Peer Effects in Water Conservation: Evidence from Consumer Migration*

Bryan Bollinger
Duke University

Jesse Burkhardt
Colorado State University

Kenneth Gillingham
Yale University

February 5, 2018

Abstract

Social interactions are widely understood to influence consumer decisions in many choice settings. This paper identifies causal peer effects in water conservation utilizing variation from consumer migration. After using a machine learning approach to classify extremely high-resolution remote sensing data, we show that the water conservation effects can be attributed to peer effects in the diffusion of dry landscaping. We provide evidence that without a price signal, peer effects are muted, highlighting an important relationship between information transmission and prices. These results inform water use policy in many areas of the world threatened by recurring drought conditions.

Keywords: social interactions; diffusion; information; water policy.
JEL classification codes: Q42, Q48, L13, L25, O33, O25.

---

*Bollinger: Fuqua School of Business, Duke University, 100 Fuqua Drive, Durham, NC 27708, phone: 919-660-7766, e-mail: bryan.bollinger@duke.edu. Burkhardt: Department of Agricultural and Resource Economics, Colorado State University, 1200 Center Ave. Mall, Fort Collins, CO 80523, phone: 503-312-8943, email: jesse.burkhardt@colostate.edu. Gillingham: School of Forestry & Environmental Studies, Department of Economics, School of Management, Yale University, 195 Prospect Street, New Haven, CT 06511, phone: 203-436-5465, e-mail: kenneth.gillingham@yale.edu. The authors are grateful for conversations and comments from Kerry Smith, Matt Kotchen, Sheila Olmstead, and Michael Hanemann, Soe Myint, and from seminar participants at Arizona State, Northwestern, and UCLA. The authors would also especially like to thank the staff at the Phoenix Metropolitan Water District, including Doug Frost, Darren Sversvold, and Jamie Campbell.

1 Introduction

Social interactions have been shown to play a pivotal role in the diffusion of many new technologies and practices, and undergird classic economic models of technology diffusion (Griliches 1957; Bass 1969; Rogers 1995). The idea that individuals learn from their peers, neighbors, or friends to adopt behaviors or technologies has been explored in settings ranging from agriculture (Foster and Rosenzweig 1995; Conley and Udry 2010) to foreclosures (Towe and Lawley 2013) to schooling (Sacerdote 2001; Graham 2008). In the environmental realm, such 'peer effects' have been shown in the adoption of solar photovoltaic panels (Bollinger and Gillingham 2012; Graziano and Gillingham 2015) and hybrid vehicles (Narayanan and Nair 2013; Heutel and Muehlegger 2015).

This paper is the first to identify causal peer effects in water consumption and landscape choice. Our identification strategy relies on quasi-experimental variation from consumer migration in which new households move into peer groups and make water consumption and landscape changes. We use water billing data from over 300,000 households in Phoenix, one of the hottest and driest cities in the United States, along with housing transaction data to find that a 10% reduction in the mean of peer water consumption leads to a 1.7% (207 gallon per month) reduction in a household's water consumption, where peers are defined as other households within a 500′ radius of the household. To understand the mechanism underlying this result, we use a machine learning approach to classify the greenness of a household's landscaping using remote sensing data. We show that the peer effects in water consumption are driven almost entirely by changes in landscaping. We also present novel evidence that a non-negligible price is necessary for these peer effects to influence consumer behavior by examining consumers eligible for extremely highly discounted Salt River Project irrigation water. Finally, we explore heterogeneity in these effects to provide guidance for targeted policy.

Access to water is a major issue in many locations around the world. Over the 20th century, global water use experienced a sixfold increase and by 2025, the U.N. Environmental Program estimates that over two-thirds of the world's population will live in water-stressed regions. Concerns about water availability are equally important in many regions of the United States. For example, the California droughts of 2012 and 2015 led to combined economic losses of approximately $5.2 billion, and only a few years earlier, droughts in the Southwest and Midwest led to losses of $20 billion.[1] The U.S. Environmental Protection Agency predicts up to a 40% decrease in snow runoff and soil moisture in parts of the Western United States by 2050, further exacerbating concerns about droughts.[2]

Residential water demand is of further interest due to rapid changes in both technologies and practices. For example, between 1990 and 2010, the diffusion of higher water efficiency appliances and a shift towards dry landscaping reduced per capita residential water use by roughly 10%.[3] This decline in per capita water usage is a positive development for concerns about water availability, but can pose challenges for municipalities

[1]See https://www.ama.org/publications/eNewsletters/Marketing-News-Weekly/Pages/economic-loss-us-droughts.aspx and http://www.cnbc.com/2015/03/03/california-drought-seen-having-worsening-3-billion-economic-impact-in-2015.html

[2]See https://19january2017snapshot.epa.gov/climate-impacts/climate-impacts-southwest_.html

[3]See https://water.usgs.gov/watuse/data/2010/index.html

that designed water systems for greater demand. Thus, understanding residential water demand and the diffusion of technologies and practices that influence demand is of major interest to policymakers.

Causal peer effects can be a major influencer of the diffusion of water-saving practices, but it is well understood that these effects are challenging to identify. Challenges include endogenous group formation leading to issues of self-selection of peers, correlated unobservables, and simultaneity (Manski 1993; Brock and Durlaf 2001; Moffitt 2001; Hartmann et al. 2008). This study addresses endogenous group formation using a rich set of time-varying geographic fixed effects. We address the other concerns using an instrumental variables strategy based on exogenous shocks to the landscape from consumer migration to a neighborhood. New residents to a particular house often change the landscaping of that property within a short period of time after the move. At the same time, there is a broader trend towards dry landscaping occurring in Phoenix, which we document. Thus, we observe a much higher probability of a switch from green to dry landscaping just after a housing transaction. Our identification strategy therefore uses shocks in consumer migration as an instrument for the average water consumption and landscape decisions of a households peer group, while focusing on the variation remaining after inclusion of time-varying geographic fixed effects.

This work contributes to several literatures. Most directly, it adds new, well-identified evidence to the large and growing literature on peer effects in the diffusion of consumer behaviors. In addition, it adds to the literature on how information transmission, in our case through social learning, can influence consumer decisions about energy and water use. Several papers explore how social norm-based messages aimed at energy conservation can reduce energy use (e.g., Allcott 2011; Allcott and Rogers 2014; Ayres et al. 2012; Costa and Kahn 2013; Dolan and Metcalfe 2015; Tsvetanov and Gillingham 2017) or how prosocial appeals influence energy use relative to economic incentives (e.g., Reiss and White 2008; Ito et al. 2017; Burkhardt et al. 2017).

Our paper is the first to explore how consumer behavior change due to social interactions is directly influenced by economic incentives. Jessoe and Rapson (2014) show that households are three standard deviations more responsive to temporary price increases when provided with high frequency information on electricity usage, and Dolan and Metcalfe (2015) show that the effect of financial incentives can disappear when information on social norms is provided. In the context of residential water demand, Ferraro and Price (2013) show that social comparison messages are the most effective among the least price sensitive households. However, in these studies, information transmission is exogenously manipulated. In contrast to these previous findings, we show that social interactions–a phenomenon involving *endogenous* information transmission–is instead muted in the absence of a non-negligible price signal.

Our findings contribute to active policy discussions. Water districts make large upfront infrastructure investments and face continual planning challenges. Understanding the speed and pattern of the diffusion of water conservation activities, such as a transition from green to dry landscaping, is immediately valuable for water planning purposes. An understanding of such diffusion can also enable better targeting of policies. Neighborhoods that begin with some landscape transitions in a given year may be expected to experience contagion from these first transitions, and observe a greater number of transitions in the future relative to other neighborhoods. One-time targeted information campaigns or subsidies for dry landscaping may be useful in fast-growing areas that are straining the water system, while such efforts may be best withheld from areas that have too much water system capacity. Furthermore, even more refined targeted campaigns may be possible. For example, we find that pet owners and households near parks display minimal peer effects in landscaping choices (perhaps due to a high preference for green space) and thus may be less responsive to direct incentives, and less likely to exhibit social spillovers through peer interventions.

The remainder of the paper is organized as follows. In the next section we provide some background on water use in Arizona and describe our unique dataset, which combines remote sensing images, water bills, and housing transaction data. Section 3 presents the model and discusses our identification strategy. Section 4 describes the results and mechanisms underlying the peer effect. Section 5 concludes.

2 Background and Data

2.1 Background on Water Use in Arizona

Residential water in Phoenix comes from three sources: wells owned by individuals or the city, the Salt and Verde rivers, and the Colorado river.[4] These sources are constrained and access to them is politically contentious. Moreover, between 1980 and 2010, the population of Phoenix increased by approximately 83%, leading to an increase in residential water use of 23%.[5]

Fortunately for Phoenix and many arid regions of the United States, while water use may be increasing, water use per household has been trending downwards over time. Figure 1 shows the average water use per household in Phoenix between 1986 and 2012. While there was a slight increase in the late 1990's, there is a clear downward trend throughout the entire time period. The broader decline in per household water use is primarily attributed to improved water efficiency of appliances, such as low-flow toilets and shower heads or water-efficient washers/driers and dishwashers. Yet there is a growing consensus among practitioners that the low hanging fruit in improvements in indoor water efficiency has mostly been picked (CUWA 2017).

