Do behavioral nudges interact with prevailing economic incentives? Pairing experimental and quasi-experimental evidence from water consumption*

Daniel A. Brent[1] and Casey J. Wichman[2]

[1]Pennsylvania State University
[2]Resources for the Future

Friday 28th December, 2018

Draft paper – please do not cite or distribute

Abstract

Social comparisons are a popular behavioral nudge to promote conservation of energy and water, partially because raising prices is politically difficult. Nudges may interact with prevailing prices, however, potentially crowding out intrinsic motivation to conserve or by increasing the salience of prices. We investigate the interaction of prices and nudges in two experiments in neighboring water utilities. First, we layer randomized behavioral treatments on top of variation in price driven by arbitrary lot-size thresholds that assign marginal prices to customers exogenously. Second, we explore whether behavioral treatments affect consumers' price sensitivity. We find no consistent evidence that social comparisons are more effective at inducing conservation at higher prices or increase consumers' price sensitivity. Ultimately, we find little empirical support that consumers respond to behavioral treatments due to private economic benefits.

Key Words: behavioral interventions, social norms, field experiments, water conservation, price sensitivity, water demand
JEL Codes: D12, C93, H42, L95, Q21, Q25

*dab320@psu.edu, wichman@rff.org. D.A.B. and C.J.W. jointly conceived the idea, performed the analysis, and wrote the paper. The authors thank Nathan Chan and workshop participants at the 2017 AERE conference and the University of Southern California.

1 Introduction

Behavioral interventions are widespread policy options for altering consumption choices. Governments and private companies around the world now look to behavioral economics to manage private and social costs. Behavioral economics has inspired policies targeting a wide range of outcomes including tax evasion (Hallsworth et al., 2017), charitable donations (Croson and Shang, 2008; Shang and Croson, 2009), education (Levitt et al., 2016), healthy eating (Hanks et al., 2012; List and Samek, 2015), and exercise (Royer et al., 2015). These interventions have motivated, and in some cases are the output, of government run "nudge units" such as the United Kingdom's Behavioural Insights Team.

Nowhere have behavioral nudges been more pervasive than for managing energy and water consumption (e.g., Allcott, 2011; Ferraro et al., 2011; Allcott and Rogers, 2014; Brent et al., 2015; Ito et al., 2018). Regulated industries, such as electricity or water and sewer service, are limited by how much they can use price as a tool of conservation. In the state of California, for example, water utilities cannot charge a price greater than cost of service, effectively rendering scarcity pricing illegal.[1] As a result, utilities often rely on nonprice demand-management tools to encourage conservation. Researchers have shown that social comparisons can be effective nonprice policies for conservation, reducing household energy and water consumption between two and five percent (Allcott, 2011; Ferraro and Price, 2013; Brent et al., 2015). At scale, these small reductions can generate substantial benefits for the service provider at relatively low cost, potentially delaying or avoiding investment in new power plants or water sources.

A notable feature of this literature on social comparisons is that the treatment estimates are causal, rising from the randomized nature of program designs implemented by companies such as OPower and WaterSmart. Reconciling these estimates with models of consumer behavior, however, is less transparent. Some have claimed that norm-based information treatments apply a *moral tax* to consumption of externality-producing goods (Levitt and List, 2007; Ferraro and Price, 2013). Others, however, have claimed that information treatments reduce the distortion in consumer's perceptions of price and quantity consumed, thus allowing for re-optimization to regulate informational "internalities" (Allcott and Taubinsky, 2015; Wichman, 2017). As such, there are competing views of whether behavioral interventions affect an individual's intrinsic motivation to conserve, provide direct economic benefits to the consumer, or both.

In line with understanding the behavioral mechanism of conservation, we posit that behavioral policies may interact with prevailing market mechanisms in an ambiguous way. Because behavioral interventions in electricity and water are always overlaid on top of con-

[1] See, e.g., http://www.latimes.com/local/orangecounty/la-me-rates-decision-20150421-story.html.

temporaneous pricing structures, existing estimates confound the interpretation of behavioral treatments and economic incentives if there is an economically significant interaction. Within the current literature, there is virtually no evidence of whether this interaction is meaningful. Our paper fills this gap.

We explore the impacts of water conservation behavior in response to social messaging experiments and large changes in marginal prices for water use. Our analysis produces causal effects by design: first, we evaluate the effects of two independent, randomized messaging experiments implemented by WaterSmart Software at different points in time for neighboring water utilities in Southern California. Second, we exploit two sources of variation that introduce price changes at the household level. One source of price variation comes from arbitrary lot-size thresholds within nonlinear water rate structures that we exploit in a regression discontinuity design. The second source of price variation arises from the utilities' rate-setting practices, included in an instrumental variables framework. Our methodology cleanly identifies the separate impact of the social comparison treatment and price on consumer behavior as well as their joint effect.

Within our unique empirical approach, we answer two questions. First, do customers facing different price levels respond more strongly to norm-based conservation campaigns? We refer to this as the *price-level effect*. We identify the price-level effect from comparing responsiveness to behavioral treatments for otherwise identical households on either side of a price discontinuity introduced by arbitrary lot-size thresholds within a utility's rate structure. Second, do norm-based conservation campaigns increase customers' price sensitivity? We refer to this as the *price-sensitivity effect*. We identify the price-sensitivity effect by estimating demand equations and observing whether our randomized behavioral treatment significantly alters our estimate of the price elasticity.

Our results show no consistent evidence that social comparisons generate more conservation for households facing an exogenously larger marginal price of water. However, in some specifications our estimate of the price-level effect is economically large, but statistically insignificant. Additionally, we find similarly weak and inconsistent evidence for a price sensitivity effect. Treatment induces small increases in the magnitude of demand elasticity in some specifications, although these effects disappear in alternative specifications.

Because norm-based policies are implemented broadly for water and electricity, the policy implications of this research are vast. Allcott and Rogers (2014) and Brent et al. (2015) both show that behavioral nudges interact with prevailing conservation policies. Additionally, Allcott (2015) shows significant heterogeneity in treatment effects, with larger treatment effects for utilities that participated earlier. Recent research shows that the mechanisms through which consumers respond to behavioral nudges has important welfare implications (Allcott and Kessler, Forthcoming; Taylor et al., 2018). Nudges generate unambiguous welfare gains if consumers conserve due to correcting internalities. However, if consumers respond due to a moral tax on consumption then welfare only increases if the price of the resource is sufficiently below the marginal social cost. Strong interactions between nudges and prices would indicate that consumers are at least in part responding to increase private benefit, given the extensive research that consumers do not have full information about prices (Sexton, 2015; Wichman, 2017; Brent and Ward, 2018) or are not responding according to standard neoclassical theory (Sallee, 2014; Allcott and Wozny, 2014; Jacobsen, 2015). Therefore, although it is difficult to directly measure the welfare benefits of behavioral interventions, we are able to shed light on potential mechanisms through which social comparisons affect behavior.

There is, however, a growing body of evidence that focuses on comparing the effects of moral and neoclassical incentives on energy and water consumption. Ito et al. (2018) explore the effectiveness of a standard moral suasion nudge relative to dynamic electricity pricing treatments. They find that moral suasion induces sizable effects in the short-run that dissipate quickly relative to dynamic prices that exhibit longer-run effects. Our project is different in that we seek to understand how the moral suasion treatment interacts with *underlying* economic incentives. Additionally, Brandon et al. (2018) implement a randomized OPower experiment in which personalized energy reports were sent to electricity customers that targeted aggregate savings or peak-load savings, and measured the response of these treatments during peak-load and non-peak load events. They find that a combination of treatments induced a larger effect than the joint effect of each treatment in isolation or, in other words, that treatments were complimentary. This result suggests an important role for exploring other policy complimentarities, particularly with respect to interactions with economic incentives because nudges can highlight the private economic benefits of conservation. Finally, in another project, List et al. (2017) show that economic incentives (via a rewards program) can better target electricity consumption reductions from low-use, low-variance households, who are typically less responsive to nudges. Importantly, electricity and water are often priced using nonlinear increasing-block rate structures where the economic benefits from conservation are positively correlated with consumption. Thus it feasible that low-use, low-variance consumers respond to nudges differently because of different private economic returns from conservation. This latter effect is precisely what we seek to estimate in this paper.

Overall, we find little evidence that moral nudges interact with underlying economic incentives. This is an important result because nearly all behavioral public policy has the potential to interact with existing neoclassical incentives. Placed alongside the previously mentioned literature, our study provides a clearer view of the mechanisms underlying responses to behavioral treatments. Behavioral nudges can be criticized for providing too many types of information to isolate the relevant mechanism for consumer behavior, but we fail to find convincing evidence that making economic incentives more salient is a relevant factor for behavioral interventions. This finding sharpens our view of past and future conservation policies because nearly all behavioral nudges for electricity and water are layered on top of prevailing rate structures.

2 Conceptual framework

To show how nudges and incentives may interact, we begin with the general framework of Allcott and Kessler (Forthcoming). Consider a consumer with income y who gains utility from the consumption of water w and numeraire good x. w generates consumption utility of $f(w;\alpha)$, where α captures consumer tastes as a demand shifter. We include an internality parameter $\gamma > 0$ that affects choice but not experienced utility, such as imperfect information, mistakes in evaluating private benefits of water consumption, or some other behavioral bias. For our purposes, it is useful to think of γ as inattention to water consumption. Consumers thus have perceived utility $\hat{f}(w;\alpha,\gamma)$, which we assume takes the form $\gamma^{-1}f(w;\alpha)$. Thus, utility is expanded for $\gamma > 1$ and contracted for $0 < \gamma < 1$.

Following Levitt and List (2007) and Ferraro and Price (2013), we include a "moral utility" term, $M = m - \mu w$, which captures nonpecuinary impacts associated with consumption of w. We define $\mu \geq 0$ as a marginal "moral tax" on consumption of w.

We summarize individual-specific parameters in the vector $\theta = \{y, \alpha, \gamma, m, \mu\}$ so that the consumer maximizes

$$\max_{x,w} \hat{U}(\theta) = x + \gamma^{-1}f(w;\alpha) + m - \mu w \tag{1}$$

subject to her budget constraint

$$y = x + pw \tag{2}$$

where $p \geq 0$ is the marginal price for water consumption. The consumer allocates all non-water expenditures to the numeraire, thus satisfying her budget constraint with equality. Because water is necessary for human survival, we are comfortable ignoring corner solutions.

Standard first-order conditions govern the consumer's choice of water consumption, $\tilde{w}$:

$$f'(\tilde{w};\alpha) = \gamma(\mu + p). \tag{3}$$

Eq. 3 states that consumers will choose consumption of $\tilde{w}$ to equalize their marginal perceived utility with the sum of perceived monetary and moral costs. Because γ introduces a wedge between experienced marginal utility and a consumer's true marginal utility, choice of $\tilde{w}$ is not required to be individually optimal. The framework so far is consistent with stylized formulations in Sexton (2015) and Wichman (2017) who model price (and quantity) misperceptions. The only difference is the inclusion of the Ferraro and Price (2013) moral cost parameter.

