Heterogeneous Treatment Effects and Mechanisms in Information-Based Environmental Policies: Evidence from a Large-Scale Field Experiment

Paul J. Ferraro
Department of Economics
Andrew Young School of Policy Studies
Georgia State University
PO Box 3992
Atlanta, GA 30302-3992
pferraro@gsu.edu

and

Juan José Miranda
Department of Economics
Andrew Young School of Policy Studies
Georgia State University
PO Box 3992
Atlanta, GA 30302-3992
jjmiranda@gsu.edu

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.

Heterogeneous Treatment Effects and Mechanisms in Information-Based Environmental Policies: Evidence from a Large-Scale Natural Field Experiment

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 evaluate heterogeneous household responses. Understanding such heterogeneity is important for improving the cost-effectiveness of non-pecuniary programs and extending them to other populations. We find little evidence of heterogeneous responses to purely technical information or traditional conservation messages, but strong evidence of heterogeneous responses to pro-social messages that highlight social norms: wealthier, owner-occupied households, and households that use more water are more responsive. In contrast, these subgroups tend to be least responsive to pecuniary incentives. Combining theory and data, we also shed light on the mechanisms through which the treatment effects arise: 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.

Keywords: program evaluation, experimental design, conditional average treatment effects, quantile average treatment effects, other-regarding preferences, social preferences.

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 Sunstein 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 Meyer 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 2010; 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; List and Price, forthcoming), to test the impacts of non-pecuniary, norm-based messages on environmental outcomes such as energy use (e.g. Ayres, Raseman and Shih 2009; Yoeli 2009; Allcott 2010; Costa and Kahn 2010) and water use (e.g. Ferraro and Price forthcoming; Ferraro, Miranda and Price 2011). These studies find that sending pro-social messages and social comparisons that contrast own consumption to peer-group consumption can reduce water and energy consumption. However, there is little evidence about (1) the characteristics of the households that are most responsive to these types of messages (studies estimate mean treatment effects rather than variations across subpopulations); and (2) the mechanisms through which the messages affect behavior.

Studying heterogeneous treatment effects yields policy and research-relevant insights (Heckman, Smith and Clements 1997; Angrist 2004; Djebbari and Smith 2008). For example, it can help policy makers more cost-effectively target the treatments to subgroups that are most responsive. It can also help strengthen the external validity of randomized controlled trials. The mean effects of the same experimental design could be different when applied in other populations with different distributions of observable characteristics (Hotz et al. 2005). Third, by combining theory and subgroup analysis, 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 beyond determining *whether* a treatment is effective to determining *why* it is effective.[1]

We study heterogeneity and mechanisms in the context of the large-scale field experiment. The experiment was run in 2007 in partnership with the Cobb County Water System in metropolitan Atlanta, Georgia, USA. To induce voluntary reductions in water use during a drought, three types of messages were sent, at random, to households. The three treatments comprise: (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 (forthcoming) report short-term average treatment effects and Ferraro, Miranda and Price (2011) extend the analysis to 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: wealthier households, owner-occupied households, and households that use more water are more responsive. Interestingly, these are among the households identified in the literature to be *least* responsive to pecuniary incentives. Also, in contrast to predictions from the psychology literature that low resource users may respond to the social comparison message by increasing their use (e.g., Schultz et

[1] Evidence of heterogeneous responses also informs future observational studies that may use instrumental variables to estimate treatment effects from information-based strategies. Heterogeneity implies that the estimates should be interpreted as local average treatment effects (LATE) rather than population average treatment effects (ATE).

al.), we find no evidence that low users or any other subgroups increase their water use, on average, in response to the social comparison message. 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 and main findings.

2. Experiments with Information-based Environmental Programs

A small number of experimental studies estimate the effects of pro-social and social comparison messages on environmental outcomes. Ayres, Raseman and Shih (2009), Allcott (forthcoming), Costa and Kahn (2010) and Yoeli (2009) focus on energy consumption, while Ferraro and Price and Ferraro, Miranda and Price 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 cause reductions in consumption.[2]

The studies by Ferraro and Price and Ferraro, Miranda and Price are described in section 4. Allcott evaluates a field experiment in Minnesota run by OPower, a firm that promotes energy efficiency for its utility partners. OPower sends home energy reports that include information on strategies to conserve energy, social comparisons of the household's consumption to consumption of geographical neighbors in homes of comparable size, and positive and negative emoticons to indicate the social desirability of the household's position in the distribution used to make the social comparison.[3] OPower's restriction of the social comparison to neighbors with comparable house sizes is intended to heighten the relevance of the social comparison, but it might also reduce the scope and impact of the comparison because consumption variability may be low within the comparison group. Allcott finds that these reports reduce energy consumption by a little over 2 percent. Ayres, Raseman and Shih

---

[2] 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.

[3] Specifically, treated households receive home energy reports containing: (i) personal use history; (ii) current period neighbor comparison; (iii) twelve-month neighbor comparison; and (iv) energy efficiency advice. For further details: http://www.opower.com/Approach/TargetedMessaging.aspx

evaluate two other OPower field experiments (in California and Washington) and find the reports reduce natural gas and electricity use by 1.2 and 2.1 percent, respectively. Yoeli examines a field experiment in California in which the decision to sign-up for a blackout prevention program was randomly varied to be private versus observable to one's neighbors. Participation is 3.6% higher in the publicly observable treatment, but the effect of the program on energy consumption is not reported.

Only three studies examine heterogeneous treatment responses. Allcott runs a quantile regression and Ferraro and Price examine how treatment responses vary as function of being below or above the median use. Both studies report that historically larger users appear to respond more, on average, to social comparison messages. Costa and Kahn use data from the California OPower experiment to test whether responses vary with political affiliation (obtained for a subsample from public voting records). Democratic households, on average, reduce their consumption, whereas Republican households, on average, do not. These three studies do not examine heterogeneity further, or probe the potential mechanisms through which treatment effects are generated.

3. Methodology

When exploring treatment effect heterogeneity, there is a substantial risk of finding statistically significant differences among subgroups when no true treatment effect heterogeneity exists because the subgroups are formed after the experiment is implemented (Imai and Strauss 2011). To mitigate this potential bias, we adopt five complementary methods 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; Heckman, Smith and Clements; Heckman, Hienrich and Smith, 2002). Finally, we isolate systematic variations among subgroups through interactions terms between the treatment variable and other covariates. We then use these estimates of heterogeneous subgroup responses, along with complementary analyses, to test hypotheses about the mechanisms through which treatment effects were achieved.

3.1 Heterogeneity in Treatment Responses

To estimate heterogeneous responses, we select covariates 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 two measures of previous water use, three household characteristics, and two neighborhood characteristics. Furthermore, in our final analysis that uses multiple hypothesis tests, we adjust the Type I error rate for sequential tests. Below, we present the covariates that define subgroups in the order we believe reflects their policy-relevance, and thus the order in which we will conduct the sequential tests.