More recently, water districts have thus been focusing on behavior change and outdoor water use. Water conservation campaigns have become popular (Ferraro and Price 2013), and some areas have changed water pricing schedules. Promoting a shift to dry

[4] Phoenix receives only 8 inches of rain a year on average. For details see http://archive.azcentral.com/members/Blog/ShaunMcKinnon/123051

[5] See https://www.biggestuscities.com/city/phoenix-arizona

landscaping, with less grass and more drought-tolerant plants, is another option that has been adopted in numerous arid water districts, including several in the Phoenix metro area. Furthermore, in some cases this transition to dry landscaping is occurring without policy action. As we will see shortly, Phoenix has seen just such a trend towards dry landscaping in recent years.

2.2 Data

The foundation of our data set is monthly water billing data from the City of Phoenix water department for all single-family households served by the water department between 2004 and 2012 (308,652 households). These contain the address of the land parcel, allowing us to geocode all parcels. They also contain the total water consumption for each household, with the exception of parcels that are eligible to receive a defined amount of extremely reduced-cost flood irrigation water from the Salt River Project (SRP), a relic from previous agricultural uses. SRP-eligible households only pay a small fee for the water delivery; currently, this fee is on the order of $5 per month for most households and has not appreciably changed in recent years. Relative to the cost of city water, this SRP water is nearly free. However, the SRP flood irrigation water is only suitable for outdoor uses and households must construct berms to direct the water.[6] For this reason, not all eligible houses use SRP water, and eligible houses must supplement SRP water with municipal water. We do not observe the consumption of SRP irrigation water, but from

[6]SRP water is provided to households based on plot size and is distributed in 45 minute increments. We find that annually, off-project houses use approximately 12 cubic feet of municipal water more than on-project houses on average, but this difference is not statistically significant once lot size is controlled for.

conversations with the water department we were able to verify that SRP water provision did not change over the time frame of our study. Roughly 43% of households in the city of Phoenix Water District are within the boundaries of the SRP.

We complement the water billing data with remote sensing images provided by the City of Phoenix in order to develop a measure of landscape greenness. These images, taken once a year in the fall, have resolutions ranging from 0.3 to 0.8 feet. All years are available at high resolution except 2010, which is only available at a lower resolution. Therefore, our remote sensing dataset spans 2004-2012, excluding 2010. Each image has three color bands (red, green, and blue). There are standard software packages for developing vegetation indices, such as the Normalized Difference Vegetation Index, but our remote sensing images lack the infrared band needed for such standard packages. Therefore, we used a machine learning approach to train the computer to find green pixels through a series of iterations. The approach we used is a supervised maximum likelihood classification approach developed by the company Imagine Software, Inc. To verify the algorithm, we compared the machine learning results to hand-coded results.[7] In comparing the results of machine learning approach to the hand-coded pixels, we found that the machine learning approach achieved an accuracy of approximately 85-90%.[8]

Figure 2 provides an example of the output of the classification process. The photo on the left shows several randomly chosen parcels while the photo on the right shows the same parcels with the pixels the computer designates as green landscaping, including

[7]Specifically, the City of Phoenix had a group of interns visually estimate the percentage of turf and greenery in 30,293 parcels.

[8]Furthermore, the means of the machine learning and hand coded parcels are not statistically different from one another using a simple t-test of differences in means.

tree tops, grasses, and other vegetation, highlighted in green. Figure 3 shows a map of Phoenix with the water department territory, Salt River project territory, and the remote sensing subsample of 71,477 parcels.

As is common in remote sensing work, further data cleaning is necessary to compare remote sensing images taken at different times. For instance, images may contain more or less shadow and more or less haze depending on when they were taken. The cameras or camera settings used may have also been slightly different in each photo. Thus, we systematically adjust the remote sensing data by focusing on the variation in the landscape greenness measure that can be explained by the water consumption data. Specifically, we regress landscape greenness on household water consumption in all 12 months of the year and use the fitted values of landscape greenness, thus allowing us to focus on the relevant variation in landscape greenness for our analysis (see Appendix A for further details on this data adjustment). Not surprisingly, landscape greenness is highly correlated with water consumption for nearly all months of the year, and especially for the drier months of the year. Figure 4 plots the average percentage greenness in the landscape in our cleaned data over time, indicating the same clear downward trend as in the water consumption data.

Our next data source is the Maricopa County Assessor's Office, which provided data on housing sales and other other physical housing characteristics including pool size, lot size, construction date, home size, and garage size. Importantly, we observe the date on which the housing transactions occur and the address of the parcel, allowing us to match these data with the previous data sets.

For both the water billing data and the landscape data, we geocode all addresses. To create a measure of peer decisions, we create a radius around the center of each household's parcel. In our primary specifications, we use a radius of 500 feet, but we also explore different radii in our robustness checks. We then create variables for the average water use and average landscape greenness for all surrounding households within the 500-foot radius. Using the migration data, we also create a variable for the fraction of households sold within that radius and during that year. Table 1 presents the summary statistics for both the water consumption data, landscape data, and moving data. It shows that just over 4% of the water consumption sample moves each year and that the mean fraction of landscape greenness is 37% (with the mean taken over household parcels). As can be seen in the summary statistics, all variables have substantial variation.

Our final data sources provide neighborhood and household demographics. First, we purchased household-level demographic information for each of the households in our remote sensing subset of data from the marketing company Acxiom. Axciom collects demographic information from a variety of sources, including public records and proprietary suppliers. These include data on physical household characteristics, such as the home size, as well as variables such as income, age, political party affiliation, and indices of interests based on magazine subscriptions. These data are useful for exploring the heterogeneity in our estimates, as well as for informing additional robustness checks. These demographic variables, as well as those from the Assessor's office, are especially useful for comparing SRP-eligible households to other households, and comparing the subsample of landscape data to the larger sample of water data.[9]

3 Empirical Specification

3.1 Challenges in Identifying Peer Effects

Peer effects are notoriously challenging to identify. A first question in any model of social interactions is how to define the peer group. Defining the peer group membership too broadly could pick up sufficient heterogeneity in the group that it leads to spurious correlations. Indeed, a careful definition of the peer group is central to identification in some studies (e.g., Bertrand et al. 2000). In our setting, we are interested in how peer effects influence the choice of landscaping and water usage. While water usage may not be visible, landscaping is usually highly visible. This lends itself to a definition of the peer group based on spatial proximity to the focal household parcel. In a similar setting, Towe and Lawley (2013) define neighbors as the nearest 13 and nearest 25 neighbors by distance. Because there is variation in parcel sizes, we prefer a measure based on the radius around the focal household parcel (our 500′ radius includes 23 neighbors on average). Using a geographic peer group is also common in the literature (Topa 2001; Arzaghi and Henderson 2007; Bell and Song 2007; Manchanda et al. 2008; Choi et al. 2010; McShane

[9] Appendix B presents summary statistics for both the Assessor's Office and Acxiom data, as well as tables of balance comparing SRP-eligible households to other households, and landscape data subsample to the larger sample of water data. Appendix B also contains further details on the steps used to convert our raw data into the final data set.

et al. 2012; Narayanan and Nair 2013). Households can naturally also be expected to have other social groups as well, such as those relating to family, friends, schools, and jobs. So we view our measure as a minimal measure of the social group relevant to water and landscape decisions.

Once the peer group has been defined, identification of peer effects also relies on addressing the further classic concerns: self-selection of peers, correlated unobservables, and simultaneity. The challenge of self-selection of peers (sometimes described as 'homophily') is a major concern for identifying the relationship between an individual's decision and the average decisions by the peer group because individuals endogenously choose their social networks. In the case of housing choice, consumers with similar preferences can be expected to sort into neighborhoods. Empirical studies have often addressed this issue with a rich set of fixed effects that remove variation stemming from endogenous selection of the peer group.

Correlated unobservables refer to the many other factors that may influence both the focal consumer and peers. For example, if there is an economic downturn facing all households in a neighborhood, their decisions may all appear to be aligned, but this alignment is due to the conditions faced by the households, rather than peer effects. Studies address this issue in various ways, including time-varying fixed effects or an instrument that provides a source of exogenous variation that shifts the peer characteristics but does not shift the individual characteristics.

Simultaneity (sometimes called 'reflection') refers to the concern that just as peers may influence the focal household, the focal household may influence peers. At the extreme this can lead to mathematical non-identification of peer effects. Simultaneity in this context can be thought of as an endogeneity issue analogous to the classic simultaneity of supply and demand. In the supply/demand context, the standard instrumental variables approach requires an exclusion restriction, such as a supply shifter to identify demand or a demand shifter to identify supply. Empirical studies often use instruments to address this issue or alternatively use lagged values of the peer group variable.

For both the issue of simultaneity and correlated unobservables, the fundamental challenge is one of research design. Angrist (2014) clearly points out the ideal characteristics of a well-identified peer effects study: "Research designs that manipulate peer characteristics in a manner unrelated to individual characteristics provide the most compelling evidence on the nature of social spillovers." Our research design overcomes the above challenges by focusing on variation within households and within zip code-by-year and exploiting exogenous shocks that fulfill the criterion laid out by Angrist (2014).

3.2 Empirical Strategy

Consider a classic linear-in-means specification that models the water consumption or landscape choice of household i in year t as a function of the peer group's aggregate choices, whether the household is moving in that year, time-invariant household characteristics, and time-varying characteristics of the local neighborhood or zip code z:

$$y_{it} = \theta \bar{y}_{it} + \delta m_{it} + \eta_i + \phi_{tz} + \epsilon_{it}, \tag{1}$$

where y_{it} is the household's choice (i.e., either water consumption or landscape choice). If we denote household i's peer group as P_i, then $\bar{y}_{it} = \frac{1}{|P_i|}\sum_{i' \in P_i} y_{it}$ is the average of the choices of the household's peers.[10] m_{it} is a dummy variable for whether the household just moved to the parcel they are occupying in that year (i.e., a housing transaction). We include m_{it} because we expect the mover's preferences to often differ from those of the previous owner and because the move event can help overcome consumer inertia in the landscaping decision. η_i contains time-invariant household characteristics, which we model as a household fixed effect (i.e., a fixed effect for each parcel x owner combination, so that there is a different fixed effect after a sale). ϕ_{tz} captures time-varying factors such as localized economic shocks or major new development in a neighborhood, and we model this with zip code x year fixed effects.