We can express changes in consumption by totally differentiating Eq 3:

$$f''(\tilde{w};\alpha)\mathrm{d}\tilde{w} = \mu\mathrm{d}\gamma + \gamma\mathrm{d}\mu + p\mathrm{d}\gamma + \gamma\mathrm{d}p. \tag{4}$$

Now, let the nudge be represented by $d\gamma$ and $d\mu$ (for clarity, we assume the nudge is corrective in that it reduces information distortions, or $d\gamma \implies \gamma \to 1$). Because the nudge sets $dp = 0$, we can express the demand effect of a nudge as

$$\mathrm{d}\tilde{w} = \frac{1}{f''(\cdot)}\left[(\mu + p)\,\mathrm{d}\gamma + \gamma\mathrm{d}\mu\right]. \tag{5}$$

Under standard assumptions of demand (i.e., diminishing marginal utility), f'' is weakly negative, which implies that the nudge will (weakly) reduce water demand in equilibrium for $\gamma < 1$.[2] Eq. 5 shows that the total effect of the nudge depends on how changes in perceptions interact with moral and explicit prices as well as how changes in moral costs interact with perception. The vast majority of research to date assumes implicitly that the increased salience of private economic benefits of conservation are negligible; in other words, these studies interpret the effect of the nudge as if $p\mathrm{d}\gamma = 0$. That is, the majority of experiments focused on exploring the effects of salience or moral suasion ignore their underlying interaction with prices. Theoretically, this omission is a potentially important oversight because behavioral interventions for water and energy use are implemented on top of prevailing prices, which are often nonlinear. Furthermore, many nudges aimed at water and energy conservation explicitly communicate the private financial benefits of conservation.

This simplified representation of demand translates directly to our first empirical hypothesis: the existence of an economically important interaction between behavioral treatments and conventional pricing mechanisms. We define this effect as the **price-level effect (PLE)**, which measures the magnitude of $\mathrm{d}\tilde{w}$ in response to the nudge that is driven by differences in price levels. Our null price-level hypothesis posits that $p\mathrm{d}\gamma = 0$. Evidence of a nonzero price-level effect would lend support to the notion that consumers change consumption in part due to changes salience of private economic benefits from conservation. We test this by comparing the effect of randomized nudges for households who face exogenously different marginal prices. We describe empirical identification in the subsequent section.

Additionally, we explore a second, complementary approach to investigate whether nudges affect consumer demand through neoclassical price mechanisms. Consider a change

[2] For $\gamma > 1$, the nudge could increase consumption if, e.g., consumers had been initially over-perceiving the costs of consumption. This stylized result is captured empirically in Wichman (2017).

in price, dp. Using Eq. 3, we can define the resulting price elasticity,

$$\hat{\varepsilon}^p = \frac{\gamma}{f''(\cdot)} \frac{p}{\tilde{w}} = \gamma \varepsilon^p \tag{6}$$

where ε^p is the neoclassical price elasticity of demand and the hat indicates "perceived" price elasticities. This formulation leads directly into our second hypothesis. We define the **price-sensitivity effect (PSE)** as the degree to which nudges affect price sensitivity. Because social comparisons operate through both channels of μ and γ, our null price-sensitivity hypothesis is $\partial \hat{\varepsilon}_p / \partial \gamma = 0$. Evidence of a nonzero price-sensitivity effect would suggest that consumers' sensitivity to price is affected by the nudge, thus providing support for the idea that consumers respond to nudges, at least in part, because of private economic benefits due to internality correction.

3 Empirical setting and strategy

3.1 Data

The data we use in this analysis are household-level water consumption records for two utilities in Southern California. We obtained these data through partnership with WaterSmart Software. We refer to the larger utility in our sample as "Large Utility" and, correspondingly, the smaller utility is "Small Utility."[3] These two utilities share a geographic border and their residents form a common labor market along with other nearby municipalities. Both utilities combine water and sewer services and also serve as the electric utility. Figure 1 shows the geographic distribution of households in the treatment and control groups in each utility.

Large and Small Utility have different pricing structures, and the water rates have changed over time. Large Utility has "budget-based" increasing-block rates in which consumption thresholds for the marginal price blocks vary with geographic region and lot size. This means different households will move to higher marginal prices at different levels of consumption. There are three geographic zones: low, medium, and high. The geographic zones refer to the water requirements based on temperature conditions; the low zone has the most moderate weather and the high zone has hotter weather. There are five lot size thresholds leading to fifteen unique sets of consumption tiers that determine marginal prices. Small Utility has a standard increasing-block rate structure. Figure 2 displays the full water rate structure over time for both utilities.

For each household in our sample, we have consumption for the given billing period and the relevant prices for the consumption. Households receive water bills every two

[3] As part of the confidentiality agreement we cannot disclose the names of these utilities.

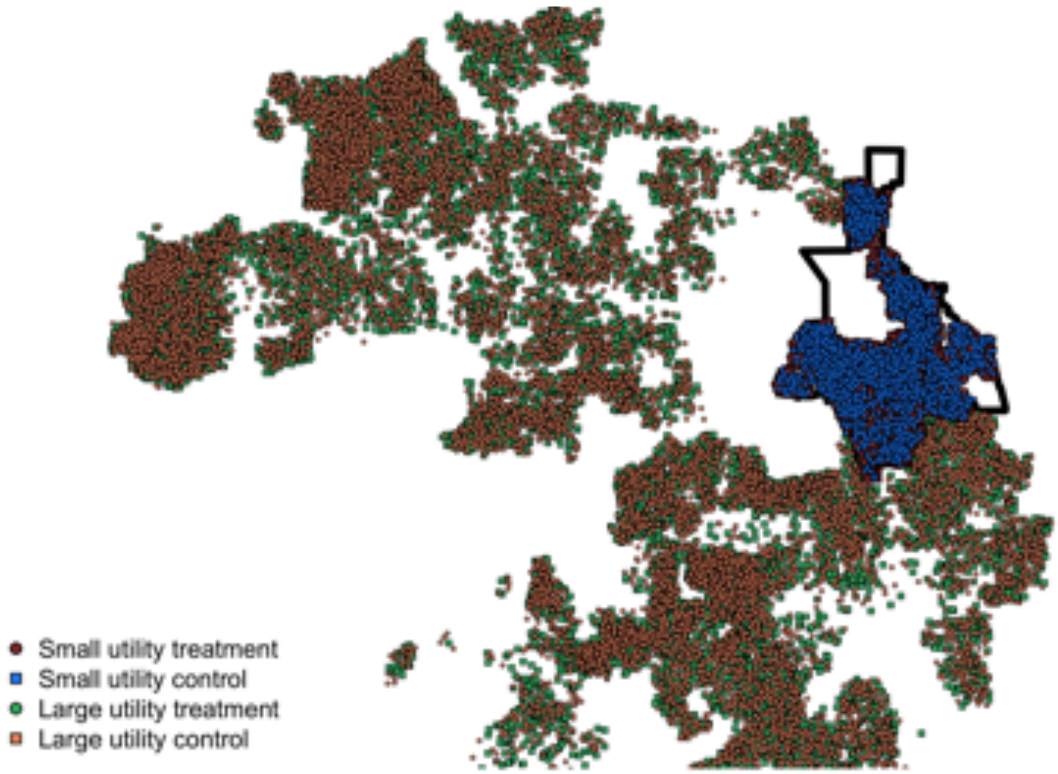


Figure 1: Water utility service area (partial), sample boundaries, and households by treatment status

Note: The small utility is contained within the black border and the large utility is outside the border. Household locations are scrambled by .001 decimal degrees to preserve anonymity.

months. To protect anonymity, geographic coordinates for each household were scrambled within 0.001 degrees (a maximum of approximately 365 feet), which permits us to identify the neighborhood of the household. Each account in our sample was randomized into the treatment or control group by WaterSmart Software. All households in each utility begin receiving HWRs at the same time. Households in both utilities are billed bimonthly leading to six billing periods each year. Treatment began during the sixth billing period of 2014 in Small Utility and during the second billing period of 2015 in Large Utility.

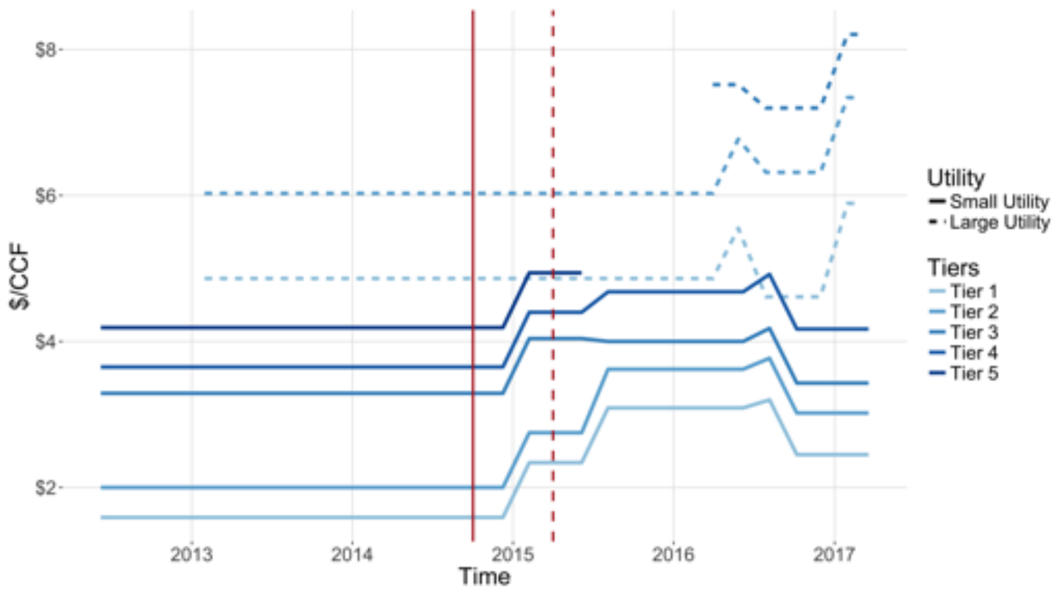


Figure 2: Marginal Prices Over Time

Note: The colors depict the marginal price for different consumption tiers. The dashed lines show prices for the Large Utility and the solid lines show marginal prices for the Small Utility. The vertical dashed and solid lines depict the treatment start date for the Large and Small Utilities respectively.

3.2 Treatment design

WaterSmart Software (henceforth WaterSmart) is a clean technology company that contracts with water utilities to help them manage demand.[4] In addition to assistance with analyzing and interpreting meter reading data, WaterSmart primarily focuses on helping utilities reduce water consumption by providing consumers with additional information through customized Home Water Reports (HWRs) (Figure 3) and an online customer account portal. For many utilities WaterSmart randomizes the assignment of households who receive HWRs in order to evaluate the causal impact on water consumption (see, e.g., Brent et al. (2015)). Since customers opt-in to view their online account, we focus here on the treatment effect for households receiving a HWR.