Ferraro and Price's analysis shows that previous water use predicts future water use and provides suggestive evidence of heterogeneous treatment responses conditional on a household's percentile in summer 2006. 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 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 (2011) 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).[4]

Davis (2010) 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 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, DiPascuale and Glaeser (1999) also found that homeowners are 12% 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).

Another measure that reflects the scope and incentives for water conservation is the age of the home. Older homes, on average, have older water-intensive capital (e.g., toilets), which are more cost-effective to replace to achieve water conservation goals, and they are more likely to have repairable leaks.[5]

---

[4] 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.
[5] We do not explore heterogeneity conditional on lot size because lot size is highly correlated with fair market value.

Environmental preferences of household occupants would likely also affect their treatment responses. We cannot observe environmental preferences, but survey evidence suggests that environmental preferences vary with education levels and race (e.g. Greenberg, 2005). We (and water utilities) cannot observe education and race at the household-level, and so 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.

There are, of course, other covariates that may moderate treatment effects in our experiment, but which we do not measure (e.g., risk preferences). While it 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). Both tests are evaluated using all the subgroup covariates described in 3.1. 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 Ferraro and Price report 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 subgroups. This question can be evaluated with the Constant CATE test. Crump et al. prove that both tests using nonparametric regression functions based on series estimators 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*, while the alternative hypothesis is that the average effect for the subpopulation with covariate values *X* is different from zero for <u>some</u> 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 ($e(\beta_1 - \beta_0)$), 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*, while 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. (p.397), we select the final 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 test results using specifications with higher degree order terms of continuous variables to improve robustness (Crump et al.), but we also ran tests without higher-order terms and with flexible coding of the covariates as dummy variables and report differences when they exist (see Referee's Appendix for complete tables. Tables A1 – A3).[6]

3.3 Quantile Treatment Effects

The nonparametric tests described in 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, τ_q,

$$\tau_q = F_{Y(1)}^{-1}(q) - F_{Y(0)}^{-1}(q) \qquad (1)$$

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

$$\tilde{\tau}_q = F_{Y(1)-Y(0)}^{-1}(q). \qquad (2)$$

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, Smith and Clements; Djebbari and Smith 2008). Randomized experiments, however, only provide the marginal distribution of treated outcomes and the marginal distribution of untreated outcomes (which permit the estimation of the ATE).

---

[6] 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/

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. Thus, 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). In this case, the impact of the treatment on the distribution would be equivalent to the distribution of treatment effects.

Paraphrasing Angrist and Pischke (2009), if we discover, 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 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% 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).[7] 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 (see Referee's Appendix, table A4). We conclude there is support for viewing our QTE results as a useful approximation to the distribution of treatment effects, but only as an approximation.

[7] 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.

3.4 Subgroup Analysis

The nonparametric and quantile regression approaches described above do not specify which subgroups are most responsive to the treatments. We explore subgroup variation through interactions terms between the treatment variables and other covariates in a regression framework (Heckman, Smith and Clements; Heckman, Heinrich and Smith; Djebbari and Smith).

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). Then 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 will only be rejected if $p<0.0075$). As noted earlier, we are estimating causal effects conditional on observable characteristics. We are 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:

- Pure Information (Treatment 1): A 'tip sheet' listing different ways to most effectively reduce water use.
- 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% of the population). Each treatment consisted of roughly 11,700 houses and the control group consisted of roughly

71,800 houses. For more details about the treatments and experimental design see Ferraro and Price and Ferraro, Miranda and Price (see also Referee's Appendix for copies of messages). Ferraro and Price find that pure information (treatment 1) had no significant effect, while the weak social norm message (treatment 2) reduced water use by about 2.5%. In contrast, the strong social norm message reduced water use by almost 5%. Ferraro, Miranda and Price (2011) find that only the strong social norm message significantly affects water use 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% of the experimental sample to the tax assessor data and 89% to the census data.

Table 1 presents descriptive statistics by treatment and control group for the pre-treatment data. Columns (1)-(3) display mean values for households assigned to treatment 1 (pure information), treatment 2 (weak social norm), and treatment 3 (strong social norm). Column (4) displays means for the control group. Column (5) shows the F-statistic and column (6) its respective 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 no statistically significant differences in pre-treatment variables, including previous water use (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 identified in 3.1. Nevertheless, given that our sample size is large, our randomization was done within small neighborhood groups and our subgroup set is small, 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 Referee's Appendix, table A5). 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 2 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 and Ferraro, Miranda and Price with our slightly smaller sample size. 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. Households consumed less water than the control group in summer 2007, with the difference statistically different from zero for treatment 2 (weak social norm) and treatment 3 (strong social norm).[8] In summer 2008, that difference remains significant only for treatment 3. In summer 2009, none of the treatment effects is significantly different from zero.[9] 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 treatment 1 had no effect in any year and treatment 2 had no effect beyond 2007, we test whether the effects of these treatments are zero for all subgroups in the relevant years (Zero CATE Test). We also test whether the effect of treatment 2 (for summer 2007) and treatment 3 (for all summers) is constant for all subgroups (Constant CATE Test). For completeness, we report test results for all treatments in all summers. Table 3 presents the results for summer water seasons 2007-2009 using the three methods of covariate choice (i.e., the three panels).

[8] 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).
[9] 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.

Result 1: *There is weak evidence of some subgroups responding in all years to treatment 1 (pure information).*

Result 2: *There is weak evidence of heterogeneous responses in 2007 to treatment 2 (weak social norm) and no evidence that any subgroups respond in later years.*

Result 3: *There is strong evidence of heterogeneous responses in all years to treatment 3 (strong social norms).*

We consider these results in more detail. Columns (1) – (5) show the results for the test of 'Zero CATE.' Columns (1) – (3) report the chi-squared distribution, the degrees of freedom and the respective p-values, and columns (4) – (5) report the normal distribution and the respective p-values. For treatment 1, the null hypothesis of zero CATE for all subgroups is rejected ($p<0.05$) in all three panels for all years. However, these results are fragile. We also ran tests without higher-order terms and with flexible coding of the covariates (see Section 3.2) and the null hypothesis is never rejected (see tables A1 – A3 in Referee's Appendix). We thus conclude that there is suggestive, but weak evidence that treatment 1 affects some subgroups of the experimental sample. For treatment 2, none of the tests for 2008 reject the null and only two out of six tests reject the null for 2009. Almost all the tests reject the null in 2007 for treatment 2, and all tests reject the null for treatment 3 in all years, which is expected given these treatments were shown to have had a causal effect in these years (same results for all alternative specifications in the Referee's Appendix).

Columns (6) – (10) show the results for the test of 'Constant CATE.' Recall our focus for this test is treatment 2 in 2007 and treatment 3 in all years. For treatment 2, only one out of the six tests reject the null hypothesis of constant CATE in 2007. Thus we conclude there is little evidence of heterogeneous treatment effects for treatment 2. In contrast, the null hypothesis of constant CATE for treatment 3 is rejected in all three panels for all years (and in all alternative specifications except for a couple using continuous variables with no higher-order terms; see Referee's Appendix).[10] Thus we conclude there is strong evidence of heterogeneous treatments effects for treatment 3.