One clear implication of this model is that as long as δ is not zero, we would expect the intensity of migration into a neighborhood to also influence $\bar{y}_{it}$, since each neighbor who moved would be more likely to change their landscaping. This suggests the use of the average migration into the peer group, $\bar{m}_{it}$, as an instrument for $\bar{y}_{it}$ in (1).

The key identification assumption in this research design is that after the inclusion of our fixed effects, there are no unobserved factors that influence both the decision of

[10]In the incomplete information framework of Manski (1993), $\bar{y}_{it}$ would be represented by $\mathbb{E}\left[y_{gt}\right]$, where g refers to the group.

potential neighbors to relocate into that neighborhood and the decision of the individual household to change their water consumption or landscape. Our time-varying localized fixed effects are crucially important for allowing us to focus on the idiosyncratic shocks (off the mean) to water consumption or landscapes after conditioning on broader neighborhood changes. The remaining variation after including time-varying localized fixed effects is due to individuals deciding to change their water consumption or landscapes after individual shocks, such as a decision to renovate, an individual-specific promotion or loss of a job, etc.

This identification strategy overcomes the three primary challenges discussed above. Household fixed effects immediately address self-selection of peers based on household-specific preferences. Both correlated unobservables and simultaneity are endogeneity concerns, and our instrumental variables strategy addresses these concerns by focusing on the remaining variation after removing the time-varying localized fixed effects. Of course, for our research strategy to work, the instrument cannot be weak.

Fortunately, just as expected, we find a strong relationship between the fraction of movers into the peer group and the peer group's decision (see Appendix C for full first-stage results). The F-statistic for the first stage is 599.9 for the water data and 37.5 for the landscape data.[11] The economic explanation for this relationship is that when people move into a new house, they are likely to use the opportunity to change things about the house, including the landscaping. We find that when there is a greater fraction of sold

[11] The instrument also consistently passes the Kleibergen-Papp under-identification and weak identification tests at the 1% level.

homes in the peer group, the changes in water use and landscape greenness are negative, suggesting that households that move tend to transition from greener landscapes to drier landscapes. This is consistent with the broader trend towards drier landscapes in the Southwest, and consumer migration provides the exogenous shock to spur the transitions by peer group members.

4 Results

4.1 Primary Water Consumption Results

We begin by estimating our primary specification to identify peer effects in water consumption (Equation 1). Table 2 presents the ordinary least squares (OLS) and instrumental variables (IV) estimates using household and year fixed effects (Columns 1 and 3) and household and year x zip code fixed effects (Columns 2 and 4).[12] The primary variable of interest is the log of the mean water consumption of the peers within 500′. Columns 3 and 4 present the IV estimates instrumenting for the peer's water consumption with the fraction of parcels within 500′ that observe a housing transaction in that year.

Our preferred specification is in column 4, which indicates that a 10% change in the peer group average water consumption results in a 1.7% change in house $i's$ water consumption. This translates to a savings of 207 gallons per month out of an average monthly consumption of 12,532 gallons. For reference, a typical load of laundry uses 30 gallons of

[12]These results are robust to substituting house/parcel fixed effects for household fixed effects, where the former is tied to the parcel while latter changes when a house is sold.

water. To calculate the effect of a change in water use by a single house in the peer group, we can divide the coefficient by the average number of houses in a peer group (25.3). We find that a 10 percentage point change in consumption for a single house in household $i's$ 500′ peer group causes a 0.065% change in household $i's$ consumption. For comparison, if we use the -0.33 water demand elasticity from Olmstead et al. (2007), water prices would have to change by 5.2% or $3.22 per month to achieve a similar change in water consumption.

To better understand the spatial nature of this peer effect, we perform a further analysis in which we separate the peer group variable into two variables. In one variable, we include only peers within 500′ that share the same street as the focal household, while in a second variable we include only those that do not share the same street. Our hypothesis is that individuals are more likely to be influenced by activities that affect water consumption (e.g., changes in landscaping) by close neighbors on the same street than by close neighbors on other streets.

The results are presented in Table 3. Column 1 displays OLS results while column 2 displays IV results. For interpreting the coefficients, it is useful to recognize that the number of houses on the same street is smaller than the number of houses not on the same street within the 500′ radius of any given house. The bottom two rows of the table show the marginal effect per household, which can be interpreted as the percent change in annual water consumption for the focal household with a 1% change for a single peer household. The results in column 2 suggest that the per-household peer effect is almost twice as large for peers on the same street as peers on other streets, 0.009 and 0.005, respectively. This provides evidence supporting our hypothesis that peer effects are stronger with closer neighbors. However, this evidence should only be taken as suggestive, as the per household marginal effects are not statistically different from one another (p=0.17).[13]

4.2 Primary Landscape Results

The regression results so far provide evidence of peer effects in aggregate water consumption, with visible landscaping as a potential mechanism. To test the hypothesis that landscaping is a primary mechanism for these peer effects, we estimate our model in (1) using the cleaned remote sensing landscape data. Table 4 presents the primary landscape results with household and year fixed effects (Columns 1 and 3) and household and year x zip code fixed effects (Columns 2 and 4). As before, columns 1 and 2 present OLS results, while columns 3 and 4 present the results where we instrument for the peer group's landscape with the fraction of houses in the peer group that were sold in a particular year. The primary variable of interest is the log of the average percentage of green landscaping for the neighbors within a 500′ radius of the given household. Just as in the water consumption results, we see statistically significant evidence of peer effects in landscape choice.

The results from the IV specification in column 4 indicate that a 10% increase in the peer group average landscaping results in a 1.4% increase in house $i's$ green landscaping. Translating this to the marginal effect of a single peer household, we find that a 10%

[13] We find similar results with the landscape data.

increase in green landscaping for a single household in house $i's$ 500′ peer group leads to a 0.06% increase in house $i's$ green landscaping.[14]

In order to determine the extent to which the water consumption peer effect is driven by outdoor landscaping decisions, we compare the water savings associated with the peer effect in water consumption to the savings associated with landscape choice. To recap, a 10% change in the peer group average water consumption results in a 1.7% or 207 gallons per month change in house $i's$ water consumption with a 90% confidence interval of 119-294 gallons per month. A 10% change in peer landscaping results in a 1.4% change in house $i's$ landscaping. When we regress landscape choice on water consumption (Table A.1 in the Appendix), we find that a 53.2 gallon change in monthly water use (0.42%) results in a 1% change in a household's landscape on average. Using this conversion, a 10% decrease in peer water use (1,253 gallons per month) maps to landscape changes that result in a 3.3% change in house $i's$ landscaping, for a savings of 176 gallons per month.[15] This turns out to be remarkably close to the savings from peer effects in water consumption regressions (207 gallons per month) and indeed the two are not statistically significantly different.

This near equivalence between the magnitude of the water peer effect and the landscape peer effect suggests that nearly all, if not all, of the water consumption peer effect is due to a peer effect in landscape choice. We view this result as eminently plausible, for outdoor water use for landscaping is likely to to be the most visible and conspicuous wa-

[14]The peer group coefficient in column 4 is statistically significant at the 10% level when year x zip code fixed effects are included. If we substitute year x 3-digit zip code fixed effects, we have statistical significance at the 1% level, with little change in the coefficient.

[15]This translates to a savings of \$0.81 per month or \$9.74 per year per household.

ter consumption, and thus is likely to be the consumption most likely to influence one's peers.

4.3 Role of Economic Incentives

With the Salt River Project's provision of heavily discounted irrigation water for outdoor use, we have an opportunity to test whether the price signal for water influences the strength of the peer effect. We use the SRP as a natural experiment to explore whether SRP-eligible households are less influenced by peer water consumption and landscape choices than non-eligible households. Jessoe and Rapson (2014) show that in the electricity context consumers respond more to prices when relevant information is available, and here we examine whether consumers respond more to peers when there is a stronger price signal in landscape decisions.

For this to be a valid analysis, we must be confident that SRP-eligible households are similar to non-eligible households. In Appendix B, we show that the characteristics of SRP and non-SRP households tend to be statistically different from one another across quite a few of the observables.[16] Thus, we perform a nearest neighbor matching approach to find a sample of non-SRP households that match the sample of SRP households. Our results are also robust to using other propensity score matching approaches, such as kernel matching.

Columns 5 and 6 of Table 4 display evidence of the impact of water prices on the peer

[16]We also run a regression of an indicator for being an SRP household on demographic variables including a subdivision fixed effect. In general, SRP and non-SRP households do not appear to be statistically different from one another within subdivisions.

effect. Column 5 displays the IV results using the nearest neighbor matched sample for non-SRP households while column 6 runs the IV specification in column 4 on SRP-eligible households only.[17] The results are striking: there is a statistically significant peer effect in landscaping for the off-project households, but no evidence supportive of a peer effect for households that were eligible for the extremely heavily discounted (nearly free) SRP irrigation water. The coefficient in column 6 (on-project) is not statistically distinguishable from zero, while the coefficient in column 5 (off-project) is 0.49, larger than the result in column 4 with the full dataset. The difference between the two estimates is statistically significant at the 1% level.