The one-page HWR as tested has three components. The main component (in the upper left of the figure) is a social comparison. WaterSmart estimates the household's total water consumption over the prior two months from utility billing records and compared that to the consumption of "average neighbors" and "efficient neighbors". "Neighbors" are defined as households that have the same number of occupants and similar irrigable area across the utility, such that the general water requirements within a peer group are comparable. "Efficient neighbors" were peers with consumption in the bottom 20%. Households with consumption above the median of their peer group receive a "Red" normative message

[4]More information is available on their website: http://www.watersmartsoftware.com/.

Table 1: Summary statistics and balance on observables

Sample	Variable	Treat	Control	Difference	KS	MW	T
Small Utility	Pre-Treatment Water	442.0	437.9	4.2	0.86	0.56	0.42
Small Utility	Pre-Treatment Water (Summer)	503.8	499.8	4.0	0.92	0.84	0.50
Small Utility	Pre-Treatment Water (Winter)	403.0	399.4	3.6	0.50	0.48	0.45
Small Utility	Lot Size	9923.5	9606.3	317.2	0.65	0.56	0.13
Small Utility	Sq. Ft.	1950.3	1947.3	3.0	0.91	0.72	0.88
Small Utility	Beds	3.0	3.0	-0.0	0.85	0.96	0.89
Small Utility	Baths	2.2	2.2	0.0	0.99	0.52	0.62
Large Utility	Pre-Treatment Water	598.7	599.2	-0.5	0.81	0.98	0.85
Large Utility	Pre-Treatment Water (Summer)	708.9	709.4	-0.5	0.87	0.79	0.90
Large Utility	Pre-Treatment Water (Winter)	544.0	544.7	-0.7	0.62	0.87	0.80
Large Utility	Lot Size	10147.8	10121.6	26.2	0.88	0.70	0.81
Large Utility	Sq. Ft.	2135.4	2159.3	-23.9	0.07	0.26	0.05
Large Utility	Beds	3.5	3.5	-0.0	0.65	0.91	0.69
Large Utility	Baths	2.5	2.5	-0.0	0.97	0.67	0.59

Note: The table shows the average values for a variety of households characteristics for the treatment and control groups in each utility. All the pre-treatment water variables are measured in gallons-per-day. Lot size and sq. ft. (indoor living space) are measured in square feet. Beds and baths are the number of bedrooms and bathrooms. The last three columns present the p-values from test statistics. KS is the non-parametric Kolmogorov-Smirnov equality of distributions test, MW is the non-parametric rank-order test, and T is the two-sided t-test for difference in means.

(shown in Figure 3), those with consumption between the median and 20^{th} percentile receive a "Yellow" message, and those below the 20^{th} percentile receive a "Green" message. (Home Water Reports showing the latter two categories are provided in the Appendix).

The second component (across the bottom of Figure 3) is a list of three personalized recommendations for strategies to save water. Recommendations include installing low-flow toilets and switching to native plants. Based on data available from the utility (described more below) or on results from a baseline household survey with limited responses, WaterSmart personalized these recommendations to the extent possible. For example, if a household had no outdoor area it was not given a recommendation regarding irrigation. The personalized recommendations provide estimates of the water savings in gallons and in dollars, and the dollar estimates rely on the marginal price the household faces. The third component (in the upper right of Figure 3) cycles between a variety of messages about water conservation and utility programs.

To show that the randomization was conducted properly we graph average water use over time across treatment groups and perform a variety of balance tests. Figure 4 shows the average historical water consumption for the treatment and control groups in both utilities. The treatment and control groups had similar consumption prior to the intervention and after treatment the treatment groups use less water. Table 1 shows that treatment and control groups are well balanced on a range of observables based on a variety of parametric and non-parametric tests. Out of the 42 tests performed none has a p-value below 0.05 and only two have a p-value below 0.1.

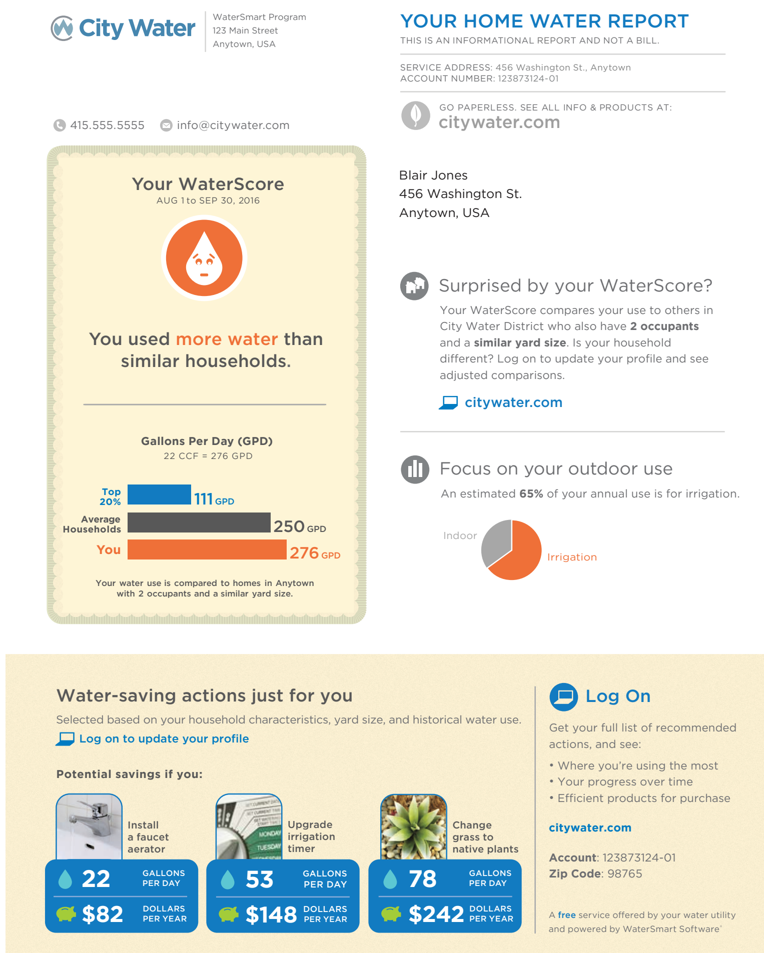


Figure 3: Home Water Report

Note: This is an example of a generic "Red" Home Water Report (HWR). These households used less than the median of their peer group.

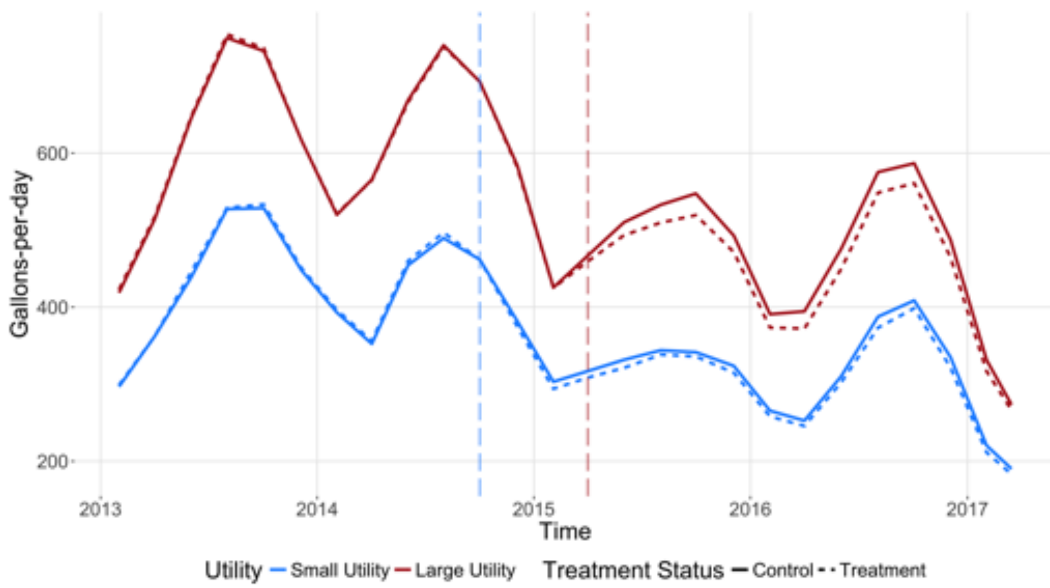


Figure 4: Historical Water Consumption by Treatment Status and Utility

Note: The graph displays the average consumption of the treatment and control groups in each utility for every billing period in the sample. The solid lines represent the control groups and the dotted lines represent the treatment groups. The vertical dashed lines designate the start of the treatment period for each utility. The line colors designate utilities.

3.3 Base treatment effects

The basic regression framework follows Allcott and Rogers (2014) where we regress normalized water use on the treatment indicator while restricting the sample to the treatment period. While this setup does not allow for the inclusion of household fixed effects we can more flexibly control for household observables and the randomization alleviates concerns of static confounding effects. Normalized water use simply divides each household's water use in gallons-per-day (GPD) by the average consumption of the control group in the post-treatment period within the same utility. This specification maintains the interpretation of coefficients as percentage changes in water consumption, but unlike the logarithmic transformation does not dampen the effect of high users. This is important in the context of social comparisons because prior research shows that most of the savings are concentrated among high users (Allcott, 2011; Brent et al., 2015).

We interact the treatment indicator with a Large Utility indicator to evaluate the differential effects across the Large and Small utilities. We refer to this as the double difference model (DD) since it examines differences in treatment and control groups as well as differences across utilities. To control for differential effects specific to the utilities that may impact the responsiveness to treatment we examine the treatment effect model restricted to households within 10 kilometers (km) from the shared utility border.

Our primary regression equation used to estimate treatment effects regresses normalized average daily water consumption ($\tilde{w}_{it}$) for household i during billing period t,

$$\tilde{w}_{it} = \alpha + \gamma_1 \text{Treat}_i + \gamma_2 (\text{Treat}_i \times \text{Large}_i) + \beta X_{it} + \tau_{it} + \varepsilon_{it}, \tag{7}$$

where Treat_i identifies the treatment group, Large_i indicates whether the household is served by the Large Utility, X_{it} is a vector of controls including weather variables and pre-treatment water consumption, τ_{it} is a period-by-utility fixed effect, and ε_{it} is the residual error term. The set of γ coefficients are estimates of the utility-specific average treatment effects of HWRs.