[10] We also ran all the tests in Table 3 and the Referee's Appendix using education and race measured at the census tract. In these tests, all specifications reject the null hypothesis of constant effects for treatment 3 and our results for the other treatments do not change.

In summary, using Crump et al.'s (2008) nonparametric tests, we find clear evidence of heterogeneous treatment responses for treatment 3. For treatment 1 and treatment 2, the evidence for heterogeneity is weak and no firm conclusions can be drawn.

5.2 Quantile Regressions

The conclusions drawn from the nonparametric tests are corroborated by the quantile regressions depicted in Figures 1-3. For each treatment over the three summer periods, each figure plots the average treatment effect (dashed line), the QTE (solid line), and the respective confidence intervals of these point estimates.

Result 4: *There is strong evidence of heterogeneous responses only for treatment 3 (strong social norms): water users at the upper end of the distribution respond more.*

For treatment 1 (Figure 1), most of the distribution lies near the zero effect line for all three years without substantial heterogeneity. For treatment 2 (Figure 2), an effect on water use is only detected in 2007, and heterogeneity in this year is confined to the upper half of the distribution. For treatment 3 (Figure 3), there is clear evidence of substantial heterogeneity in 2007, with the greatest water reductions in the upper 20% of the distribution. Summer 2008 also shows greater reduction by high water users, but not as much in 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.

Furthermore, if one were willing to assume rank preservation (see section 3.3) and interpret the figures as estimates of the distribution of treatment effects, one would infer that the treatment messages either reduces water use or has no effect. Nowhere in the distribution is there evidence of statistically significant increases in water use as a result of receiving any treatment message.

5.3 Subgroup Analysis

We define subgroups using the median: a household is thus labeled either as "high value" or "low value." For ownership status, the subgroups are owners and renters. In Section 4, we demonstrated that pre-treatment mean water use across treatment and control groups are 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.

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 treatments 1 and 2, we focus on treatment 3. For completeness, we present subgroup analyses for all the treatments, but would caution the reader against interpreting any statistically significant results as evidence of heterogeneous treatment effects for treatment 1 and 2.

In regression models that include the treatments, the subgroups, and all interactions of the treatments and the subgroups (see Referee's Appendix, Table A6), we can 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 ($p \approx 0.15$). For treatments 1 and 2, we cannot reject this null hypothesis for any year.

Following Heckman, Heinrich and Smith, we simplify the presentation by running independent regressions for each subgroup covariate. Table 4 presents the results for 2007. 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 3.4, we draw the same inferences. We also draw the same inferences from the full regression with interaction terms (see Table A6 in Referee's Appendix).

Result 5: *For treatment 3 (strong social norms), greater responses are observed in households that use more water in the past, live in more expensive homes and are occupied by owners rather than renters.*

The evidence of heterogeneity across these characteristics is strongest in 2007. In 2008, we observe statistically significant heterogeneity conditional on previous water use and fair market value. In 2009, all the tests fail to reject the null hypothesis of no differences across subgroups (there is some weak evidence of heterogeneity based on fair market value). See Referee's Appendix A6-A8.

5.4 Targeting Information Campaigns

Ferraro and Price (forthcoming) show that treatment 3 is the most cost-effective treatment among the three treatments tested. They then demonstrate that by targeting only those households at or above the median use for the previous summer, CCWS could obtain 88% of the original reduction for 65% of the original cost.[11] Under this targeting rule, 2007 summer water use would have been expected to decline by approximately 163 million gallons – the equivalent of shutting off the water to about 4500 households — at a cost of $0.43 per thousand gallons reduced.

Could the information from sections 5.1-5.3 be used to further improve targeting? For example, if instead of targeting households based on their use during the previous year's summer, the utility were to instead target households based on their use in the two months before the campaign, it could obtain 80% of the reduction for 48% of the original cost (i.e., a reduction of 149 million gallons at a cost of $0.35 per thousand gallons reduced[12]).

Result 6: *By targeting on households identified as being more responsive to treatment, the water utility can reduce the overall program cost by over 50%, and the cost per gallon reduced by almost 40%, with less than a 20% decline in the total number of 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 towards $0.30 per thousand gallons reduced (e.g., target large water users who own their home and who reduce, on average, over 3,000 gallons/household). Combined with information on the benefits from reduced water use, heterogeneous treatment effect estimates could be used to determine an optimal targeting strategy.

---

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

[12] When contrasting cost-effectiveness across different conservation and augmentation policy options, one must remember to also consider the persistence of the treatment effects (Ferraro, Miranda and Price, 2011).

5.5 Mechanisms of the Strong Social Norm Message (Treatment 3)

Among the experimental treatments, treatment 3 had the largest and most persistent effect, and the strongest evidence of heterogeneous treatment effects. In this section, we explore the mechanisms through which this treatment affects household behavior. 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).

The first hypothesis is relevant for understanding the long-term effects of the program. This hypothesis can also 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, even after the current inhabitants depart.

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. Previous empirical work implies outdoor water use is more price elastic (e.g., Mansur and Olmstead 2011). Thus, one might predict that, after receiving a message, households would first look to reduce water from outdoor use, just as they would respond to a price increase.

The third hypothesis has not, to our knowledge, been raised in the literature on social comparisons. Rather than working through social preferences, as assumed in the literature, the social comparison may work simply by conveying 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 emphasize the social comparison, rather than treatment 3 in its entirety, because (1) Ferraro and Price 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 "How did the addition of the social comparison reduce water use further?"

5.4.1. Recurring behavioral changes versus one-shot technological investments

To examine the first hypothesis about the nature of the actions taken by households, we use three pieces of information. First, if households were to reduce water use mainly through one-shot investments (e.g., fix leaks, install low-flow toilets), one would expect relatively constant treatment effects across years within seasons. Yet, as indicated in Table 5, the effects wane over time within season (the same waning occurs also in spring months). However, inter-year variations in other factors could also explain this 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). Table 4, however, shows that there is no difference in the responses between older and newer houses.

A 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 6 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: 1,900 versus 1,700 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). Together, these three pieces of evidence suggest a seventh result:

Result 7: *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.*

5.4.2. Outdoor versus indoor water use changes

To examine the second 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 theory and previous empirical evidence suggesting that outdoor water demand is more price elastic, and indirect empirical evidence from our experiment that shows 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% increase. We thus believe a contrast of treatment effects in December and July captures differences in indoor versus outdoor use.

Table 7 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). In order to argue that these results do not imply most of the treatment effect was coming from outdoor watering, one would have to argue that waning only occurs with indoor water use (and thus the July 2008 effect is the same as the outdoor effect in July 2007). Such an argument is difficult to maintain for two reasons. First, about 60% of indoor water use is for toilets, washing and bathing,[13] and 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. 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. We therefore believe that the evidence supports an eighth result:

Result 8: *The empirical 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.*

5.4.3. Social versus private preferences

To examine the hypothesis that the social comparison works through social preferences rather than private preferences, 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: 9,000 gallons/month (from $2.21 to $2.55/1,000 gallons) and 16,000 gallons/month (to $2.88/1,000 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

---

[13] http://www.epa.gov/WaterSense/pubs/indoor.html

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 (see Referee's Appendix for summary, Table A9 – A11). 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).

