 693–708.

Costa, D., Kahn, M., 2013. Energy Conservation "Nudges" and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment. Journal of the European Economic Association 11 (3).

Crump, R., Hotz, J., Imbens, G., Mitnik, O., 2008. Nonparametric tests for treatment effect heterogeneity. The Review of Economics and Statistics 90 (August (3)), 389–405.

Davis, L., 2012. Evaluating the slow adoption of energy efficient investments: are renters less likely to have energy efficient appliances? In: Fullerton, D., Wolfram, C. (Eds.), The Design and Implementation of U.S. Climate Policy. University of Chicago Press, Chicago, IL.

Deaton, A., 2010. Instruments, randomization, and learning about development. Journal of Economic Literature 48 (2), 424–455.

DiPasquale, D., Glaeser, E., 1999. Incentives and social capital: are homeowners better citizens? Journal of Urban Economics 45, 354–384.

Djebbari, H., Smith, J., 2008. Heterogeneous impacts in PROGRESA. Journal of Econometrics 145, 64–80.

Duflo, E., Saez, E., 2003. The role of information and social interactions in retirement plan decisions: evidence from a randomized experiment. The Quarterly Journal of Economics 118 (3), 815–842.

Ferraro, P., Miranda, J.J., Price, M., 2011. The persistence of treatment effects with non-pecuniary policy instruments: evidence from a randomized environmental policy experiment. American Economic Review: Papers and Proceedings 101 (3), 318–322.

Ferraro, P.J., Price, M., 2013. Using Non-pecuniary Strategies to Influence Behavior: evidence from a large-scale field experiment. The Review of Economics and Statistics 95 (1), 64–73.

Firpo, S., 2007. Efficient semiparametric estimation of quantile treatment effects. Econometrica 75 (1), 259–276.

Fischer, C., 2008. Feedback on household electricity consumption: a tool for saving energy? Energy Efficiency 1 (1), 79–104.
Goldstein, N.J., Cialdini, R.B., Griskevicius, V., 2008. A room with a viewpoint: using social norms to motivate environmental conservation in hotels. Journal of Consumer Research 35, 472–482.
Greenberg, M., 2005. Concern about environmental pollution: how much difference do race and ethnicity make? A New Jersey case study. Environmental Health Perspectives 113 (4), 369–374.
Greenstone, M., Gayer, T., 2009. Quasi-experimental and experimental approaches to environmental management. Journal of Environmental Economics and Management 57 (1), 21–44.
Habyarimana, J., Jack, W., 2011. Heckle and chide: results of a randomized road safety intervention in Kenya. Journal of Public Economics 95, 1438–1446.
Heckman, J., Smith, J., Clements, N., 1997. Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts. Review of Economic Studies 64, 487–535.
Heckman, J., Heinrich, C., Smith, J., 2002. The performance of performance standards. Journal of Human Resources 37 (Autumn (4)), 778–811.
Hotz, J., Imbens, G., Mortimer, J., 2005. Predicting the efficacy of future training programs using past experiences at other locations. Journal of Econometrics 125, 241–270.
House of Lords, Science and Technology Select Committee, 2011. Behaviour Change, 2nd Report of Session 2010–12. The Stationary Office Limited, London.
Imai, K., Strauss, A., 2011. Estimation of heterogeneous treatment effects from randomized experiments, with application to the optimal planning of the get-out-the-vote campaign. Political Analysis 19 (Winter (1)), 1–19.
Imai, K., Ratkovic, M., 2013. Estimating Treatment Effect Heterogeneity in Randomized Program Evaluation. Annals of Applied Statistics 7 (1), 443–470.
Imbens, G., Wooldridge, J., 2009. Recent developments in the econometrics of program evaluation. Journal of Economic Literature 47 (11), 5–86.
Jin, G., Leslie, P., 2003. The effect of information on product quality: evidence from restaurant hygiene grade cards. Quarterly Journal of Economics 118 (2), 409–451.
Mansur, E., Olmstead, S., 2012. The Value of Scarce Water: Measuring the Inefficiency of Municipal Regulations. Journal of Urban Economics 71 (3), 332–346.
Lehmann, M.-C., 2010. Spatial Externalities of Social Programs: Why Do Cash Transfer Programs Affect Ineligible's Consumption? Paris School of Economics, Working Paper.
Ludwig, J., Kling, J., Mullainathan, S., 2011. Mechanism experiments and policy evaluations. Journal of Economic Perspectives 25 (3), 17–38.
Manski, C., 2004. Statistical treatment rules for heterogeneous populations. Econometrica 72 (4), 1221–1246.
Pawson, R., Tilley, N., 1997. Realistic Evaluation. Sage Publications. Thousand Oaks, CA.
Rohe, W., Van Zandt, S., McCarthy, G., 2001. The social benefits and costs of homeownership: a critical assessment of the research. Low-Income Homeownership Working Paper Series LIHO-01.12, Joint Center for Housing Studies, Harvard University.
Wesley, S.P., Nolan, J.M., Cialdini, R.B., Goldstein, N.J., Vladas, G., 2007. The constructive, destructive, and reconstructive power of social norms. Psychological Science 18, 429–434.
Shogren, J., Taylor, L., 2008. On Behavioral-Environmental Economics. Review of Environmental Economics and Policy 2, 26–44.
Smith, K., Desvousges, W., 1990. Risk communication and the value of information: radon as a case study. The Review of Economics and Statistics 72 (1), 137–142.
Smith, K., Desvousges, W., Johnson, R., Fisher, A., 1990. Can public information programs affect risk perceptions? Journal of Policy Analysis and Management 9 (1), 41–59.
Thaler, R., Sustein, C., 2008. Nudge. Improving Decisions About Health, Wealth and Happiness. Penguin Books, New York, NY.
Yoeli, E., 2009. Does Social Approval Stimulate Prosocial Behavior? Evidence from a Field Experiment in the Residential Electricity Market, University of Chicago, Working Paper.