The notable difference between the SRP and non-SRP peer coefficients suggests that economic incentives to adopt dry landscaping are important for the operation of peer effects. In neighborhoods where green landscaping is costly, neighbors may discuss xeriscaping (landscaping with slow-growing, drought-tolerant plants) as a money-saving tool. There are at least two possible channels by which economic incentives affect the peer effects. Households may be more susceptible to peer effects in landscaping when there are economic incentives because they are more likely to be looking to save money on their water bill. Alternatively, neighbors may be more likely to discuss and share information about dry landscaping when there is an obvious monetary benefit. While we cannot differentiate between these two hypotheses, we can develop a deeper understanding of how the peer effects work by estimating heterogeneous peer effects.

[17] Appendix C compares the results estimated on the nearest neighbor matched sample with the preferred IV approach on non-SRP households only. The coefficient using the nearest neighbor matching approach is larger in magnitude but the two estimates are not statically different from one another.

4.4 Heterogeneous Peer Effects

We explore heterogeneity in peer effects for two main reasons. First, by understanding which households are most susceptible to peer influence, we can shed light on the mechanisms underpinning peer effects. Second, such an analysis provides useful information for policymakers who might consider targeted information campaigns to 'seed' a neighborhood with dry landscaping. This may be particularly useful for fast-growing neighborhoods or other areas with growing water demand. To examine heterogeneous peer effects, we first estimate the IV specification described by (1) on four subsamples.

The first subsample we examine is a subsample of households with children. Column 1 in Table 5 shows evidence of statistically significant peer effects in landscaping for households that have children. The coefficient is slightly less than the coefficient in column 3 of Table 4, but it not statistically significantly different from that estimate. This result indicates that peer effects in landscaping do occur for families with children, perhaps facilitated by interactions with neighbors through children.

The second and third subsamples are restricted to households with a pet. One might expect households with a pet to prefer green lawns and be less susceptible to peer influences relating to landscaping. Indeed in column 2 of Table 5, we restrict the sample to households with a dog and see no statistically significant peer effect. This result is simply a failure to reject the null, which means that we cannot make a definitive statement. However, the sample size is still quite large, so we view this result as suggesting that households with dogs are less amenable to landscaping changes influenced by peers. In column 3, we see a similar result for cats, again suggesting that households with pets are less like to switch landscaping due to the influence of peers.

The fourth subsample is restricted to households with a pool. This is a common occurrence in Phoenix; 42% of our sample has a pool. In column 4 of Table 5 we show a larger and statistically significant peer effect, consistent with a hypothesis that pools foster social interactions.

Last, we interact our peer greenness variable with the distance to the closest park and the school. We hypothesize that households further away from parks and schools have less access to publicly available green spaces and therefore are more likely to be affected by their neighbors' landscape choices, which may lead them to be more influenced by their neighbors' choices. The specification with interactions that we run is the following:

$$y_{it} = \theta \bar{y}_{it} + \sum_{k \in \mathcal{K}} \gamma_j D_{ik} + \sum_{k \in \mathcal{K}} \vartheta_j D_{ik} * \bar{y}_{it} + \rho m_{it} + \mu_i + \xi_t + \epsilon_{it}, \tag{2}$$

where D_{ik} refers to distance variables and $\mathcal{K} = \{park, school\}$. Because our demographics are perfectly collinear with zip code x year fixed effects, we use household i and year t fixed effects in this specification, making it comparable to column 3 in Table 4.

Column 5 of Table 5 shows that increasing the distance to a park enhances the peer effect in landscape choices, consistent with our hypothesis. An alternative explanation is that houses closer to parks have a revealed preference for green space, and are thus less affected. The effect is quite strong, as the units for the distance to a park is 1,000′ (the mean distance is 2,725′ and standard deviation is 1,918′). So going from a -1 standard deviation to +1 standard deviation from the mean changes the magnitude of the effect of all households in the peer group adding a 1% change by 0.172%. This change is larger than our effect in our primary specification in Table 4.

To help understand why we might see such heterogeneity, we also explore the correlational relationships between demographic variables and household preferences for landscape choice by regressing the estimated household fixed effects on the demographic variables. These results, included in Appendix C, indicate that households furthest from the nearest park and school are correlated with less-green landscaping, consistent with households sorting into residential location in part based on distance to green space. However, this result did not hold for SRP households, suggesting that public green spaces and schools may be substitutes for private green spaces, but only when there are low costs for private green space. Dog and cat ownership, as well as having a pool are also correlated with more green space. For policy, these correlational findings are useful benchmarks that identify groups of households with the most green space that could be converted to drier landscaping.

4.5 Implications for Policy and Targeted Interventions

Policies to promote dry landscaping may have substantial spillover effects due to peer effects. Consider a dry landscaping subsidy that leads a single household to reduce its landscaping greenness by 10%. To calculate the total impact of such a subsidy, we also need to calculate the effect on the household's peers. Using the results from our regression with heterogeneous peer effects (column 5 of Table 5), we first calculate the peer effect for each household from a 10% landscape transition by its entire peer group. This is equal to 0.1 times the fitted peer effect from the regression, which is the peer effect coefficient plus the interaction effects. We then divide by the number of peer houses to calculate the 'susceptibility' of each household to the actions of a single peer. This can be interpreted as the expected landscape change the household would make in response to a peer landscape change of 10%. From our interaction results, we know that houses with children and those further from parks are more susceptible to peer effects. Houses with fewer peers are also more susceptible to a landscape change by a single peer because each peer of a household with a small peer group has a larger effect on the average than for households with large peer groups.

We calculate the average susceptibility of households in our data from a landscape change by a single peer of 10% to be 0.124%. If we aggregate the peer susceptibility values for each focal household, considering that each focal household could be targeted for a dry landscaping subsidy, the average total susceptibility of the peer group is 0.0275.[18] This yields an average water savings for the entire peer group *from the peer effect* of 146 gallons per month. This is 27.5% of the water savings from the original 10% intervention–a substantial water savings from peer effects alone. Extrapolating, if 1,000 randomly distributed households across Phoenix were influenced by policy to reduce water use by 10%, in addition to the reduction by the targeted households of 532 thousand gallons, the

[18] If we instead were to multiply by the average peer group size of 23.2, this yields a value of 0.0288 instead of 0.0275 – the 5% difference results from the network structure.

associated peer effects would lead to an additional reduction of approximately 146 thousand gallons of water per month by the peers–the equivalent of the monthly water use of 13 households.

This illustrative calculation highlights the importance of accounting for peer influence in assessing interventions, but it also ignores the ability to target. We have shown that there is significant heterogeneity in landscaping peer effects, which would suggest that targeted policies may be significantly more effective. Rather than a random set of households, if we instead targeted the 1,000 households whose peers had the highest susceptibility to peer influence, the average total peer influence exerted by these top influencers from a 10% decrease in greenness of landscaping is 0.174, more than six times the average and 74% larger than the size of the original, hypothetical intervention. Targeting these 1,000 households would have a spillover effect on water savings of 928 thousand gallons of water per month, the usage of 84 households.[19] Even a simple targeting strategy of off-project houses only increases the aggregate influence of a single house from 0.028 to .031, and if we look at a random set of these that are also at least 2,000 feet from a park, the aggregate effect increases further to 0.041, a 47% increase over the baseline average and almost half the size of the original intervention. Of course such a targeted policy would be most cost-effective if it also focused on areas most in need of future upgrades to the water system due to increasing demand.

[19] The effect is actually slightly larger since this calculation ignores the additional effects on the peers-of-peers.

4.6 Robustness Checks

In this section we provide a series of robustness checks to build further evidence that our peer effect estimate is well-identified and robust to various assumptions.

We begin by examining alternative definitions of the peer group. Our preferred specification uses a 500′ radius, which captures about 23 homes on average. This assumption aligns with the number of neighbors we would expect a typical household to interact with (even if it is just seeing the landscaping) and is in the same range as other peer group specifications in the literature. But we recognize that one could argue for a broader or narrower peer group. Table 6 presents the results of estimating our preferred specifications (column 4 of Table 2 and column 4 of Table 4) using a 400′, 600′, and 700′ radius. The first three columns present the results using the water consumption data, while the second three columns present the results using the landscape data. The key result from this robustness check, as can be seen in the last two rows, is that the peer effect only changes very modestly with distance, especially in the water consumption results. We lose some statistical significance with a smaller radii in the landscape results, but the marginal effects for a change by a single peer household remain very similar. In general, the standard errors render the results at the different radii statistically indistinguishable from one another.[20]

One potential concern that might lead to a difference in results for the landscaping and water use analyses is the fact that the houses in the remote sensing sample are a subset of

[20]We extended the radius to 1000 feet and found that we lose statistical significance around a radius of 800 feet. Likewise, peer groups defined by radii smaller than 300 feet do not have a statistically significant effect because they contain too few houses, a problem that occurs even at 400′ in the landscape data.

the households in the water consumption sample. Sample selection bias in the landscaping data is therefore a natural concern. In the Appendix, Table A.7 presents the results of estimating the models in Table 2, our preferred water consumption regressions, on the remote sensing sample of households. The estimates are statistically indistinguishable indicating the remote sensing sample does not suffer from sample selection bias.