Next we combine the results from 5.3 and 5.4.1 with assumptions about heterogeneous responses when private preferences motivate water use reductions. In 5.4.1, 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 contrast, we observe in 5.3 that owners are more responsive to the treatment (even after conditioning on all the other covariates; see Table A6 in Referee's Appendix).

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 and thus the norm created by the social comparison to neighbors is more relevant or salient. 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 characteristics and 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 8. 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.

Recall that treatment 3 augments treatment 2 with a social comparison treatment, and treatment 2 affects water use significantly less than treatment 3. Combining this observation with the evidence in this subsection, we arrive at our final result:

Result 9: *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.*

6 Conclusions

Non-pecuniary, information-based strategies are increasingly being used to influence individual decision-making to achieve policy objectives. Despite the increasing application of these strategies, their conceptual and empirical foundations remain under-researched. Although a growing number of scholars are conducting randomized experiments to test information-based strategies, many of them 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 our study of an information-based environmental program aimed at inducing voluntary reductions in the use of a common pool resource, we find strong evidence of heterogeneous treatment effects for a message that augments pure information with pro-social language and social comparisons (strong social norm message). In contrast, the 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) is weak. The social psychology literature on social norms predicts heterogeneous responses to social comparison messages (e.g., Schultz et al.), but this predicted heterogeneity is in the form of a "boomerang" effect, whereby low users discover through the social comparison that they are low users and, in response, increase their use. Based on this literature, 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. Assuming rank preservation is a good approximation in our study (section 3.3), there is no evidence of statistically significant increases in water use anywhere in the distribution as a result of receiving the social comparison message. Likewise, our subgroup analysis reveals no evidence that any subgroup, on average, increases its water consumption as a result of receiving the message. Complementary evidence comes from Ferraro and Price who show that, on average, below-median users reduce their water use upon receiving the social comparison, rather than increase it.

Turning to our mechanism hypotheses, the evidence suggests that the strong social norm message operates through behavioral changes with recurring variable costs rather than one-shot, fixed-cost investments, and through changes in outdoor watering rather than indoor watering. 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 simply 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. A better understanding of the mechanisms through which norm-based messages operate is important for future attempts to estimate the full welfare implications of information-based policies and programs.

Finally, 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% with less than a 20% decline in the aggregate impact.

Third, as suggested by Ferraro and Price, 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), 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).

7 References

Angrist, Joshua and Jörn-Steffen Pischke (2009), Mostly Harmless Econometrics: An Empiricist's Companion, Princeton University Press.

Allcott, Hunt (forthcoming), Social Norms and Energy Conservation, *Journal of Public Economics*.

Allcott, Hunt and Sendhil Mullainathan (2010), Behavioral and Energy Policy, *Science*, 5 March 2010, Vol. 327, No. 5970, pp. 1204–1205.

Angrist, Joshua (2004), Treatment Effect Heterogeneity in Theory and Practice, *Economic Journal*, Vol. 114, March, pp. C52–C83.

Ayres, Ian; Sophie Raseman and Alice Shih (2009), Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage, NBER Working Paper # 15386.

Bennear, Lori and Sheila Olmstead (2008), The impacts of the ''right to know'': Information disclosure and the violation of drinking water standards, *Journal of Environmental Economics and Management*, Vol. 56, pp. 117–130.

Bitler, Marianne; Jonah Gelbach and Hilary Hoynes (2005), Distributional impacts of the Self-Sufficiency Project, NBER Working Paper # 1626.

Bitler, Marianne; Jonah Gelbach and Hilary Hoynes (2006), What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments, *American Economic Review*, Vol. 96, No. 4, pp. 988–1012.

Bitler, Marianne; Jonah Gelbach and Hilary Hoynes (2008), Distributional impacts of the Self-Sufficiency Project, *Journal of Public Economics*, Vol. 92, pp. 748–765.

Bjørner, Thomas B., Lars G. Hansen and Clifford S. Russell (2004), Environmental Labeling and Consumers' Choice: an empirical analysis of the effect of the Nordic Swan, *Journal of Environmental Economics and Management*, Vol. 47, No. 3, pp.411-434.

Bui, Linda and Christopher Mayer (2003), Regulation and Capitalization of Environmental Amenities: Evidence from the Toxic Release Inventory in Massachusetts, *The Review of Economics and Statistics*, Vol. 85, No. 3, pp. 693–708.

Costa, Dora and Matthew Kahn (2010), Energy Conservation 'Nudges' and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment, NBER Working Paper # 15939.

Crump, Richard; Joseph Hotz; Guido Imbens and Oscar Mitnik (2008), Nonparametric Tests for Treatment Effect Heterogeneity, *The Review of Economics and Statistics*, August 2008, Vol. 90, No. 3, pp. 389–405.

Davis, Lucas (2010), Evaluating the Slow Adoption of Energy Efficient Investments: Are Renters Less Likely to Have Energy Efficient Appliances?, NBER Working Paper # 16114.

Deaton, Angus (2010), Instruments, Randomization, and Learning about Development, *Journal of Economic Literature*, Vol. 48, No. 2, pp. 424–455.

DiPasquale, Denise and Edward Glaeser (1999), Incentives and Social Capital: Are Homeowners Better Citizens? *Journal of Urban Economics*, Vol. 45, pp. 354–384.

Djebbari, Habiba and Jeffrey Smith (2008), Heterogeneous Impacts in PROGRESA, *Journal of Econometrics*, Vol. 145, pp. 64–80.

Duflo, Esther and Emmanuel Saez (2003), The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment, *The Quarterly Journal of Economics*, Vol. 118, No. 3, pp. 815–842.

Ferraro, Paul and Michael Price (forthcoming), Using Non-Pecuniary Strategies to Influence Behavior: Evidence from a large-scale field experiment, *The Review of Economics and Statistics*.

Ferraro, Paul, Juan Jose Miranda and Michael Price (2011), The Persistence of Treatment Effects with Non-Pecuniary Policy Instruments: Evidence from a Randomized Environmental Policy Experiment, *American Economic Review: Papers and Proceedings*, Vol. 101, No. 3: 318–322.

Firpo, Sergio (2007), Efficient Semiparametric Estimation of Quantile Treatment Effects, *Econometrica*, Vol. 75, No. 1, pp. 259–276.

Fischer, Corinna (2008), Feedback on Household Electricity Consumption: a Tool for Saving Energy?, *Energy Efficiency*, Vol. 1, No. 1, pp. 79–104.

Goldstein, Noah J., Robert B. Cialdini, and Vladas Griskevicius (2008), A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels, *Journal of Consumer Research*, 35, pp. 472 – 482.