3.4 Identifying price-level and price-sensitivity effects

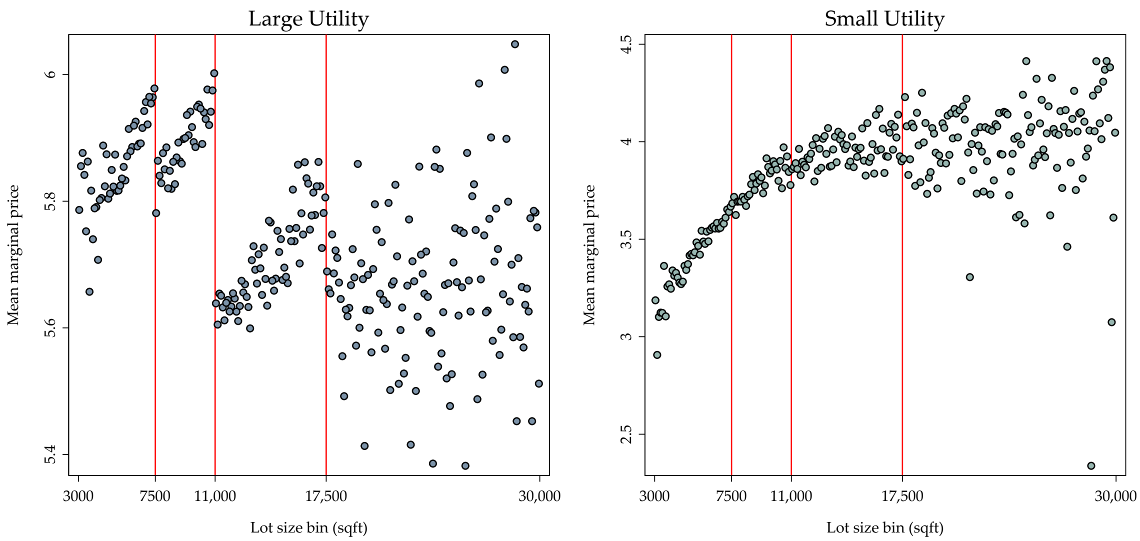
In order to identify the differential impact of HWRs for households facing different prices we estimate a difference-in-discontinuity model (DD) that exploits a discontinuity in the rate structure of the Large Utility. As described above Large Utility uses a budget-based increasing-block rate structure, where the tier thresholds depend on the climate zone and lot size (in square feet). There are five lot size tiers (0-7499, 7500-10999, 11000-17499, 17500-43559, $\geq$ 43560), and households with lower lot sizes are allocated less water before moving to a higher pricing tier. Therefore, houses that are just below the lot size tier (e.g. 7499 sq.

ft.) on average face higher prices than houses just above a tier (e.g. 7500 sq. ft.). Because this budget-based billing only occurs in the Large Utility, we use a third difference (in a difference-in-difference-in-discontinuity design, or DDD) to compare similar households above and below the lot-size threshold across utility boundaries. These households below the lot-size discontinuity will only face higher prices in the Large Utility. We restrict the analysis to various lot-size bandwidths such that the households are relatively close to the lot size thresholds (within 1000 square feet).

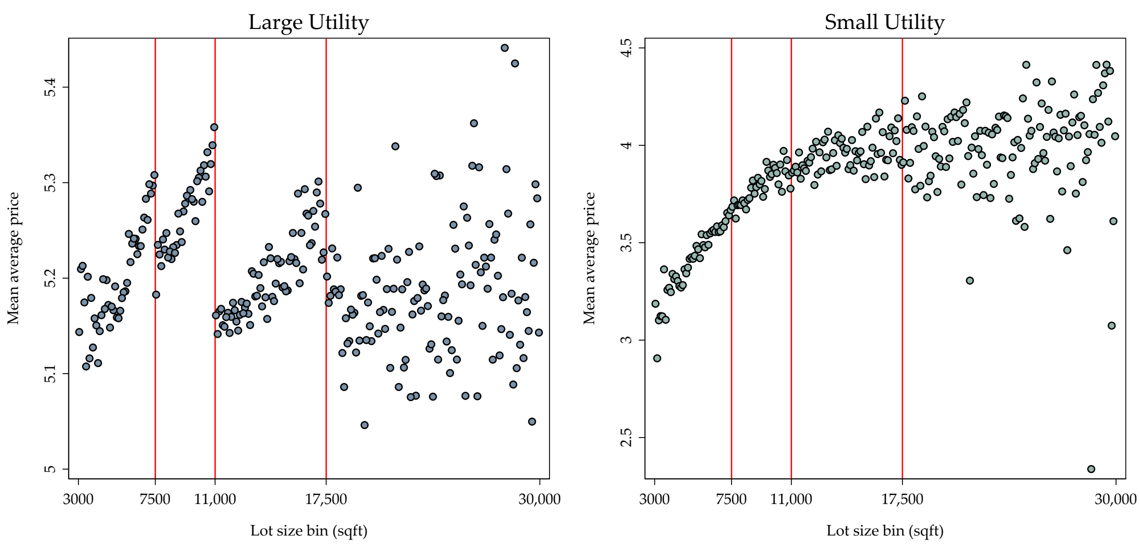
To clarify, the primary differences we exploit are as follows. The first difference is identified via random assignment of treatment status, the second difference is identified by the discontinuity in rate structure, and, as a robustness check, we include a third difference across utilities. The primary motivation for the robustness check across utilities is that our discontinuity depends on lot size, which in turn affects water consumption. Since many studies find that households that use more water are more responsive to social comparisons, we also estimate a difference in discontinuities model that nets out any primary effect of the lot size threshold.

Figure 5 shows how the lot size threshold has a differential effect on marginal prices in each utility. These figures present mean marginal (average) prices relative to lot size in 100 sq. ft. bins. There is a distinct jump in the expected marginal price for households just below the lot-size threshold in the Large Utility (panel (a)), but not in the Small Utility (panel (b)). To highlight the difference in typical marginal prices induced by the lot-size threshold we define "low" households as those who less than 1000 feet below a lot-size threshold (e.g. 6499-7499 sq. ft.). We define "high" households as those less than 1000 feet above a lot-size threshold (7500-8500 sq. ft.). The raw data shows that in the Small Utility the average marginal price for low and high size households is \$3.65 and \$3.77 respectively—so high-size households on average pay more for water because they are higher users. In the Large Utility the average marginal price for low- and high-size households is \$5.98 and \$5.81 respectively—low-size households pay more for water *despite* the fact that they are lower users. Note that while the average price difference in the Large Utility from the lot-size threshold is only \$0.20, the marginal price increase from moving to the higher tier is more than \$1. The average marginal prices reflect the both the change in marginal prices and the probability that a household moves into the higher consumption tier. Therefore, some households will face significant marginal price increases due to the lot-size threshold discontinuity.

In Figure 6, we present a different way to visualize the price variation that we are exploiting. Here we show the different rate structures for households in three different lot-size groups. Each lot-size group faces the same set of marginal prices, but larger lots are allocated a larger proportion of bi-monthly consumption at the lower marginal price. As shown, a household with a 7400 sq. ft. lot is bumped into the second price tier at 28 ccf,



(a) RD treatment in marginal prices at lot-size thresholds for both utilities



(b) RD treatment in average prices at lot-size thresholds for both utilities

Figure 5: Discontinuities in average and marginal price driven by lot-size

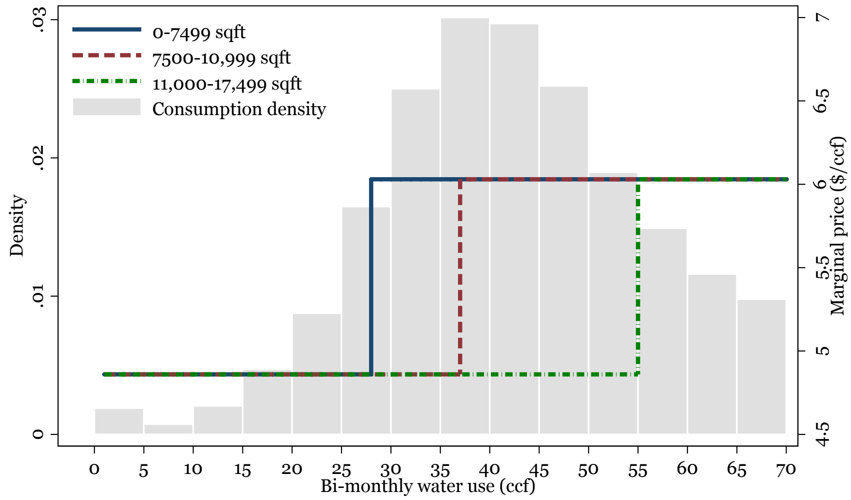


Figure 6: Changes price structure driven by lot-size groups

whereas a household with a 7500 sq. ft. lot is bumped into the second price tier at 37 ccf. Further, a household with an 11,000 sq. ft. lot isn't bumped into the second price tier until after 55 ccf. Moreover, these inframarginal price differences are not trivial: households with lots smaller than 7500 sq. ft. face a marginal price increase of 19.4% nine units of consumption sooner than do households with slightly larger lot sizes. At the 11,000 sq. ft. threshold, this inframarginal price difference is sustained for 18 ccf every two months.

The standard identifying assumptions in RD frameworks are that (a) other covariates move smoothly through the discontinuity induced by the running variable and (b) the running variable cannot be manipulated. The latter assumption is satisfied by noting that lot sizes are effectively fixed over time. In our setting, if other variables associated with water consumption changed discontinuously then we would worry that (a) is not satisfied. As a visual test of this assumption, we present in Figure 7 three relevant variables for water consumption across our RD threshold: irrigable area of lot, indoor square footage of home, and number of bathrooms. Notably, irrigable area, which we anticipate to be highly correlated with lot size, moves nearly linearly through the lot-size discontinuities, which provides convincing support for the RD assumptions. We observe no obvious discontinuity in square footage and number of bathrooms at the discontinuities either. This analysis for additional covariates is presented in Figure A.3.

Formally the DD model interacts the variables in equation 7 with an indicator for whether the household is below the lot size threshold:

$$\tilde{w}_{it} = \alpha + \gamma_1 \text{Treat} + \gamma_2(\text{Treat} \times \text{Low}) + \theta_1 \text{Low} + \beta X_{it} + \tau_{it} + \varepsilon_{it}, \tag{8}$$