To address any remaining concerns about simultaneity (reflection), we estimate our preferred specifications (column 4 of Table 2 and column 4 of Table 4) using the lag of the log of peer water consumption or the lag of the log of peer landscaping in place of contemporaneous peer water consumption and peer landscaping. We then instrument each of the lagged peer variables with the lag of the proportion of houses sold in each peer group (the lagged version of our primary instrumental variable). The results, presented in columns 1 and 3 of Table A.10, show that the lagged peer effect estimates for water consumption and landscaping are lower than the contemporaneous peer effect estimates by 0.1 and 0.044 respectively. This might be expected if peer effects diminish over time. However, the lagged and contemporaneous estimates are not statistically significantly different from one another.[21]

Finally, despite controlling for housing transactions, one might be concerned that landscape changes and water consumption changes for a focal household might be driven more by whether they moved themselves rather than peer consumption and peer landscape changes. To address this concern, we estimate our preferred specifications on

[21]The lagged landscaping peer effect is not statistically significant using five-digit zip code by year fixed effects. It is significant using three-digit zip code by year fixed effects however, but slightly larger in magnitude. We present the five-digit zip code results to be consistent with our primary specifications.

houses that were not sold in year t. The results are presented in columns 2 and 4 of Table A.10. Both peer effect estimates remain statistically significant and the coefficients are largely unchanged.

5 Conclusions

In this paper, we estimate causal peer effects in residential water consumption and landscaping using a unique IV strategy that leverages consumer migration into the peer group. In this strategy, we exploit within household and within zip code-year variation, so that our key identifying assumption is that peer housing transactions are an exogenous shock to the individual's peer group. Any factors that vary over time and are zip code-specific as well as any factors that are household-specific will be accommodated by our fixed effects, allowing our IV strategy to address all three known issues in identifying peer effects: homophily, correlated unobservables, and simultaneity.

For our analysis of peer effects in landscaping, we use machine learning techniques on high resolution remote sensing data to develop a robust measure of landscape greenness. Our primary result is that a 10% change in peer water consumption results in a 1.7% change in the individual household's water consumption, while a 10% change in peer landscape greenness results in a 1.4% change in the individual household's landscape greenness. These results translate to a savings of roughly 200 gallons or roughly $0.81 per household per month, or 2,112 gallons or $9.74 per household per year. Through a simple calculation, we show that the peer effect in landscape transitions appears to fully or nearly fully explain the peer effect in total water consumption, providing evidence that the peer effect works through the channel of outdoor water use.

One novel finding of this study is that the effect of social interactions depends on economic incentives. By exploiting a natural experiment offered by the existence of the Salt River Project's provision of heavily discounted irrigation water, we show that the landscaping peer effect is close to zero and not statistically significant for households eligible for the discounted water. This provides further evidence that the mechanism behind the water consumption peer effect is outdoor water use and demonstrates the power of economic incentives on the influence of peers.

Our results have clear policy implications. The presence of a social spillover suggests that policies may have disproportionately larger indirect effects. For example, a dry landscaping subsidy that causes 1,000 distributed households within Phoenix to decrease their green landscaping by 10% would lead to a social spillover effect equivalent to the monthly water use of 13 households. We also showed that targeting may be especially important. The social spillovers not only appear to be non-existent for the Salt River Project households, but they appear much weaker for households near parks and households with pets. It is also clear that water districts in dry regions would only want to consider such policies to encourage dry landscaping when there is a strain on parts of the water system from increasing demand or reduced supply. Thus, optimal policy design will inherently involve a consideration of both water district constraints and the potential spillovers possible in the target audience.

References

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

Allcott, H. and Rogers, T. (2014), 'The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation', *The American Economic Review* **104**(10), 3003–3037.

Angrist, J. D. (2014), 'The perils of peer effects', *Labour Economics* **30**, 98–108.

Arzaghi, M. and Henderson, J. V. (2007), 'Networking off Madison Avenue', *Review of Economic Studies* **75**(4), 1011–1038.

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

Bass, F. M. (1969), 'A new product growth model for consumer durables', *Management Science* **15**, 215–227.

Bell, D. R. and Song, S. (2007), 'Neighborhood effects and trial on the internet: Evidence from online grocery retailing', *Quantitative Marketing and Economics* **5**, 361–400.

Bertrand, M., Luttmer, E. F. P. and Mullainathan, S. (2000), 'Network effects and welfare cultures', *Quarterly Journal of Economics* **115**(3), 1019–1055.

Bollinger, B. and Gillingham, K. (2012), 'Peer effects in the diffusion of solar photovoltaic panels', *Marketing Science* **31**(6), 900–912.

Brock, W. and Durlaf, S. (2001), *Handbook of Econometrics*, Vol. 5, Elsevier, chapter 4, Interaction-based Models, pp. 3297–3380.

Burkhardt, J., Gillingham, K. and Kopelle, P. (2017), 'How do households respond to critical peak pricing? experimental evidence on the role of information and incentives', *Yale University Working Paper* .

Choi, J., Huiand, S. K. and Bell, D. R. (2010), 'Spatiotemporal analysis of imitation behavior across new buyers at an online grocery retailer', *Journal of Marketing Research* **47**(1), 75–89.

Conley, T. and Udry, C. (2010), 'Learning about a new technology: Pineapple in Ghana', *American Economic Review* **100(1)**, 35–69.

Costa, D. L. and Kahn, M. E. (2013), 'Energy conservation nudges and environmentalist ideology: Evidence from a randomized residential electricity field experiment', *Journal of the European Economic Association* **11**, 680702.

CUWA (2017), 'Adapting to change: Utility systems and declining flows', *California Urban Water Agencies Working Paper* .

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J. and Pasteris., P. P. (2008), 'Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous united states', *International journal of climatology* **28**(15), 2031–2064.

Deschnes, O. and Greenstone, M. (2007), 'The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather', *The American Economic Review* **97**(1), 354–385.

Dolan, P. and Metcalfe, R. (2015), Neighbors, knowledge, and nuggets: Two natural field experiments on the role of incentives on energy conservation.

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

Foster, A. and Rosenzweig, M. (1995), 'Learning by doing and learning from others: Human capital and technical change in agriculture', *Journal of Political Economy* **103**(6), 1176–1209.

Graham, B. (2008), 'Identifying social interactions through conditional variance restrictions', *Econometrica* **76(3)**, 643–660.

Graziano, M. and Gillingham, K. (2015), 'Spatial patterns of solar photovoltaic system adoption: The influence of neighbors and the built environment', *Journal of Economic Geography* **15(4)**, 815–839.

Griliches, Z. (1957), 'Hybrid corn: An exploration in the economics of technological change', *Econometrica* **25**(4), 501–522.

Hartmann, W. R., Manchanda, P., Nair, H., Bothner, M., Dodds, P., Godes, D., Hosanagar, K. and Tucker, C. (2008), 'Modeling social interactions: Identification, empirical methods and policy implications', *Marketing Letters* **19**, 287–304.

Heutel, G. and Muehlegger, E. (2015), 'Consumer learning and hybrid vehicle adoption', *Environmental and resource economics* **62**(1), 125–161.

Ito, K., Ida, T. and Tanaka, M. (2017), 'Moral suasion and economic incentives: Field experimental evidence from energy demand', *American Economic Journal: Economic Policy* **forthcoming**.

Jessoe, K. and Rapson, D. (2014), 'Knowledge is (less) power: Experimental evidence from residential energy use', *The American Economic Review* **104**(4), 1417–1438.

Loarie, S. R., Duffy, P. B., Hamilton, H., Asner, G. P., Field, C. B. and Ackerly, D. D. (2009), 'The velocity of climate change', *Nature* **462**(7276), 1052–1055.

Manchanda, P., Xie, Y. and Youn, N. (2008), 'The role of targeted communication and contagion in product adoption', *Marketing Science* **27**(6), 961–976.

Manski, C. (1993), 'Identification of endogenous social effects: The reflection problem', *Review of Economic Studies* **60**, 531–542.

McShane, B., Bradlow, E. T. and Berger, J. (2012), 'Visual influence and social groups', *Journal of Marketing Research* **49(6)**, 854–871.

Moffitt, R. (2001), *Social Dynamics*, MIT Press, chapter 3, Policy Interventions, Low-level Equilibria, and Social Interactions, pp. 45–82.

Narayanan, S. and Nair, H. (2013), 'Estimating causal installed-base effects: A bias-correction approach', *Journal of Marketing Research* **50**(1), 70–94.

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

Reiss, P. C. and White, M. W. (2008), 'What changes energy consumption? prices and public pressures', *The RAND Journal of Economics* **39**(3), 636–663.

Rogers, E. (1995), *Diffusion of Innovations*, The Free Press.

Sacerdote, B. (2001), 'Peer effects with random assignment: Results for Dartmouth roommates', *Quarterly Journal of Economics* **116**, 681–704.

Topa, G. (2001), 'Social interactions, local spillovers and unemployment', *Review of Economic Studies* **68**(2), 261–295.

Towe, C. and Lawley, C. (2013), 'The contagion effect of neighboring foreclosures', *American Economic Journal: Economic Policy* **5**(2), 313–335.

Tsvetanov, T. and Gillingham, K. (2017), 'Nudging energy efficiency audits: Evidence from a field experiment', *Yale University Working Paper* .