Greenberg, Michael (2005), Concern about Environmental Pollution: How Much Difference Do Race and Ethnicity Make? A New Jersey Case Study, *Environmental Health Perspectives*, Vol. 113, No. 4, pp. 369–374.

Greenstone, Michael, and Ted Gayer (2009), Quasi-experimental and Experimental Approaches to Environmental Management, *Journal of Environmental Economics and Management* Vol. 57, No. 1, pp.21-44.

Habyarimana, James and William Jack (2011), Heckle and Chide: Results of a randomized road safety intervention in Kenya, Journal of Public Economics, Vol. 95, pp. 1438–1446.

Heckman, James; Jeffrey Smith and Nancy Clements (1997), Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts, *Review of Economic Studies*, Vol. 64, pp. 487–535.

Heckman, James, Carolyn Heinrich and Jeffrey Smith (2002), The Performance of Performance Standards, *Journal of Human Resources*, Vol. 37, No. 4, Autumn 2002, pp. 778–811.

Hotz, Joseph, Guido Imbens and Julie Mortimer (2005), Predicting the efficacy of future training programs using past experiences at other locations, *Journal of Econometrics*, Vol. 125, pp. 241–270.

House of Lords, Science and Technology Select Committee (2011), Behaviour Change, 2nd Report of Session 2010-12, London: The Stationary Office Limited.

Imai, Kosuke and Aaron Strauss (2011), Estimation of Heterogeneous Treatment Effects from Randomized Experiments, with Application to the Optimal Planning of the Get-Out-the-Vote Campaign, *Political Analysis*, Vol. 19, No. 1 (Winter), pp. 1-19.

Imbens, Guido and Jeffrey Wooldridge (2009), Recent Developments in the Econometrics of Program Evaluation, *Journal of Economic Literature*, Vol. 47, No. 11, pp. 5–86.

Jin, Ginger and Phillip Leslie (2003), The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards, *Quarterly Journal of Economics*, Vol. 118, No. 2, pp. 409–451.

Mansur, Erin and Sheila Olmstead (2011), The Value of Scarce Water: Measuring the Inefficiency of Municipal Regulations, Yale University, Working Paper.

Lehmann, Michael-Christian (2010), Spatial Externalities of Social Programs: Why Do Cash Transfer Programs Affect Ineligible's Consumption?, Paris School of Economics, Working Paper.

List, John A. and Michael K. Price (forthcoming), Using Field Experiments in Environmental and Resource Economics, *Review of Environmental Economics and Policy*

Rohe, William, Shannon Van Zandt and George McCarthy (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.

Schultz, P. Wesley, Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius. 2007. The Constructive, Destructive, and Reconstructive Power of Social Norms, *Psychological Science*, 18, pp. 429 – 434.

Smith, Kerry and William Desvousges (1990), Risk Communication and the Value of Information: Radon as a Case Study, *The Review of Economics and Statistics*, Vol. 72, No. 1, pp. 137–142.

Smith, Kerry; William Desvousges; Reed Johnson, and Ann Fisher (1990), Can Public Information Programs Affect Risk Perceptions? *Journal of Policy Analysis and Management*, Vol. 9, No. 1, pp. 41–59.

Thaler, Richard and Cass Sustein (2008), Nudge. Improving Decisions About Health, Wealth, and Happiness, Penguin Books.

Yoeli, Erez (2009), Does Social Approval Stimulate Prosocial Behavior? Evidence from a Field Experiment in the Residential Electricity Market, University of Chicago, Working Paper.

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
Pre-Treatment Data						
Water Use in Jun-Nov 2006 1/	58.286	58.012	58.381	58.142	0.200	0.897
Water Use in Apr-May 2007 1/	15.952	15.841	15.957	15.867	0.380	0.768
House's Fair Market Value	257,824	260,984	260,888	258,647	0.950	0.415
Age of House	20.753	20.830	20.710	20.723	0.230	0.878
% Owner Occupiers	0.844	0.836	0.835	0.844	3.190	0.023
% Population 25 years ≥ Bachelor Degree 2/	0.728	0.727	0.728	0.728	0.020	0.997
% of Households White 2/	0.842	0.842	0.843	0.842	0.160	0.924

1/ In thousands of gallons.
2/ At census block group level.
Sources: Experimental Data, 2007 Cobb County Tax Assessor Database, 2000 US Census.

Table 2
Post-Treatment Water Use Descriptive Statistics
(in thousand of gallons)

Variable	(1) Pure Information (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							
Water Use in Summer 2007 1/	36.35	35.39	34.87	36.40	-0.05	-1.00 ***	-1.53 ***
Water Use in Summer 2008 1/	25.51	25.33	24.99	25.49	0.02	-0.17	-0.50 **
Water Use in Summer 2009 1/	27.78	27.38	27.18	27.42	0.36	-0.04	-0.24
Water Use in Winter 07/08 2/	21.63	21.58	21.43	21.71	-0.08	-0.13	-0.28 **
Water Use in Winter 08/09 2/	21.83	21.57	21.63	21.79	0.04	-0.22	-0.16

1/ Summer season comprises July to October use.
2/ Winter season comprises December to March use.
Source: Experimental Data.

Table 3: Test of Zero CATE & Constant CATE for Summer 2007, 2008, 2009 (including higher order terms)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Zero CATE					Constant CATE				
	Chi-Sq	dof	*p-val*	Normal	*p-val*	Chi-Sq	dof	*p-val*	Normal	*p-val*
Top Down Selection of Covariates										
Summer 2007										
Treatment 1	36.47	13	0.001	4.60	0.000	25.44	12	0.013	2.74	0.003
Treatment 2	43.79	12	0.000	6.49	0.000	19.49	11	0.053	1.81	0.035
Treatment 3	230.32	15	0.000	39.31	0.000	175.10	14	0.000	30.44	0.000
Summer 2008										
Treatment 1	52.53	15	0.000	6.85	0.000	52.07	14	0.000	7.20	0.000
Treatment 2	14.21	11	0.222	0.68	0.247	12.91	10	0.229	0.65	0.258
Treatment 3	191.62	13	0.000	35.03	0.000	176.44	12	0.000	33.57	0.000
Summer 2009										
Treatment 1	67.98	11	0.000	12.15	0.000	67.16	10	0.000	12.78	0.000
Treatment 2	21.48	8	0.006	3.37	0.000	20.84	7	0.004	3.70	0.000
Treatment 3	189.92	11	0.000	38.15	0.000	185.52	10	0.000	39.25	0.000
Bottom Up Selection of Covariates										
Summer 2007										
Treatment 1	34.76	9	0.000	6.07	0.000	25.04	8	0.002	4.26	0.000
Treatment 2	37.57	10	0.000	6.16	0.000	13.06	9	0.160	0.96	0.170
Treatment 3	203.97	12	0.000	39.19	0.000	156.30	11	0.000	30.98	0.000
Summer 2008										
Treatment 1	48.47	10	0.000	8.60	0.000	48.12	9	0.000	9.22	0.000
Treatment 2	14.68	9	0.100	1.34	0.090	11.40	8	0.180	0.85	0.197
Treatment 3	126.37	14	0.000	21.24	0.000	120.26	13	0.000	21.04	0.000
Summer 2009										
Treatment 1	67.07	9	0.000	13.69	0.000	66.17	8	0.000	14.54	0.000
Treatment 2	20.72	8	0.008	3.18	0.001	19.73	7	0.006	3.40	0.000
Treatment 3	54.34	9	0.000	10.69	0.000	54.17	8	0.000	11.54	0.000
All Covariates										
Summer 2007										
Treatment 1	35.02	20	0.020	2.37	0.009	31.87	19	0.032	2.09	0.018
Treatment 2	23.43	20	0.268	0.54	0.294	19.90	19	0.401	0.15	0.442
Treatment 3	78.62	20	0.000	9.27	0.000	64.76	19	0.000	7.42	0.000
Summer 2008										
Treatment 1	50.30	20	0.000	4.79	0.000	49.94	19	0.000	5.02	0.000
Treatment 2	13.10	20	0.873	-1.09	0.862	12.68	19	0.855	-1.03	0.153
Treatment 3	71.26	20	0.000	8.10	0.000	66.81	19	0.000	7.76	0.000
Summer 2009										
Treatment 1	82.75	20	0.000	9.92	0.000	82.74	19	0.000	10.34	0.000
Treatment 2	14.71	20	0.793	-0.84	0.799	14.66	19	0.744	-0.70	0.241
Treatment 3	59.21	20	0.000	6.20	0.000	56.61	19	0.000	6.10	0.000