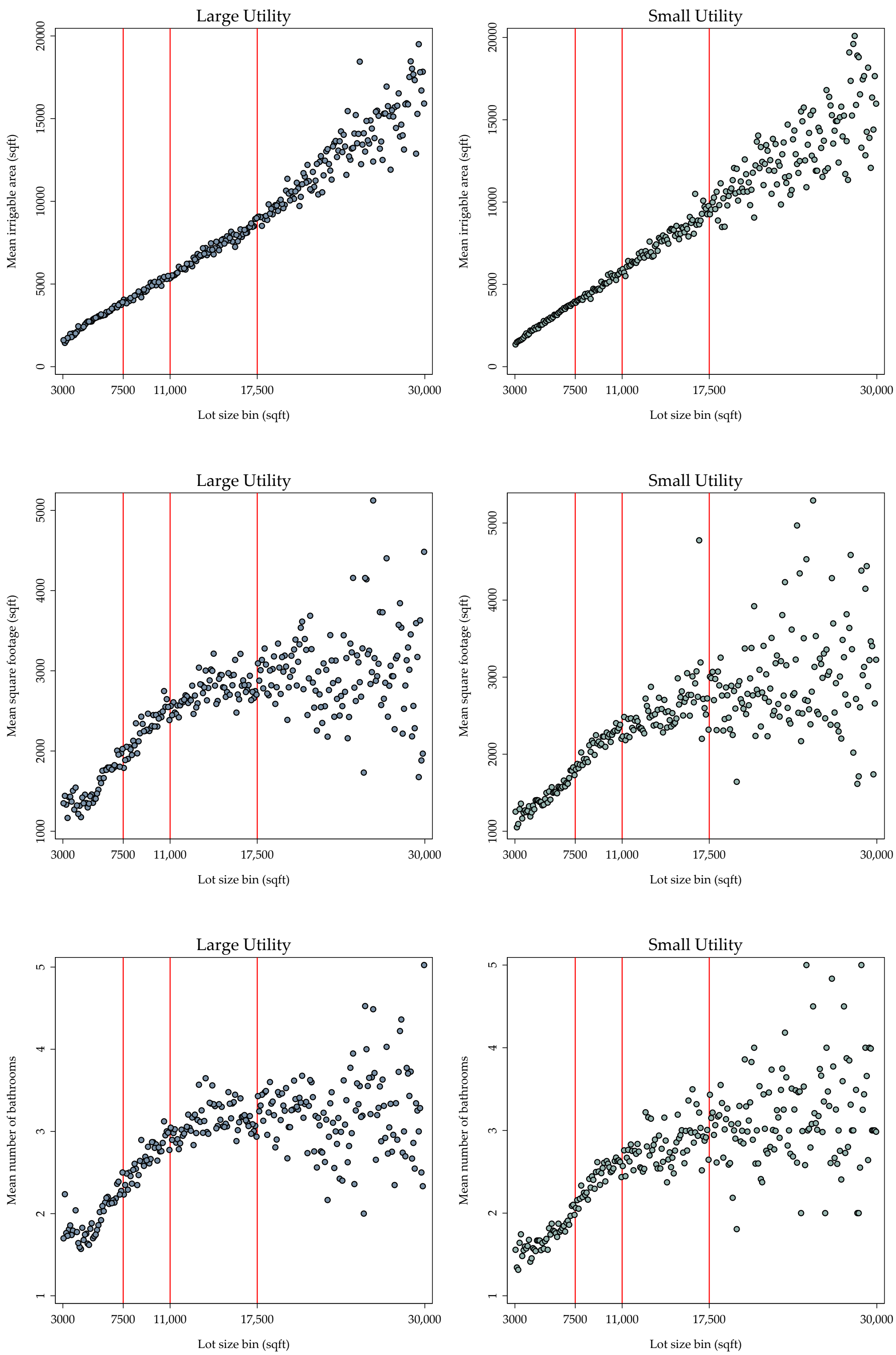


Figure 7: Covariate distributions across lot-size thresholds for both utilities

where all variables are the same as in equation 7, except we add a new indicator (Low), which signifies that a household is below any of the lot-size thresholds in the Large Utility. We estimate this equation using Large Utility households only. We vary the bandwidth of lot-size from $+/-1000$ sq. ft., $+/-750$ sq. ft., and $+/-500$ sq. ft. of lot-size thresholds. In additional specifications, we include a continuous lot size variable interacted both with the lot size and treatment dummies to control for differential trends in water use on either side of the lot-size threshold and across treatment and control households. This model with lot size interactions mimics the standard local linear regression discontinuity design. We also estimate this model for the three primary lot-size discontinuities (at 7000 sq. ft., 11,000 sq. ft., and 17,500 sq. ft.) individually.

The *price-level effect*, the amount of the HWR treatment effect that is driven by exogenously different marginal price levels, is given by γ_2. This model allows us to test the hypothesis that $\gamma_2 = 0$. The regressions include both household and weather controls (X_{it}) and billing period-by-utility (τ_{it}) fixed effects. As a robustness check, we add a third difference with the Large Utility in the following framework:

$$\begin{aligned} \tilde{w}_{it} =& \alpha + \gamma_1 \text{Treat} + \gamma_2(\text{Treat} \times \text{Large}) + \gamma_3(\text{Treat} \times \text{Large} \times \text{Low}) + \\ & \theta_1 \text{Large} + \theta_2 \text{Low} + \theta_3(\text{Large} \times \text{Low}) + \beta X_{it} + \tau_t + \varepsilon_{it}, \end{aligned} \tag{9}$$

In this setup, γ_3 is our estimate of the PLE. The θ parameters $(\theta_1, \theta_2, \theta_3)$ control for lot size and utility-specific variation in the treatment period across both treatment and control groups.[5]

The *price-sensitivity effect* is how treatment induces differential responses to price changes. We exploit price changes over time and across the utilities in order to estimate the demand elasticity and then interact this with the treatment variables. Our demand elasticity regressions take the following form:

$$\begin{aligned} \ln(w_{it}) =& \beta_1 \ln(\hat{p}) + \beta_2(\ln(\hat{p}) \times \text{Treat}) + \beta_3(\ln(\hat{p}) \times \text{Period}) + \\ & \gamma_1 \text{Treat} + \theta_1 \text{Period} + \alpha_i + \tau_t + \varepsilon_{it} \end{aligned} \tag{10}$$

The *price sensitivity effect* is β_2, and this model allows us to test the price level hypothesis if $\beta_2 = 0$. Studies debate whether marginal or average price is the relevant price signal when consumers face increasing block rates (Nataraj and Hanemann, 2011; Ito, 2014; Wichman, 2014) so we model price as both average and marginal price.[6] The presence of

[5] One might wonder why we did not consider the utility boundary as a spatial regression discontinuity similar to Ito (2014). In our setting, water utility boundaries also serve as political boundaries that induce numerous other changes in tax rates, city regulations, and so forth, thus we did not believe the abrupt change in prices at utility borders would provide a viable identification strategy.

[6] We define average price as the volumetric proportion of the bill divided by quantity consumed that month.

increasing block rates also makes price endogenous because the marginal price the consumers faces depends on the quantity consumed. The model is therefore estimated using two-stage-least squares (2SLS) where price and the associated interactions are endogenous variables. Following Olmstead (2009) and Wichman et al. (2016), we instrument for the actual price the consumer faces (either marginal or average) with the full set of marginal prices from the rate structure. All marginal price instruments are transformed by natural logarithms. Therefore, our identification comes from variation in water rates set by the utility as opposed to changes in the households' consumption. The regressions include both household (α_i) and billing period (τ_i) fixed effects. We also estimate the demand model with interactions with utility indicators to capture utility-specific price elasticities.

4 Results and discussion

4.1 Baseline treatment and demand models

We first summarize our initial results from our baseline treatment effects. In Table 2, we present average treatment effects of home water reports. In the first two columns, our treatment effects for the Large and Small utilities are -4.5% and -3.5% reductions in water consumption due to randomized HWRs. In the third column, we pool both utilities, but allow for different treatment responses by including an interaction between our treatment variable and and an indicator for Large Utility. In the final column, we restrict the sample of the Large Utility to households within 10km of the Small Utility's border to ensure common support across both utilities. Overall, we find consistent evidence in line with previous research that HWRs reduce water consumption by 3–5% (Ferraro and Price, 2013; Brent et al., 2015).

Additionally, we present our initial demand specifications in Table 3. In the first two columns, we present naïve models using endogenous marginal and average price variables. As expected with increasing block-rate structures, we observe positive price elasticities. Our IV approach, in columns (3) and (4), performs comparatively better, providing sensible demand elasticities (-0.25 for MP and -0.17 for AP) well within the range of previous estimates (Dalhuisen et al., 2003). In the present analysis, we do not take a stand on whether average or marginal price responsiveness is the correct specification, rather we model them side-by-side. In columns (5) and (6), we restrict the sample to within 10km of the shared border, and the results are essentially the same.

Overall, our initial analysis produces estimates of average treatment effects for HWRs and estimates of price elasticities that are squarely within the results of previous studies. This consistency provides us with confidence that the experiments were conducted accurately and that our identification of price effects is valid.

Table 2: Baseline treatment effects

	(1) Large	(2) Small	(3) Both	(4) 10km
Treat	-0.045***	-0.035***	-0.032***	-0.035***
	(0.004)	(0.007)	(0.007)	(0.007)
Treat*Large			-0.014*	-0.017
			(0.008)	(0.011)
Observations	275,173	258,535	533,708	326,028
Households	26,104	19,119	45,223	25,602
Household FEs	–	–	–	–
Sample	Full	Full	Full	10km
Period-by-utility FEs	Y	Y	Y	Y

Notes: Dependent variable is average daily water consumption normalized by utility-specific control group consumption. All specifications control for evapotranspiration, precipitation, and pre-treatment water consumption. Robust standard errors are clustered at the household level. *p<0.1; **p<0.05; ***p<0.01

Table 3: Base demand models

	(1) MP	(2) AP	(3) MP	(4) AP	(5) MP	(6) AP
ln(MP)	0.549***		-0.246***		-0.278***	
	(0.007)		(0.012)		(0.017)	
ln(AP)		0.564***		-0.169***		-0.187***
		(0.008)		(0.011)		(0.016)
Observations	939,775	929,842	928,032	921,537	572,528	570,981
Households	43,133	43,132	43,124	43,120	25,990	25,987
Household FEs	Y	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y	Y
IV	Y	Y	Y	Y	Y	Y
Sample	Full	Full	Full	Full	10km	10km
First-stage F-stat			35,572	49,538	15,432	29,797

Note: Dependent variable is the natural log of average daily water consumption. All specifications control for evapotranspiration and precipitation. Prices are instrumented with full set of marginal prices from the utility rate schedule and associated interactions with exogenous variables. Robust standard errors are clustered at the household level. *p<0.1; **p<0.05; ***p<0.01

4.2 Price-level effects

We now turn to Table 4, in which we present our primary results of the price-level effect. Recall, the PLE in our setting is the interaction between the treatment effect of the HWR and the exogenous assignment of a higher marginal price via the lot-size discontinuity. We present results only for the Large Utility because the Small Utility does not have discontinuous changes in price due to lot-size (see Fig. 5). In our framework, the coefficient on the interaction Treat*Low is our estimate of the PLE. We vary the bandwidth (distance from

Table 4: Price-level effect: Large utility only

	(1) 1000sqft	(2) 750sqft	(3) 500sqft
Treat	-0.050***	-0.050***	-0.043***
	(0.008)	(0.009)	(0.010)
Treat*Low	0.003	-0.006	-0.006
	(0.011)	(0.013)	(0.015)
Observations	124,067	97,015	70,068
Households	12,302	9,607	6,920
Sample	Large only	Large only	Large only
Period-by-utility FEs	Y	Y	Y

Notes: Dependent variable is average daily water consumption normalized by utility-specific control group consumption. All specifications control for evapotranspiration, precipitation, and pre-treatment water consumption. Robust standard errors are clustered at the household level. *p<0.1; **p<0.05; ***p<0.01

the lot-size discontinuity) in each of the columns. For all bandwidths, we estimate a precise zero effect. In our narrowest bandwidth, the interacted coefficient is -0.006 with a standard error (robust to within-household correlation) of 0.015. This estimate is based on a sample of nearly 7000 households totaling more than 70,000 observations, thus we have sufficient power to identify a statistically meaningful effect. We do not, however, find any evidence of a PLE with a magnitude that is economically meaningful.

It is possible that our PLE estimate in Table 4 is biased because we pool all lot-size discontinuities in the rate structure together. In Table 5, we estimate PLE effects for each discontinuity separately, again for the Large Utility only. For the 7500 and 11,000 sq. ft. discontinuities in panels (a) and (b), we again find precisely estimated null effects, with standard errors increasing slightly with smaller bandwidths. For the larger discontinuity at 17,500 sq. ft., we observe both a substantially larger base treatment effect ($9-12\%$ reductions in daily consumption) and a larger PLE within 500 sq. ft. of the discontinuity. The PLE, however, is estimated with large confidence intervals, due in part to the smaller amount of households near this discontinuity.