Tables & Figures

Table 1: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.
Panel A: Water Consumption (N=2,533,715; 308,652 households)				
household annual consumption	177.4	111.6	15.0	718.0
annual consumption of households within 500′	177.0	63.9	16.0	717.0
1(housing transaction)	0.045	0.207	0	1
fraction of houses sold within 500′	0.045	0.051	0	1
1(SRP-eligible)	0.43	0.50	0	1
Panel B: Landscape (N=531,650; 71,477 households)				
fraction of landscape green	0.37	0.10	0.19	0.68
fraction green for households within 500′	0.37	0.06	0.19	0.68
1(housing transaction)	0.04	0.20	0	1
fraction of households sold within 500′	0.04	0.05	0	1
1(SRP-eligible)	0.38	0.49	0	1
Lot Size (ft^2)	9,706	570	1,537	299,200

Notes: An observation is a household-year for both datasets. The units of household annual water consumption are in hundred cubic feet (ccf) per year (1 ccf is 748 gallons). Figure 3 illustrates the parcels covered in the landscape subsample and Appendix B provides evidence on the representativeness of this sample. Summary statistics for other radii besides 500′ are also given in Appendix B.

Table 2: Primary Specifications for Water Consumption

	(1) OLS	(2) OLS	(3) IV	(4) IV
log(consumption within 500′)	0.18***	0.12***	0.31***	0.17***
	(0.01)	(0.01)	(0.03)	(0.04)
1(housing transaction)	-0.23***	-0.22***	-0.23***	-0.22***
	(0.003)	(0.003)	(0.003)	(0.003)
Household Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y	N	Y	N
Year x Zip Code Fixed Effects	N	Y	N	Y
R-squared	0.77	0.77	0.77	0.77
N	2,533,715	2,533,715	2,533,715	2,533,715

Notes: The dependent variable is the log of annual water consumption. An observation is a household parcel-year. log(consumption within 500′) is the log of the average annual consumption of the neighbors within a 500-foot radius. On average, there are 25.3 houses within a 500′ radius of any household in our study. Column 1 and 2 present OLS peer effect results. Columns 3 and 4 instrument for log(consumption within 500′) using the fraction of parcels with housing transactions within 500′. Standard errors are clustered at the subdivision level. *** denotes significance at 1% level, ** at 5% level, * at 10% level.

Table 3: Street Level Analysis for Water Consumption

	(1) OLS	(2) IV
log(same street peer consumption)	0.054***	0.067***
	(0.004)	(0.023)
log(other street peer consumption)	0.051***	0.091***
	(0.005)	(0.025)
1(housing transaction)	-0.25***	-0.25***
	(0.004)	(0.004)
Household Fixed Effects	Y	Y
Year x Zip Code Fixed Effects	Y	Y
R-squared	0.78	0.78
N	1,959,519	1,959,519
Marginal effect per peer household		
Same street peer consumption	0.007	0.009
Other street peer consumption	0.003	0.005

Notes: The dependent variable is the log of annual water consumption. An observation is a household parcel-year. log(same street peer consumption) is the log of the average annual consumption of the neighbors within a 500-foot radius who are on the same street as the focal household. log(other street peer consumption) refers to those not on the same street. Column 2 instruments for the same street log consumption using the fraction of parcels with housing transactions within 500′ that are on the same street, while the other street log consumption is instrumented with the fraction of parcels with transactions on other streets. On average, there are 7.3 houses on the same street within a 500′ radius of any house in our study and 17.5 houses not on the same street within a 500′ radius of any house. The final two rows show the calculated marginal effect per household for same and other street peers. Standard errors are clustered at the subdivision level. *** denotes significance at 1% level, ** at 5% level, * at 10% level.

Table 4: Primary Specifications for Landscape Data

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS		IV		IV	
	All	All	All	All	Off SRP	On SRP
log(% green within 500′)	0.018***	0.018***	0.315***	0.140*	0.489***	-0.058
	(0.003)	(0.003)	(0.105)	(0.073)	(0.148)	(0.084)
1(housing transaction)	-0.007***	-0.007***	-0.007***	-0.007***	-0.01***	-0.006***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.002)	(0.0003)
Household Fixed Effects	Y	Y	Y	Y	Y	Y
Year Fixed Effects	Y	N	Y	N	N	N
Year x Zip Code Fixed Effects	N	Y	N	Y	Y	Y
R-squared	0.04	0.08	0.02	0.03	0.09	0.03
N	531,650	531,650	531,650	531,650	230,372	203,030

Notes: The dependent variable in each regression is the log of the percentage of green pixels in a given parcel. An observation is a household parcel-year. log(% green within 500′), the variable of interest, is the log of the average percentage of green of the neighbors within a 500′ radius. On average, there are 23.2 houses within a 500 foot radius of any household in our study. Column 1 and 2 present OLS results. Columns 3-6 instrument the peer variable using the fraction of houses in the peer group that had a housing transaction. Columns 5 and 6 divide the sample by whether the household is eligible for Salt River Project (SRP) discounted water. Column 5 uses a nearest neighbor matched sample (see Appendix C for IV results using the entire sample). The number of households used in columns 5 and 6 is 29,602. The number of observations do no match because the panel is unbalanced. Standard errors are clustered at the subdivision level (block bootstrapped for column 5). *** denotes significance at 1% level, ** at 5% level, * at 10% level.

Table 5: Heterogeneous Peer Effects in Landscaping

	(1) has kids	(2) has dog	(3) has cat	(4) has pool	(5) interactions
log(% green within 500′)	0.25**	0.04	-0.11	0.44***	0.10
	(0.11)	(0.18)	(0.21)	(0.15)	(0.09)
1(housing transaction)	-0.008***	-0.008***	-0.007***	-0.005***	-0.008***
	(0.0003)	(0.0006)	(0.0008)	(0.0004)	(0.0003)
% green x distance to park					0.05***
					(0.02)
% green x distance to school					0.03
					(0.03)
Household Fixed Effects	Y	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y	Y
R-squared	0.01	0.04	0.03	0.01	0.01
N	251,001	99,619	57,679	230,303	339,030

Notes: The dependent variable in each regression is the log of the percentage of green pixels in a given parcel. An observation is a household parcel-year. log(% green within 500′), the variable of interest, is the log of the average percentage of green of the neighbors within a 500′ radius. Column 1 restricts to the subsample of households with children; 2 to households with a dog; 3 to households with a cat; 4 to households with a pool. Column 5 interacts the variable of interest with the distance to the closest park and school (in units of 1,000′). The main effects are also included in the interactions specification, but are not shown for brevity. Standard errors are clustered at the subdivision level. *** denotes significance at 1% level, ** at 5% level, * at 10% level.

Table 6: Robustness Checks Comparing Different Radii

	(1)	(2)	(3)	(4)	(5)	(6)
	Water Consumption			Landscape		
	400′	600′	700′	400′	600′	700′
log(consumption within radius)	0.14***	0.17***	0.17***			
	(0.03)	(0.04)	(0.04)			
log(% green within radius)				0.10	0.19*	0.22**
				(0.07)	(0.07)	(0.08)
1(housing transaction)	-0.22***	-0.22***	-0.22***	-0.007***	-0.007***	-0.007***
	(0.003)	(0.003)	(0.003)	(0.0003)	(0.0003)	(0.0003)
Household Fixed Effects	Y	Y	Y	Y	Y	Y
Year x Zip Code Fixed Effects	Y	Y	Y	Y	Y	Y
R-squared	0.77	0.77	0.77	0.032	0.023	0.019
N	2,532,336	2,534,459	2,534,922	531,461	531,704	531,759
Marginal effect of a 10% change for a single peer household						
consumption within radius	0.06%	0.06%	0.06%			
% green within radius				0.05%	0.07%	0.08%

Notes: In the first three columns, the dependent variable is water consumption. In the second three, the dependent variable is the % landscape greenness. All columns run our preferred IV specification using the fraction of movers in the peer group as the instrument for the peer group variable. Each column uses a different radius for the peer group definition. On average, there are respectively 22.9, 26.25, and 26.7 houses within a 400, 600, and 700 foot radius of any household in our study. Standard errors are clustered at the subdivision level. *** denotes significance at 1% level, ** at 5% level, * at 10% level.

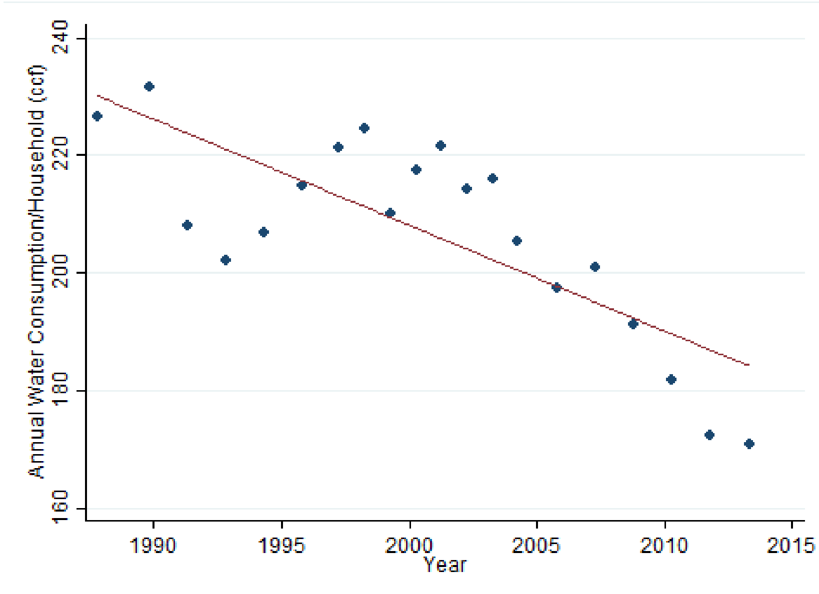

Figure 1: Binscatter plot of the annual water consumption per household in the Phoenix City Water District along with a linear best-fit trendline. Source: City of Phoenix Water Services Department.