Zero CATE.

H_0: Average Effect for subpopulation with covariates value X is equal to zero for all X.

H_1: Average Effect for subpopulation with covariates value X is different from zero for some X.

Constant CATE.

H_0: Average Effect for subpopulation with covariates value X is equal to ATE for all X.

H_1: Average Effect for subpopulation with covariates value X is different from ATE for some X.

Note: For the zero and constant conditional average treatment effect test, the chi-sq column is equal to the square root of 2K times the normal column plus K, where K is the degrees of freedom. For the column with the zero average treatment effect results, the chi-sq column is equal to the square of the normal column.

Table 4
Subgroup Analysis for Water Consumption (Summer 2007)

	Dependent Variable: Summer 2007						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Previous water use (June - Nov 2006)	Previous water use (April - May 2007)	Fair Market Value	Owners/Renters 1/	Age of Home	% White	% with higher degree
Treatment 1 (Pure Information)	-0.320	-0.136	0.171	0.399	0.121	-0.550	-0.551
	(0.171)	(0.196)	(0.264)	(0.976)	(0.472)	(0.342)	(0.342)
Treatment 2 (Weak Social Norm)	-0.444*	-0.559**	-0.732**	-0.310	-0.651	-0.678	-0.925*
	(0.185)	(0.197)	(0.268)	(0.777)	(0.442)	(0.352)	(0.371)
Treatment 3 (Strong Social Norm)	-0.653**	-0.722**	-0.772**	0.248	-1.533**	-1.421**	-1.100**
	(0.159)	(0.198)	(0.251)	(0.754)	(0.405)	(0.322)	(0.340)
Subgroup var. (high = 1)	27.45**	27.22**	16.54**	2.634**	-4.275**	9.694**	8.642**
	(0.200)	(0.208)	(0.212)	(0.341)	(0.221)	(0.227)	(0.228)
Treat1*high subgrop var.	0.445	0.420	-0.415	-0.530	-0.294	0.894	0.876
	(0.559)	(0.589)	(0.595)	(1.027)	(0.616)	(0.628)	(0.630)
Treat2*high subgrop var.	-0.774	-0.614	-0.383	-0.802	-0.773	-0.775	-0.296
	(0.519)	(0.541)	(0.552)	(0.835)	(0.570)	(0.587)	(0.588)
Treat3*high subgrop var.	-1.994**	-2.176**	-1.406**	-2.101**	0.0419	-0.334	-0.887
	(0.485)	(0.503)	(0.523)	(0.808)	(0.543)	(0.555)	(0.560)
Constant	23.02**	23.87**	28.11**	34.17**	38.51**	31.28**	31.82**
	(0.0636)	(0.0746)	(0.0965)	(0.320)	(0.165)	(0.130)	(0.134)
Observations	102,887	102,887	102,871	102,869	102,461	94,833	94,833
R-squared	0.222	0.217	0.080	0.001	0.006	0.029	0.023
p-value equal impact T1	0.426	0.476	0.486	0.606	0.633	0.154	0.164
p-value equal impact T2	0.136	0.257	0.488	0.337	0.175	0.187	0.614
p-value equal impact T3	0.000	0.000	0.007	0.009	0.939	0.547	0.113

Note: All water consumption variables are in thousands of gallons.
1/ In the case of Owners (=1) / Renter (=0), interaction terms are for owner group rather than high group.
Robust standard errors in parentheses
** p<0.01, * p<0.05

Table 5
Linear Regressions of Water Seasons
(with meter route fixed effects)

	(1) Summer07	(2) Winter0708	(3) Summer08	(4) Winter0809	(5) Summer09
Treatment 1 (Technical Advice)	-0.237	-0.121	-0.0702	0.0335	0.241
	(0.189)	(0.121)	(0.166)	(0.189)	(0.169)
Treatment 2 (Weak Social Norm)	-0.991**	-0.108	-0.190	-0.239	-0.0604
	(0.171)	(0.122)	(0.184)	(0.181)	(0.167)
Treatment 3 (Strong Social Norm)	-1.741**	-0.359**	-0.637**	-0.223	-0.344*
	(0.166)	(0.129)	(0.161)	(0.181)	(0.162)
Water Use from June - November 2006	0.347**	0.0258**	0.127**	0.0379**	0.170**
	(0.0130)	(0.00485)	(0.00981)	(0.00913)	(0.0106)
Water Use in April and May 2007	0.829**	0.335**	0.414**	0.237**	0.427**
	(0.0450)	(0.0171)	(0.0246)	(0.0160)	(0.0251)
Constant	1.874	16.21**	7.086**	15.79**	14.14**
	(1.595)	(0.736)	(0.841)	(0.720)	(1.156)
Observations	106,669	106,669	106,669	106,669	106,669
R-squared	0.634	0.129	0.248	0.021	0.333

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
** $p<0.01$, * $p<0.05$

Table 6
Linear Regressions: Movers and Non-Movers
(with meter route fixed effects)