We include several additional analyses to support our results. In Table A.1, we re-run our primary PLE specifications but we include interaction terms with lot-size, as is typical in RD designs. Results are virtually unchanged: we find precisely estimated null effects for the PLE. Additionally, we perform a falsification test in the Large Utility at false discontinuities of 9000 sq. ft. and 13,000 sq. ft. We choose these thresholds because they are near our true thresholds without overlapping at the largest bandwidths (1000 sq. ft.). These falsification tests examines whether our lot size thresholds would partially pick up the smaller treatment effects (in absolute value), associated with smaller lots that use less water. If the true PLE

Table 5: Price-level effect: Large utility only, no lot size interactions, at individual discontinuities

(a) 7500 sq. ft. discontinuity

	(1) 1000sqft	(2) 750sqft	(3) 500sqft
Treat	-0.041***	-0.038***	-0.036***
	(0.009)	(0.010)	(0.011)
Treat*Low	0.000	-0.009	-0.008
	(0.012)	(0.014)	(0.016)
Observations	79,351	61,908	43,096
Households	7,934	6,181	4,290
Sample	Large only	Large only	Large only
Period-by-utility FEs	Y	Y	Y

(b) 11,000 sq. ft. discontinuity

	(1) 1000sqft	(2) 750sqft	(3) 500sqft
Treat	-0.041***	-0.043***	-0.036**
	(0.015)	(0.016)	(0.018)
Treat*Low	-0.005	-0.011	-0.003
	(0.024)	(0.028)	(0.033)
Observations	32,471	25,263	19,719
Households	3,180	2,472	1,924
Sample	Large only	Large only	Large only
Period-by-utility FEs	Y	Y	Y

(c) 17,500 sq. ft. discontinuity

	(1) 1000sqft	(2) 750sqft	(3) 500sqft
Treat	-0.123***	-0.136***	-0.092**
	(0.037)	(0.041)	(0.043)
Treat*Low	0.022	0.021	-0.024
	(0.053)	(0.061)	(0.065)
Observations	11,855	9,517	7,002
Households	1,150	922	681
Sample	Large only	Large only	Large only
Period-by-utility FEs	Y	Y	Y

Notes: Dependent variable is average daily water consumption normalized by utility-specific control group consumption within 1,000 sq. ft. of the lot-size discontinuity. All specifications control for evapotranspiration, precipitation, and pre-treatment water consumption. Robust standard errors are clustered at the household level. *p<0.1; **p<0.05; ***p<0.01

is negative (households who face higher prices are more responsive to HWRs) the small lot size effect will bias our estimates of the PLE towards zero. These results are presented in Table A.2. Here again, we find statistical zeros, and the point estimates switch between positive and negative values.

In an additional robustness check, we add a third difference to our difference-in-discontinuity design because it is possible that response to HWRs is greater for households with higher consumption levels, which is correlated positively with lot size (our running variable in the regression discontinuity). To implement the third difference, we estimate Equation 10 on a sample including both utilities. We present these results in Table 6. In these specifications, the PLE is the coefficient on Treat*Low*Large, or, the marginal change in the treatment effect effect due to facing an exogenously higher marginal price by being just below the lot-size threshold relative to similar households in the small utility who face no price discontinuity. In these specifications, we find no statistical evidence of a PLE, which is a precisely estimated zero for bandwidths of 1000 sq. ft. and 750 sq. ft. Within 500 sq. ft. of the lot-size threshold, however, we find an economically large PLE of -0.057. Although not significant at the $p < 0.1$ level, this estimate is larger in magnitude than our initial treatment effect. Because we did not see such an effect in our Large Utility Only RD (Table 2), we suspect that this is an artifact of unexplained increases in consumption in the Small Utility near the lot-size threshold.

Table 6: Price-level effect: Full sample, no lot size interactions

	(1) 1000sqft	(2) 750sqft	(3) 500sqft
Treat	-0.054***	-0.062***	-0.077***
	(0.019)	(0.022)	(0.028)
Treat*Large	0.003	0.011	0.034
	(0.021)	(0.024)	(0.030)
Treat*Low	0.006	0.013	0.051
	(0.026)	(0.031)	(0.039)
Treat*Low*Large	-0.003	-0.018	-0.057
	(0.029)	(0.033)	(0.041)
Observations	194,802	149,703	106,029
Households	17,752	13,658	9,686
Sample	Full	Full	Full
Period-by-utility FEs	Y	Y	Y

Notes: Dependent variable is average daily water consumption normalized by utility-specific control group consumption. All specifications control for evapotranspiration, precipitation, and pre-treatment water consumption. Robust standard errors are clustered at the household level. *p<0.1; **p<0.05; ***p<0.01

We explore this result further in Figure 8, in which we plot the base treatment effect coefficients interacted with 250 sq. ft. lot-size bins near the lot-size thresholds. We do so for both utilities. Evidence of a nonzero PLE would be revealed by a discontinuous jump in treatment effect estimates at the lot-size thresholds (represented by solid red vertical lines). Specifically, in the presence of a signficant PLE we expect the treatment effect immediately to the left of the threshold to be larger in magnitude than the treatment effect immediately to the right of the threshold. For the 7500 sq. ft. discontinuity, we observe the treatment effect move smoothly through the discontinuity for both utilities. All estimates are statistically similar, shown by overlapping 95% confidence intervals. The results for the 11,000 and 17,500 sq. ft. thresholds are noisier: point estimates jump around a bit more, but confidence intervals also overlap for all estimates within a utility. The majority of households included in our RD samples are located near the 7500 sq. ft. threshold.

To recap, we find no statistical evidence that exogenously assigned differences in marginal prices affect the responsiveness of HWRs. This result is somewhat surprising because the HWRs make the private economic benefits of water conservation more salient (e.g., bottom panel in Figure 3). HWRs provide cost-savings information that consumers might expect from changing behavior or technology. Consumers just above/below the price discontinuities we use for identification would thus face nontrivial differences in expected cost-savings despite being otherwise similar types of households, but we observe no statistical different in their response to HWRs. Our analysis thus far suggests that the primary mechanism for the HWR operates through channels of increasing (the salience of) the moral costs of water consumption.

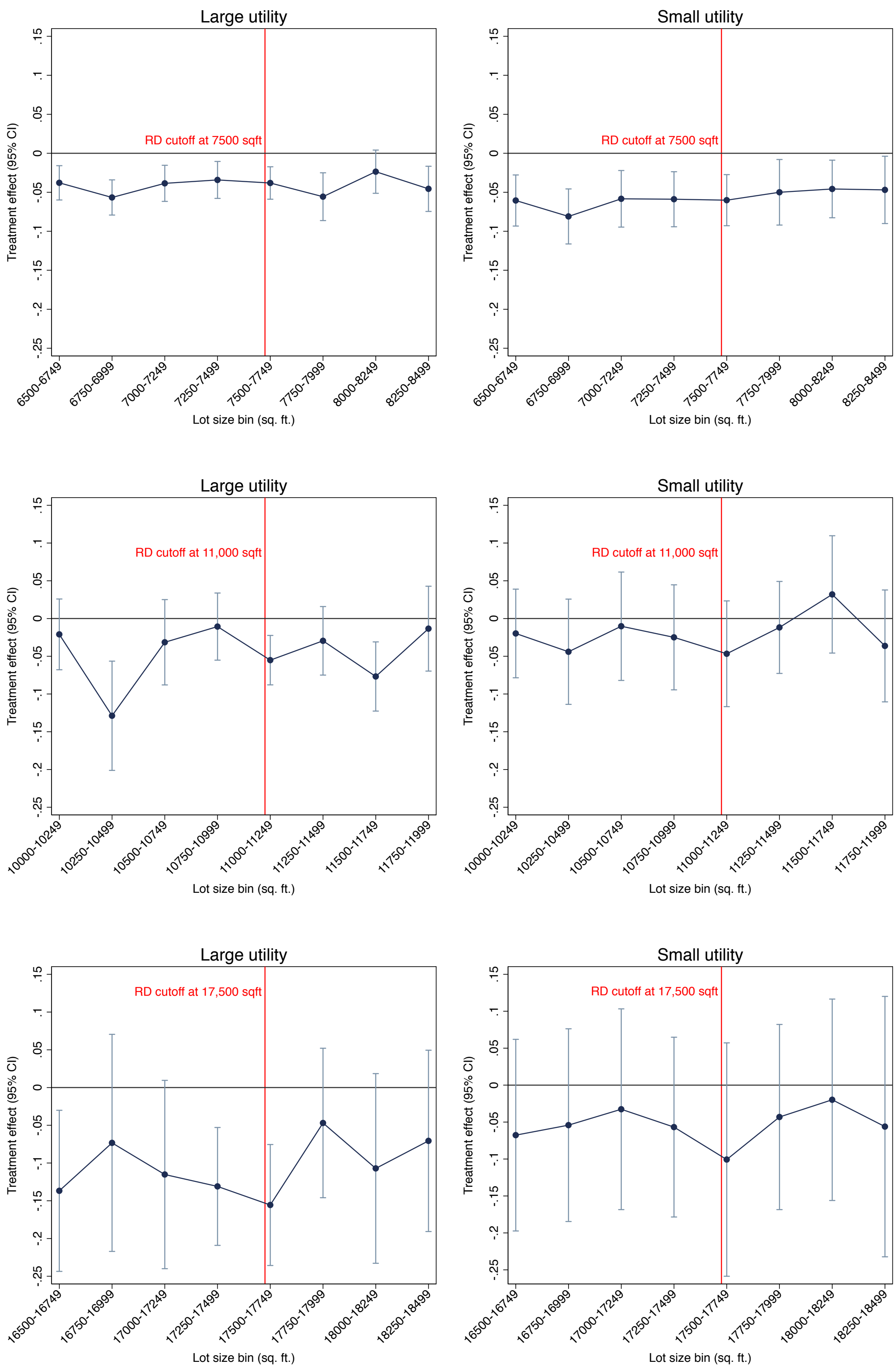


Figure 8: Treatment effects by lot-size bin

4.3 Price-sensitivity effects

Lastly, we present results for the PSE in Table 7. Recall, the PSE is the degree to which HWRs increase consumers' price sensitivity, e.g., by making the costs of consumption more salient. Identification of this effect is straightforward: we estimate price elasticities of water demand equation as in Equation 10 and interact our price variables with the randomized HWR treatment. The resulting coefficient on that interaction is the PSE.

In columns (1) and (2) of Table 7, we report PSE estimates for our pooled sample. We find statistically significant evidence that HWRs increase price sensitivity under the marginal price, but not the average price, demand specification. The PSE increases price sensitivity by approximately 13% for the MP specification. We might be concerned, that in addition to price variation, there is significant variation in unobservables across utilities that could impact price elasticity. In an attempt to control for these cross border differences, we restrict the sample to households within 10km of the border. Columns (3) and (4) present the PSE estimates in the restricted sample, and we find no evidence of a signficant PSE in either the MP or AP specification. These results are less decisive than our PLE results, but still do not find conclusive evidence that social comparisons have meaningful interactions with prevailing economic incentives.