Figure 2: Illustrative remote sensing images demonstrating the classification of green space. Panel A on the right shows what our remote sensing images look like, while Panel B on the left shows how the machine learning algorithm codes the pixels of green space.

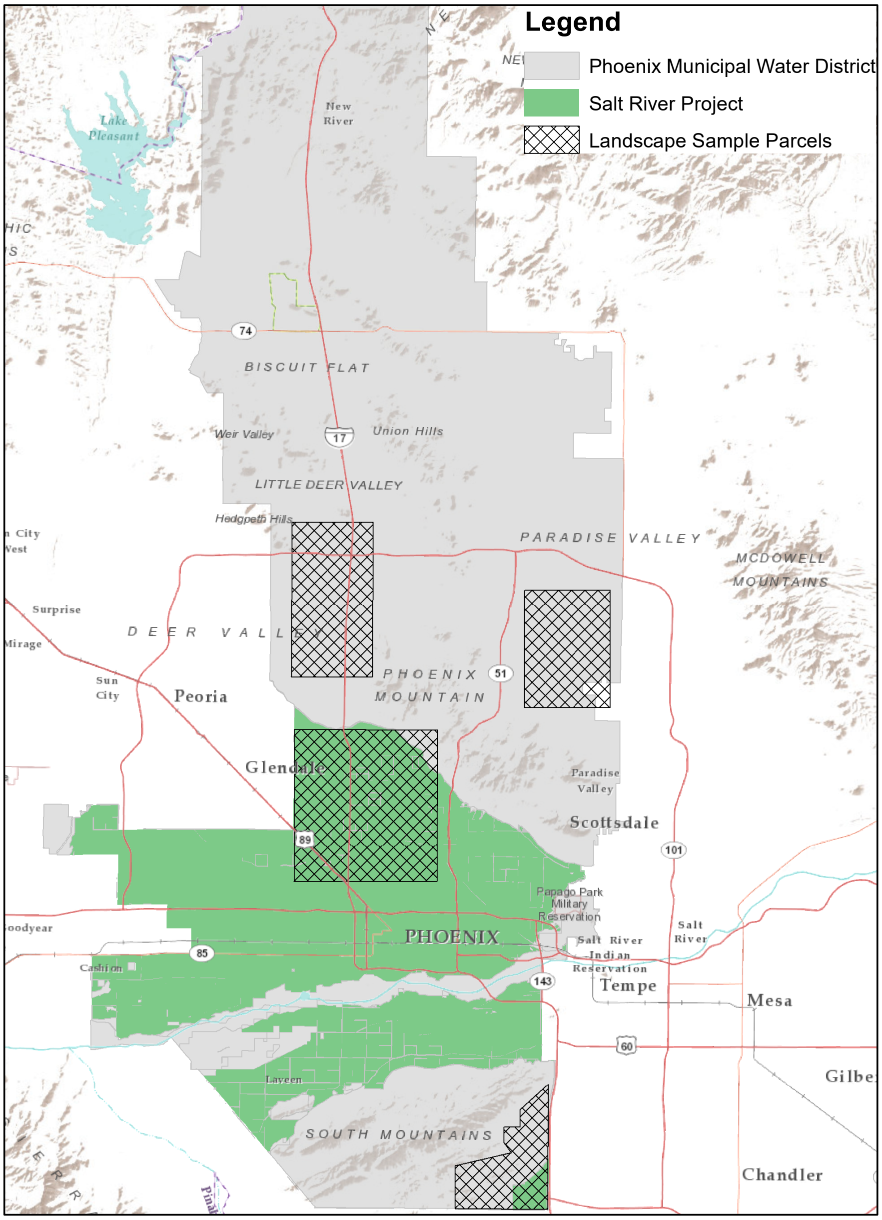

Figure 3: Phoenix water district boundary, along with identification of areas under the Salt River Project and landscape sample parcels.

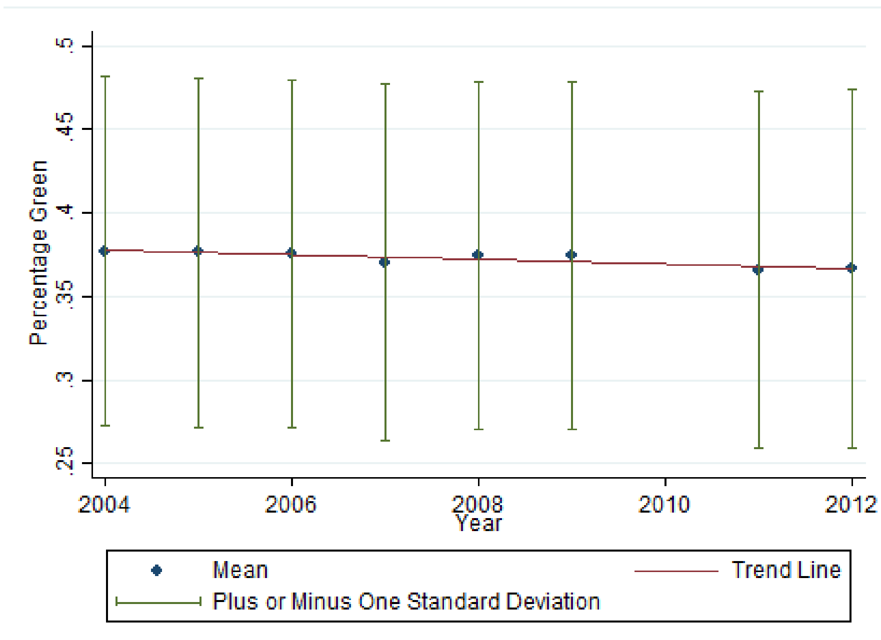

Figure 4: Mean landscape greenness by year with one standard deviation whiskers, based on on the remote sensing imagery.

Online Appendix

A Remote Sensing Data Cleaning Methodology

This appendix lays out the issue we are faced with in our remote sensing data, why an IV strategy does not solve it, and how we adjust the remote sensing data to develop a consistent and robust measure of landscape greenness.

One common issue in remote sensing imagery data is that it contains inherent measurement error from a variety of sources. These include the exact time and date that the image was taken, the weather on those dates, the type of camera, and even the camera settings. Thus, in remote sensing work it is common to adjust the imagery using other data sources to develop a consistent time series of images. Our methodology uses the water consumption data in a regression-based approach to adjust the remote sensing imagery data. This approach can be thought of as somewhat analogous to the PRISM model developed by Daly et al. (2008), which is used in both the scientific (Loarie et al. 2009) and economic (Deschnes and Greenstone 2007) literatures on climate change. The PRISM interpolation process assumes a linear relationship between elevation and the climate variables at each grid point (i.e., an intercept and a slope). The key assumption is that "for a localized region, elevation is the most important factor in the distribution of temperature and precipitation" (see section 4.1 in the linked data file). Thus elevation is used to create the fitted climate variables at every grid point. In our setting, we use observed water consumption as the most important determinant of the percentage of greenness in a the landscaping of any given parcel in Phoenix, a natural approach based on the climate in Phoenix. We then use the fitted landscape values, just as the the PRISM approach uses fitted climate values.

To further explain why we might have an issue, let y_{it} be the true fraction of green landscape on a given parcel i in year t. With the remote sensing data, we observe $\tilde{y}_{it}$, which is equal to y_{it} with a measurement error u_{it}:

$$\tilde{y}_{it} = y_{it} + u_{it}.$$

Substituting in from (1), we have

$$\tilde{y}_{it} = \theta\bar{\tilde{y}}_{it} + \delta m_{it} + \eta_i + \phi_{tz} + \epsilon_{it} - \theta\bar{u}_{it} + u_{it}.$$

where we define $\bar{\tilde{y}}_{it} = \frac{1}{|P_i|}\sum_{i' \in P_i} \tilde{y}_{it}$ and $\bar{u}_{it} = \frac{1}{|P_i|}\sum_{i' \in P_i} u_{it}$.

The fundamental identification issue in this setting is that the remote sensing measurement error can enter through $\bar{u}_{it}$ and at the same time may be correlated with factors that influence peers to move into certain neighborhoods. For example, if people are more likely to move to neighborhoods with larger lots and larger lots have more measurement error, then the peer measurement error $\bar{u}_{it}$ will be correlated with the instrument (the fraction of peers who moved into the peer group), leading to an invalid IV strategy. We might expect larger lots to have more measurement error because in larger lots a larger fraction of the lot is not covered by the house itself and one would expect that the non-house pixels are more subject to measurement error (it is easy to identify the house in the remote sensing imagery).

We thus recreate our landscape measures drawing upon the water consumption data. We know (from basic botany) that the greenness of a lot in a climate zone such as Phoenix is a direct function of how much water per square foot is used for irrigation. This implies:

$$\begin{aligned} S_i * y_{it} &= \sum_{m \in t} q^{out}_{itm} \psi_i + \xi^{out}_{it} + \omega_t \\ &= \sum_{m \in t} q_{itm} \psi_{im} - \sum_{m \in t} \psi_{im} q^{in}_{itm} + \xi^{out}_{it} + \omega_t \\ &\equiv \sum_{m \in t} q_{itm} \psi_{im} + \zeta_i + \omega_t + \xi_{it}, \end{aligned} \tag{3}$$

where S_i is the lot size of house i (since more water is needed for larger lots), q^{out}_{itm} is outdoor water usage in month m of year t, q^{in}_{itm} is monthly indoor water usage, q_{itm} is total monthly water usage, ζ_i is a household fixed effect, and ω_t is a year fixed effect.