	(1)	(2)	(3)	(4)
	Summer 2007		Summer 2008	
	Mover 1/	Non Mover	Mover 1/	Non Mover
Treatment 1 (Pure Information)	-1.938	-0.127	-1.367	-0.00830
	(1.058)	(0.192)	(0.968)	(0.169)
Treatment 2 (Weak Social Norm)	-1.329	-0.970**	-0.357	-0.175
	(1.051)	(0.174)	(1.104)	(0.187)
Treatment 3 (Strong Social Norm)	-1.931*	-1.695**	0.826	-0.671**
	(0.959)	(0.169)	(0.999)	(0.163)
Water Use from June - November 2006	0.226**	0.352**	0.124**	0.126**
	(0.0293)	(0.0133)	(0.0322)	(0.0102)
Water Use in April and May 2007	0.985**	0.817**	0.133*	0.423**
	(0.0872)	(0.0462)	(0.0587)	(0.0259)
Constant	-12.94*	2.679	6.177*	7.292**
	(5.905)	(1.622)	(2.424)	(0.861)
Observations	3,667	102,811	3,667	102,811
R-squared	0.525	0.640	0.216	0.254

Note: All water consumption variables are in thousands of gallons.
1/ New residents between December 2007 – September 2008.
Robust standard errors in parentheses
** p<0.01, * p<0.05

Table 7
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)	-0.0535	-0.0495	-0.0673
	(0.0618)	(0.0453)	(0.0623)
Treatment 2 (Weak Social Norm)	-0.334**	-0.0572	-0.0542
	(0.0618)	(0.0453)	(0.0623)
Treatment 3 (Strong Social Norm)	-0.548**	-0.0870	-0.220**
	(0.0618)	(0.0453)	(0.0623)
Water Use from June - November 2006	0.103**	0.00763**	0.0360**
	(0.000652)	(0.000478)	(0.000658)
Water Use in April and May 2007	0.228**	0.0927**	0.119**
	(0.00221)	(0.00162)	(0.00223)
Constant	-0.899*	-1.215**	3.481**
	(0.386)	(0.283)	(0.389)
Observations	106,669	106,669	106,669
R-squared	0.555	0.150	0.193

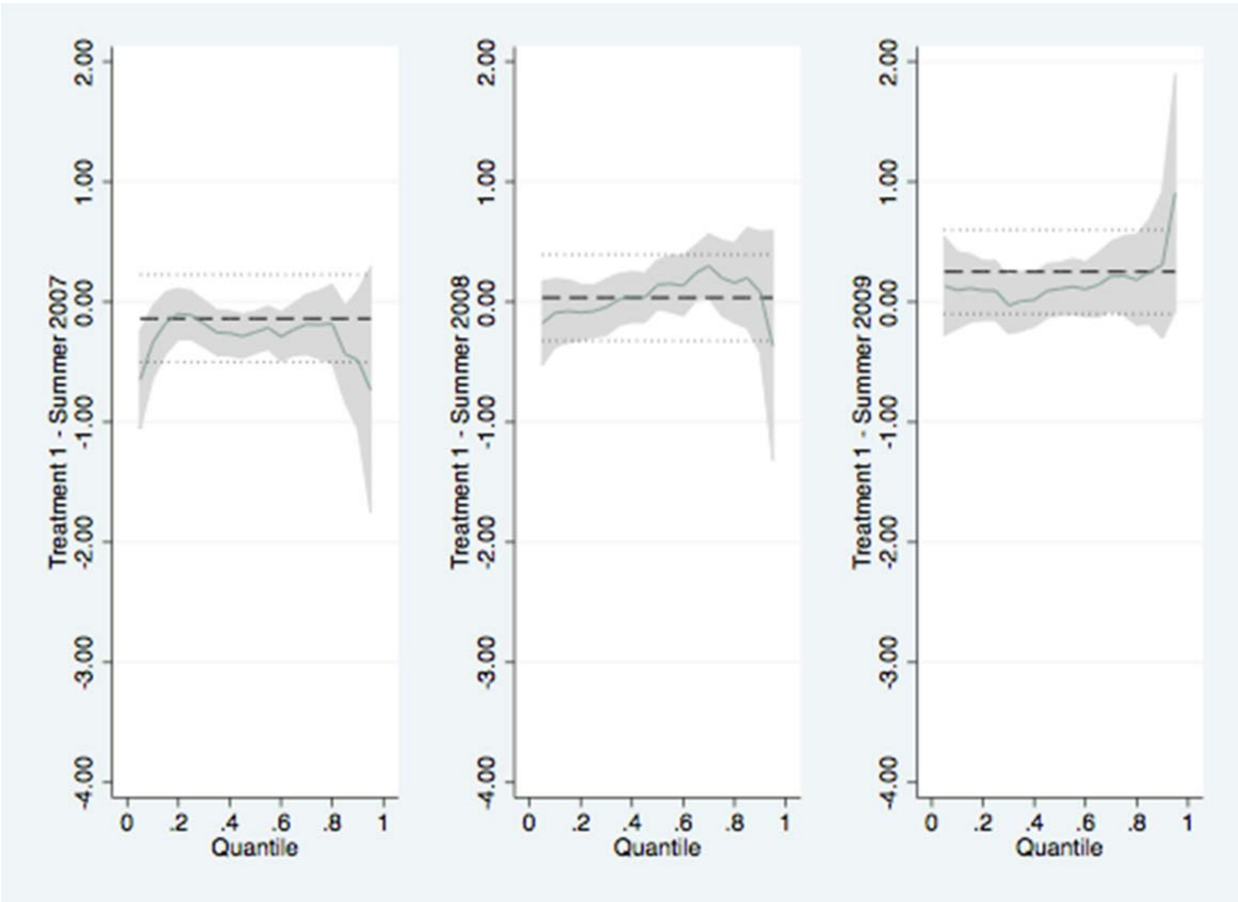
Standard errors in parentheses
** p<0.01, * p<0.05

Table 8
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 - November 2006	0.335**
	(0.0144)
Water Use in April and May 2007	0.823**
	(0.0517)
Ownership status	0.701**
	(0.189)
Fair Market Value	1.79e-05**
	(3.12e-06)
Age of Home	0.0265*
	(0.0112)
Constant	-2.422
	(1.799)
Observations	95,233
R-squared	0.638