Table 7: Price-sensitivity effect

	(1) MP	(2) AP	(3) MP	(4) AP
Treat	-0.035***	-0.042***	-0.041***	-0.044***
	(0.003)	(0.003)	(0.005)	(0.005)
ln(MP)	-0.221***		-0.236***	
	(0.011)		(0.015)	
ln(MP)*Treat	-0.030**		-0.012	
	(0.014)		(0.020)	
ln(AP)		-0.191***		-0.189***
		(0.010)		(0.014)
ln(AP)*Treat		-0.002		0.002
		(0.012)		(0.016)
Observations	928,032	921,537	572,528	570,981
Households	43,124	43,120	25,990	25,987
Household FEs	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y
IV	Y	Y	Y	Y
Sample	Full	Full	10km	10km
First-stage F-stat	11,253	13,298	4,819	6,034

Note: Dependent variable is the natural log of average daily water consumption. All specifications control for evapotranspiration and precipitation. Prices are instrumented with full set of marginal prices from the utility rate schedule and associated interactions with exogenous variables. Interactions with indicators for treatment periods are included but coefficients are not reported for clarity. Robust standard errors are clustered at the household level. *p<0.1; **p<0.05; ***p<0.01

5 Concluding remarks

Behavioral nudges do not exist in a vacuum. While the randomized deployment of many behavioral nudges produces strong internal validity for the estimation of causal effects, as Allcott (2015) shows the treatment effects of any location may be a function of the underlying characteristics of the specific population. In order to ensure that the estimates from any one location are externally valid, it is critical to identify the sources of heterogeneity and adjust the magnitude based on the characteristics of the target population. This is challenging when the entire experimental sample faces the same set of existing policies. We focus on how variation in prevailing water prices affects consumer responsiveness to HWRs aimed at water conservation that include social comparisons. We do not find any evidence that the response to this prevalent behavioral nudge has any meaningful interactions with underlying water rates.

In addition to external validity, our results have implications for the behavioral mechanisms through which nudges operate. Finding no evidence of heterogeneity due to different private benefits of conservation leads us to conclude that consumers are not primarily responding to social comparisons due to private financial motivations. This has important implications for the welfare effect of nudges as shown by Allcott and Kessler (Forthcoming). If nudges are essentially just a moral tax, they will only be welfare enhancing if the social cost of energy/water exceeds the current private costs. This suggests a re-thinking of behavioral policies that specifically target welfare as opposed to simply changing behavior. Given substantial evidence of behavioral biases in energy and water markets (Allcott and Wozny, 2014; Sexton, 2015; Wichman, 2017; Brent and Ward, Forthcoming, 2018) it is worthwhile to find ways to promote pro-social behavior that also improves private decisions.

References

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

—, "Site selection bias in program evaluation," *The Quarterly Journal of Economics*, 2015, *130* (3), 1117–1165.

— **and Dmitry Taubinsky**, "Evaluating Behaviorally-Motivated Policy: Experimental Evidence from the Lightbulb Market," *American Economic Review*, 2015, *105* (8), 2501–2538.

— **and Judd B. Kessler**, "The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons," *American Economic Journal: Applied Economics*, Forthcoming.

— **and Nathan Wozny**, "Gasoline prices, fuel economy, and the energy paradox," *Review of Economics and Statistics*, 2014, *96* (5), 779–795.

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

Brandon, Alec, John A. List, Robert D. Metcalfe, Michael K. Price, and Florian Rundhammer, "Testing for crowd out in social nudges: Evidence from a natural field experiment in the market for electricity," *Proceedings of the National Academy of Sciences*, 2018, p. 201802874.

Brent, Daniel A and Michael B Ward, "Price Perceptions in Water Demand," Technical Report 2018.

— **and** —, "Energy Efficiency and Financial Literacy," *Journal of Environmental Economics and Management*, Forthcoming.

Brent, Daniel A., Joseph Cook, and Skylar Olsen, "Social comparisons, household water use and participation in utility conservation programs: Evidence from three randomized trials," *Journal of the Association of Environmental and Resource Economists*, 2015, *2* (4), 597–627.

Croson, Rachel and Jen Yue Shang, "The impact of downward social information on contribution decisions," *Experimental Economics*, 2008, *11* (3), 221–233.

Dalhuisen, Jasper M., Raymond J.G.M. Florax, Henri L.F. De Groot, and Peter Nijkamp, "Price and income elasticities of residential water demand: a meta-analysis," *Land Economics*, 2003, *79* (2), 292–308.

Ferraro, Paul J. and Michael K. Price, "Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment," *The Review of Economics and Statistics*, 2013, *95* (1), 247–264.

—**, Juan Jose Miranda, and Michael K. Price**, "The Persistence of Treatment Effects with Norm-Based Policy Instruments: Evidence from a Randomized Environmental Policy Experiment," *American Economic Review, Papers and Proceedings*, 2011, *101* (3), 318–322.

Hallsworth, Michael, John A. List, Robert D. Metcalfe, and Ivo Vlaev, "The behavioralist as tax collector: Using natural field experiments to enhance tax compliance," *Journal of Public Economics*, 2017, *148*, 14–31.

Hanks, Andrew S., David R. Just, Laura E. Smith, and Brian Wansink, "Healthy convenience: nudging students toward healthier choices in the lunchroom," *Journal of Public Health*, 2012, *34* (3), 370–376.

Ito, Koichiro, "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing," *American Economic Review*, 2014, *104* (2), 537–563.

_ , Takanori Ida, and Makoto Tanaka, "Moral suasion and economic incentives: Field experimental evidence from energy demand," *American Economic Journal: Economic Policy*, 2018, *10* (1), 240–67.

Jacobsen, Grant D., "Do energy prices influence investment in energy efficiency? Evidence from energy star appliances," *Journal of Environmental Economics and Management*, 2015, *74*, 94–106.

Levitt, Steven D. and John A. List, "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?," *Journal of Economic Perspectives*, June 2007, *21* (2), 153–174.

_ , _ , Susanne Neckermann, and Sally Sadoff, "The behavioralist goes to school: Leveraging behavioral economics to improve educational performance," *American Economic Journal: Economic Policy*, 2016, *8* (4), 183–219.

List, John A. and Anya Savikhin Samek, "The behavioralist as nutritionist: leveraging behavioral economics to improve child food choice and consumption," *Journal of health economics*, 2015, *39*, 135–146.

_ , Robert D. Metcalfe, Michael K. Price, and Florian Rundhammer, "Harnessing Policy Complementarities to Conserve Energy: Evidence from a Natural Field Experiment," Technical Report, National Bureau of Economic Research 2017.

Nataraj, Shanthi and W. Michael Hanemann, "Does marginal price matter? A regression discontinuity approach to estimating water demand," *Journal of Environmental Economics and Management*, March 2011, *61* (2), 198–212.

Olmstead, Sheila M., "Reduced-form versus structural models of water demand under nonlinear prices," *Journal of Business & Economic Statistics*, 2009, *27* (1), 84–94.

Royer, Heather, Mark Stehr, and Justin Sydnor, "Incentives, commitments, and habit formation in exercise: evidence from a field experiment with workers at a fortune-500 company," *American Economic Journal: Applied Economics*, 2015, *7* (3), 51–84.

Sallee, James M., "Rational Inattention and Energy Efficiency," *Journal of Law and Economics*, 2014, *57* (3), 781–820.

Sexton, Steven E., "Automatic bill payment and salience effects: Evidence from electricity consumption," *The Review of Economics and Statistics*, 2015, *97* (2), 229–241.

Shang, Jen and Rachel T.A. Croson, "A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods," *Economic Journal*, 2009, *119* (540), 1422–1439.

Taylor, Michael H., Kimberly Rollins, and Corey Lott, "Exploring the behavioral and welfare implications of social-comparison messages in residential water and electricity," *Economics Letters*, 2018, *168*, 65–69.

Wichman, Casey J., "Perceived price in residential water demand: Evidence from a natural experiment," *Journal of Economic Behavior & Organization*, 2014, *107*, 308–323.

Wichman, Casey J, "Information provision and consumer behavior: A natural experiment in billing frequency," *Journal of Public Economics*, 2017, *152*, 13–33.

Wichman, Casey J., Laura O. Taylor, and Roger H. von Haefen, "Conservation policies: Who responds to price and who responds to prescription?," *Journal of Environmental Economics and Management*, 2016, *79*, 114 – 134.

Appendix: Additional Results

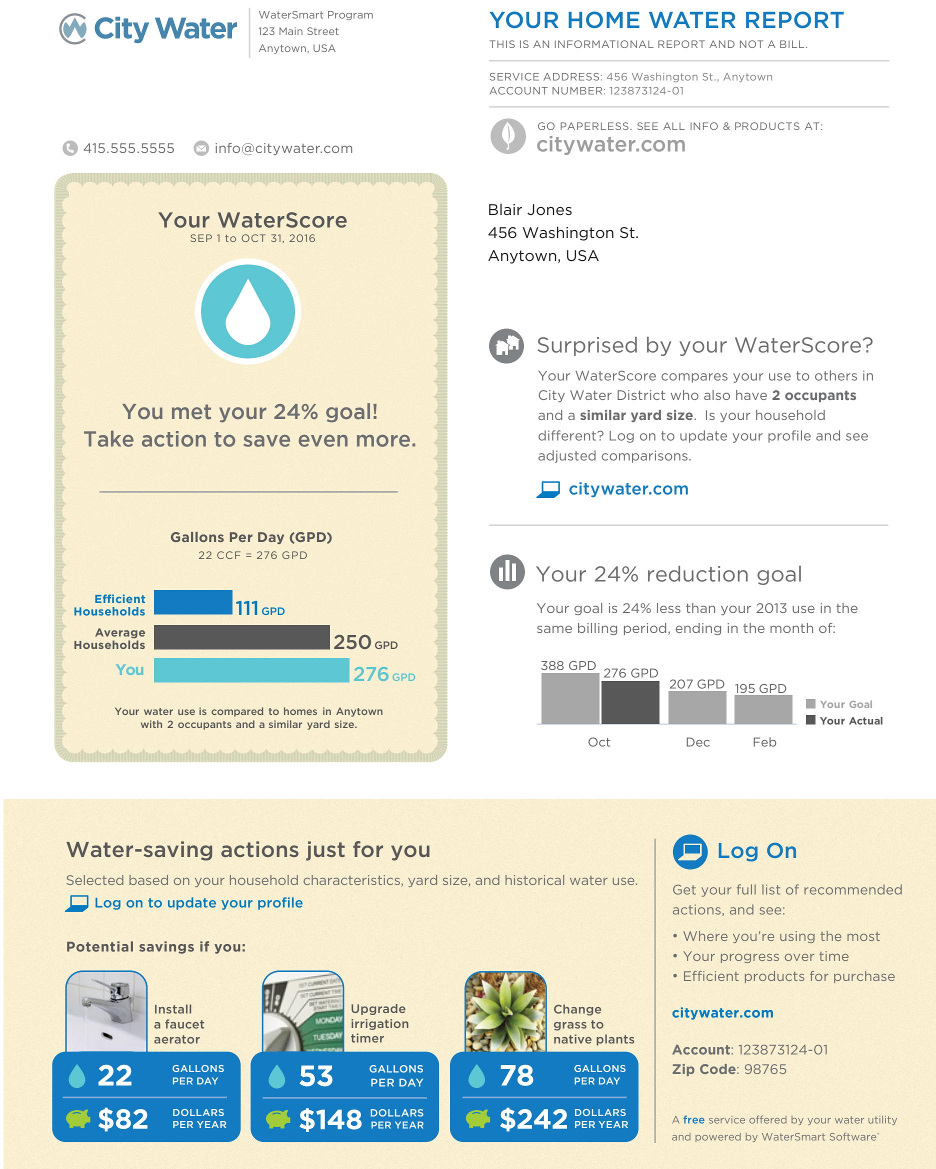


Figure A.1: Home Water Report

Note: This is an example of a generic "Yellow" Home Water Report (HWR). Households receiving this report used less water than their peer group average.