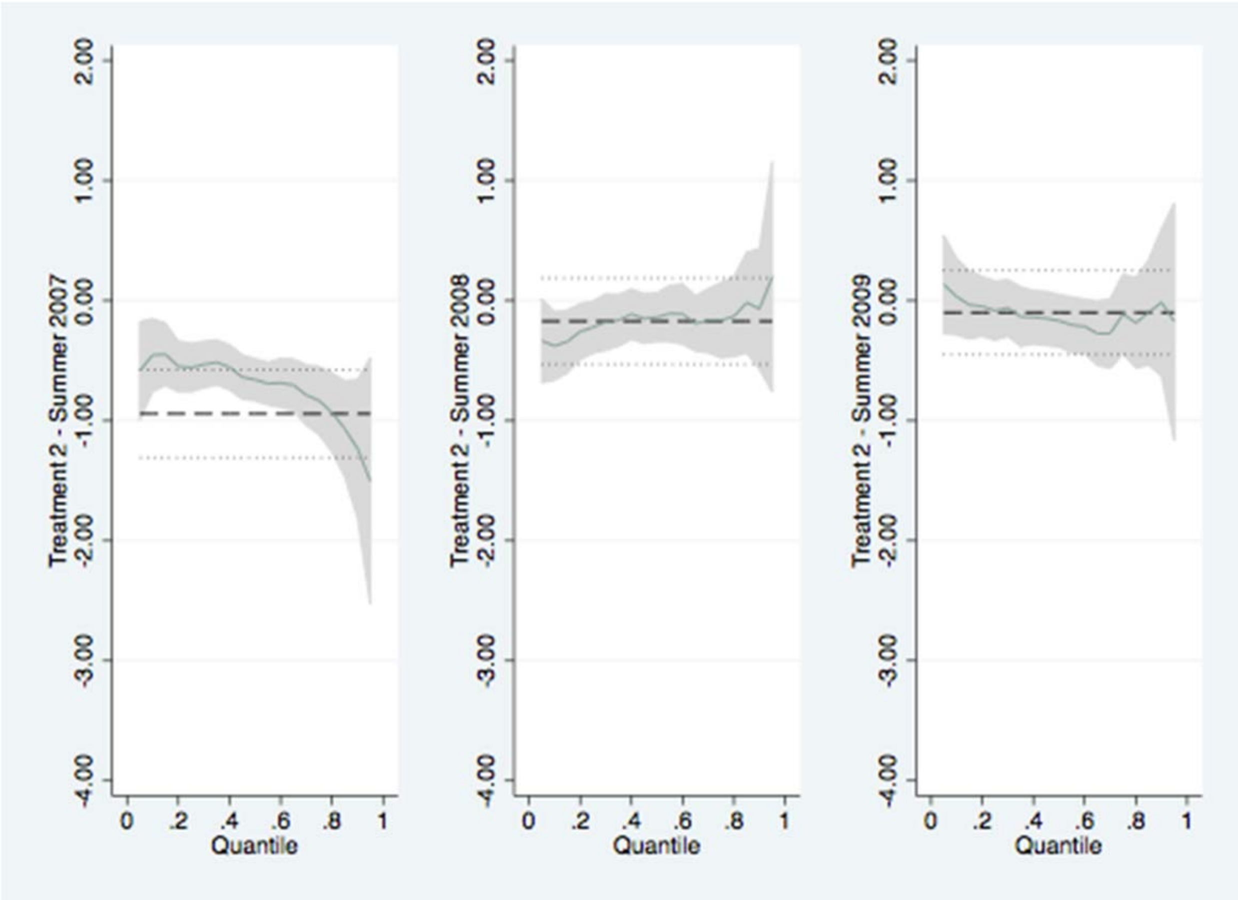
Note: All water consumption variables are in thousands of gallons.
Robust standard errors in parentheses
** p<0.01, * p<0.05

Figure 1
Quantile Treatment Effects for Treatment 1: Pure Information
(summer 2007, summer 2008, summer 2009)



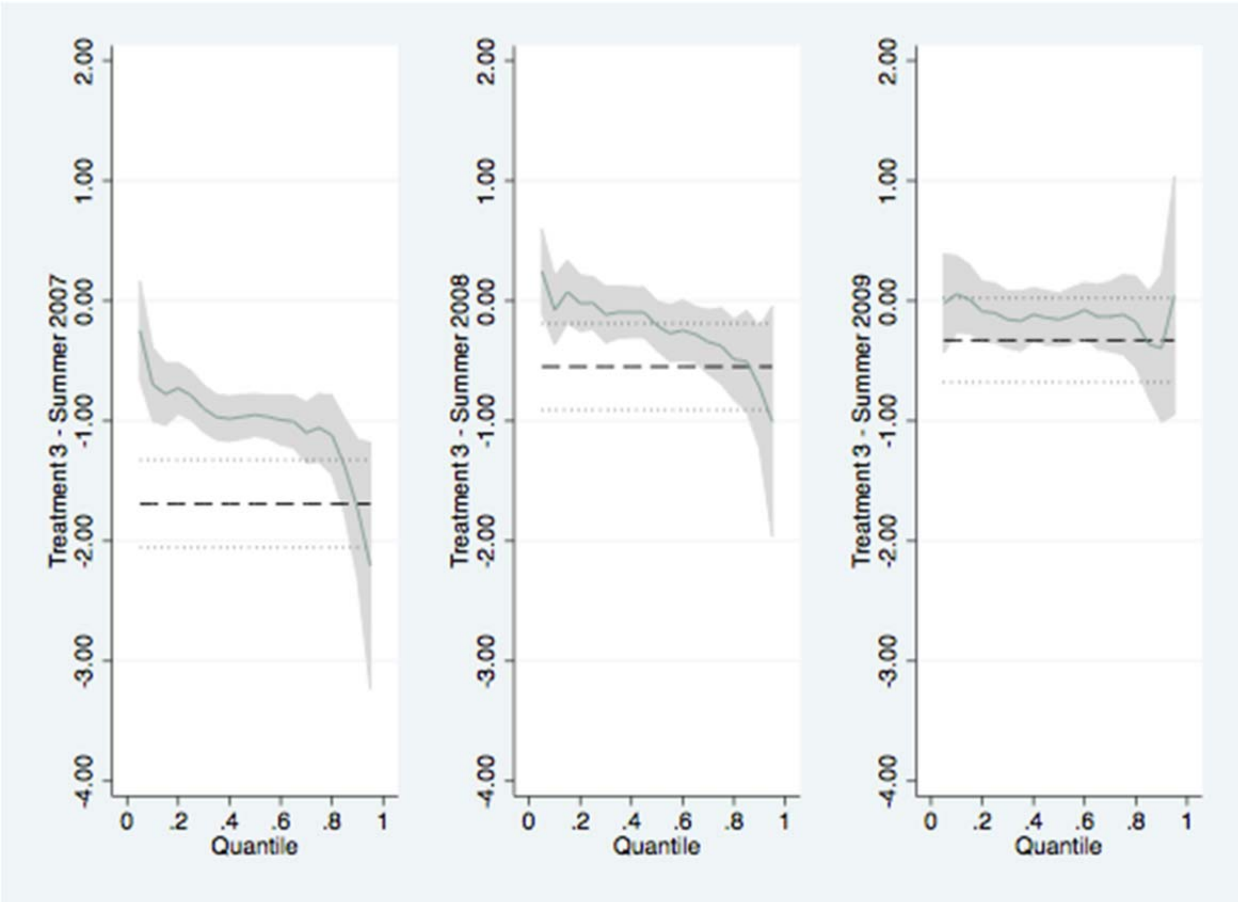
Note: Graphs plot quantile estimates and mean treatment effects across water use distribution. Dashed line depicts the Average Treatment Effect in a linear regression (OLS) framework, while dotted line represents its confidence interval. Solid line depicts the Quantile Treatment Effect and the shadowed area represents its confidence interval.

Figure 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 confidence interval. Solid line depicts the Quantile Treatment Effect and the shadowed area represents its confidence interval.

Figure 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 confidence interval. Solid line depicts the Quantile Treatment Effect and the shadowed area represents its confidence interval.