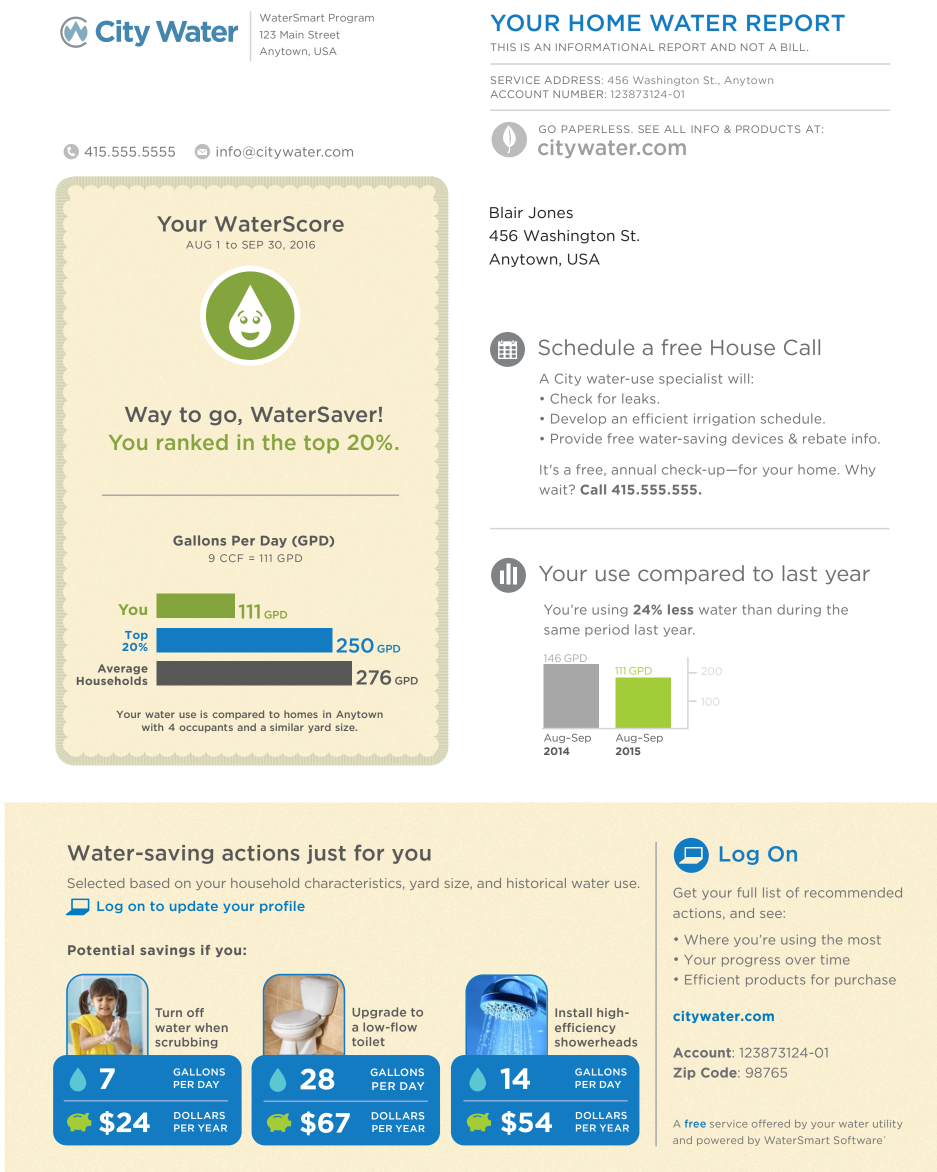


Figure A.2: Home Water Report

Note: This is an example of a generic "Green" Home Water Report (HWR). Households receiving this report were in the bottom 20% of their peer group.

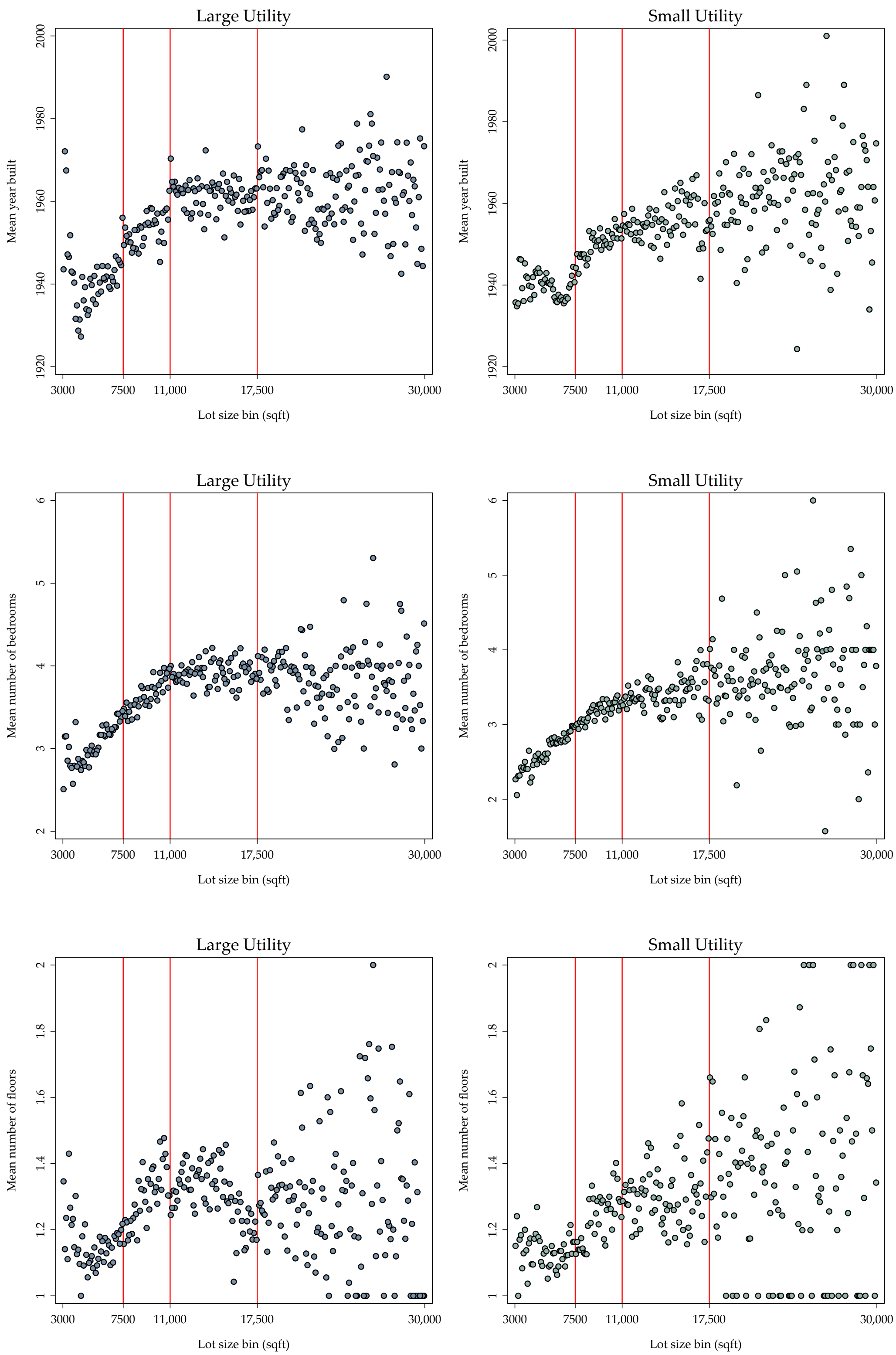


Figure A.3: Additional covariate distributions across lot-size thresholds for both utilities

Table A.1: Price-level effect: Large utility only, with lot-size interactions

	(1)	(2)	(3)
	1000sqft	750sqft	500sqft
Treat	-0.050***	-0.050***	-0.042***
	(0.008)	(0.009)	(0.010)
Treat*Low	-0.002	-0.012	-0.011
	(0.014)	(0.015)	(0.017)
Treat*Sq.ft.	-0.000	-0.017	0.013
	(0.039)	(0.044)	(0.053)
Treat*Low*Sq.ft.	-0.041	-0.049	-0.065
	(0.065)	(0.067)	(0.081)
Observations	124,067	97,015	70,068
Households	12,302	9,607	6,920
Sample	Large only	Large only	Large only
Period-by-utility FEs	Y	Y	Y

Notes: Dependent variable is average daily water consumption normalized by utility-specific control group consumption. All specifications control for evapotranspiration, precipitation, and pre-treatment water consumption. Robust standard errors are clustered at the household level. *p$<$0.1; **p$<$0.05; ***p$<$0.01

Table A.2: Price-level effect falsification test: Large utility only, no lot size interactions, at individual discontinuities

(a) False 9000 sq. ft. discontinuity

	(1)	(2)	(3)
	1000sqft	750sqft	500sqft
Treat	-0.035**	-0.043**	-0.051***
	(0.016)	(0.017)	(0.019)
Treat*Low	-0.006	0.003	0.022
	(0.020)	(0.022)	(0.027)
Observations	37,810	27,019	17,709
Households	3,763	2,691	1,758
Sample	Large only	Large only	Large only
Period-by-utility FEs	Y	Y	Y

(b) False 13,000 sq. ft. discontinuity

	(1)	(2)	(3)
	1000sqft	750sqft	500sqft
Treat	-0.072**	-0.080**	-0.080*
	(0.029)	(0.033)	(0.041)
Treat*Low	-0.011	-0.009	0.004
	(0.040)	(0.046)	(0.059)
Observations	15,963	11,498	7,415
Households	1,577	1,139	734
Sample	Large only	Large only	Large only
Period-by-utility FEs	Y	Y	Y

Notes: Dependent variable is average daily water consumption normalized by utility-specific control group consumption within 1,000 sq. ft. of the lot-size discontinuity. All specifications control for evapotranspiration, precipitation, and pre-treatment water consumption. Robust standard errors are clustered at the household level. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$