By substitution we know that

$$S_i * \tilde{y}_{it} = \sum_{m \in t} q_{itm} \psi_m + \zeta_i + \omega_t + \xi_{it} + S_i u_{it}. \tag{4}$$

We can then estimate $\hat{\psi}_m$ with a standard fixed effects regression under the assumption that changes in indoor water use are not correlated with the measurement error (after conditioning on each household's fixed effect). We thus obtain the the fitted $\hat{y}_{it}$:

$$\hat{y}_{it} \equiv \frac{1}{S_i} \left(\sum_{m \in t} q_{itm} \hat{\psi}_m + \zeta_i + \omega_t \right).$$

This fitted $\hat{y}_{it}$ relies on the variation in the landscape greenness that stems from variation in monthly water consumption. Intuitively, this is useful because variation in water consumption will not have the measurement error inherent in the remote sensing images. This fitted variable is the one we use in all of our landscape estimation, both for the focal and peer variables.

To see why this approach addresses the issue of potentially correlated measurement error, first note that by rearranging (4) and plugging in $\hat{y}_{it}$, we have:

$$y_{it} = \hat{y}_{it} + \frac{1}{S_i}\xi_{it} + e_{it}, \tag{5}$$

where e_{it} is a mean-zero estimation error.

Now, by substitution of (5) into (1), we have:

$$\hat{y}_{it} = \theta\bar{\hat{y}}_{it} + \delta m_{it} + \eta_i + \phi_{tz} + \nu_{it},$$

where we define $\bar{\hat{y}}_{it} = \frac{1}{|P_i|}\sum_{i' \in P_i} \hat{y}_{it}$, $\bar{\xi}_{it} = \frac{1}{|P_i|}\sum_{i' \in P_i} \xi_{it}$, and $\nu_{it} = \epsilon_{it} + \frac{1}{S_j}\bar{\xi}_{it} - \frac{1}{S_i}\xi_{it} + \frac{1}{S_j}e_{it} - e_{it}$.

What is useful about this equation is that this specification no longer has u_{it}. Thus, with u_{it} removed, the fraction of households that moved into the peer group in a given year is again a valid instrument, as long as the unobservable in the indoor water use regression, ξ_{it}, is uncorrelated with the move decision after conditioning on the fixed effects. Moreover, this approach cleans the data in a way that develops a more sensible and consistent time series of landscape data over time.

We believe that this triangulation of the remote sensing image data and water consumption data is essential to provide an accurate measure of actual landscape values. When we run our preferred IV strategy on the uncorrected landscape data prior to the data cleaning exercise, we largely just find noise–the peer landscape coefficient is spurious and very poorly identified. In our conversations with remote sensing experts, they predicted that this could be the case.

The only final wrinkle is that ψ_{it} is likely to be different for SRP-eligible households than non-eligible households, since less water is used for outdoor purposes for SRP houses. Thus, we perform the data cleaning regression separately for each group of households. Table A.1 below shows the results of the regression used to obtain $\hat{y}_{it}$ as described above in (4). One immediate finding is that we explain 84% of the variation in landscaping through the water consumption data. The fact that so much variation is explained is not surprising, but it is reassuring that the methodology is retaining much of the variation in landscaping. In addition, the results can be interpreted as showing strongly positive correlations between the monthly water consumption in the driest months of the year and the fraction of green landscaping.

Table A.1: Regression Used to Obtain $\hat{y}_{it}$

	(1) non-SRP	(2) SRP	(3) Full Sample
Water Consumption			
January	-1.51***	-0.56***	-1.09***
	(0.17)	(0.18)	(0.12)
February	-3.05***	-1.56***	-2.29***
	(0.22)	(0.21)	(0.16)
March	3.10***	-0.40	1.77***
	(0.40)	(0.49)	(0.31)
April	0.92***	0.29**	0.58***
	(0.15)	(0.13)	(0.10)
May	1.32***	1.45***	1.46***
	(0.28)	(0.33)	(0.22)
June	2.90***	1.92***	2.52***
	(0.28)	(0.27)	(0.20)
July	-0.04	1.73***	0.31**
	(0.17)	(0.33)	(0.15)
August	0.61***	0.17	0.56***
	(0.15)	(0.24)	(0.13)
September	-0.97***	-0.61***	-0.77***
	(0.18)	(0.22)	(0.14)
October	2.37***	1.33***	1.76***
	(0.22)	(0.17)	(0.14)
November	-0.22	0.88***	0.51***
	(0.16)	(0.15)	(0.11)
December	3.27***	1.51***	2.14***
	(0.20)	(0.16)	(0.12)
Household Fixed Effects	Y	Y	Y
Year Fixed Effects	Y	Y	Y
R-squared	0.84	0.84	0.84
N	340,643	214,239	554,882

Notes: Dependent variable is fraction of green landscaping in a given parcel per square foot, i.e., lot size x percentage green. The covariates are the water consumption for each month of the year. Standard errors are clustered at the subdivision level. *** denotes significance at 1% level, ** at 5% level, * at 10% level.

B Further Data Details

B.1 Details on Data Cleaning

The raw water consumption data set contains 2,623,751 observations and 311,049 households. Often the utility will make billing adjustments by manually changing a household's water consumption in a particular month to either credit or charge an account. Hence, negative and extremely large consumption values are either errors or billing adjustments and do not reflect actual consumption. Accordingly, we drop annual consumption below the 1st percentile of consumption and above the 99th percentile of consumption, as these are very likely to be outliers that do not reflect actual consumption. The remaining data set has 2,572,534 observations and 310,040 households. Finally, including household-level fixed effects leads to 33,022 singleton observations, resulting in an unbalanced panel in the final dataset of 2,533,715 observations and 308,652 households.

The raw fitted landscaping data set contains 544,882 observations and 74,112 households. Again, we drop observations below the 1st percentile of consumption and above the 99th percentile of consumption for similar reasons as above. The remaining data set contains 540,451 observations and 72,007 households. Finally, including household-level fixed effects leads to 7,313 singleton observations, resulting in a final data set of 531,650 observations and 71,477 households.

B.2 Further Summary Statistics

This appendix section contains summary statistics for some of the variables used in our analysis of heterogeneity of the peer effect. We also include a histogram of our cleaned landscape variable, which demonstrates substantial variation.

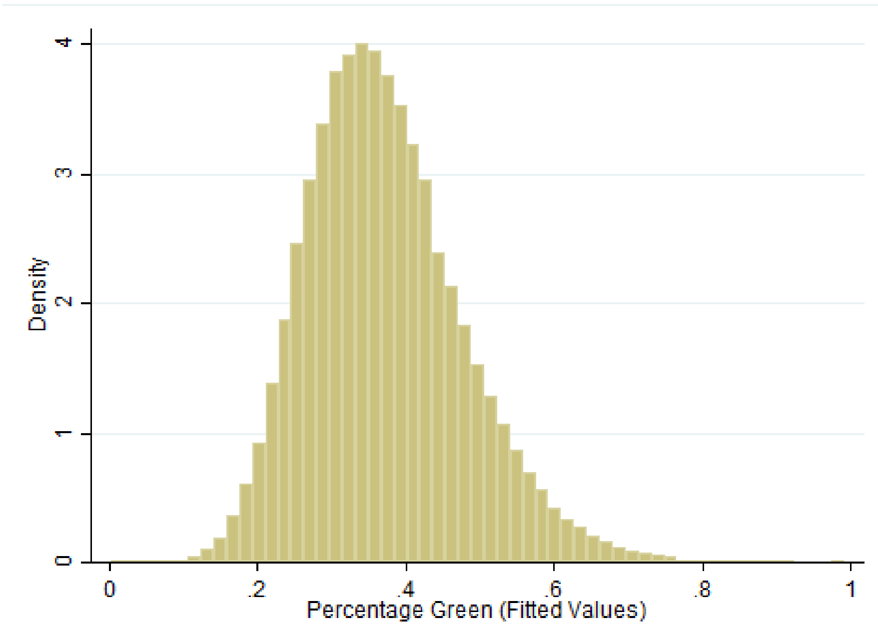

Figure A.1: Histogram of the percentage of the land area of a parcel that is green space, as coded by the machine learning algorithm and adjusted to address remote sensing measurement error. Values truncated at zero and one for clarity.

Table A.2: Summary Statistics for Household Characteristics

Variable	Mean	Std. Dev.	Min.	Max.	N
# bathroom fixtures	7.09	2.60	1	40	308,256
home size (ft^2)	1791.2	713.2	240	21012	308,263
construction date	1978.6	19.2	1900	2012	308,263
pool dummy	0.33	0.47	0	1	308,652
1(SRP project)	0.43	0.50	0	1	308,652
lot size (ft^2)	8,808	6,036	531	222,790	308,652

Table A.3: Summary Statistics for Demographic Variables

Variable	Mean	Std. Dev.	Min.	Max.	N
lot size (ft^2)	9,868.6	6,256.6	1,536.9	211,027.6	71,477
income bracket	7.9	3.6	1	13	59,139
age (years)	56.9	14.9	18	99	45,074
home value	10.2	4.2	1	19	58,567
number of kids	0.52	0.84	0	6	59,139
house size (ft^2)	1821.8	665.9	288	10343	71,386
construction date	1971.3	15.4	1900	2012	71,386
1(pool)	0.42	0.50	0	1	71,477
1(cat)	0.13	0.34	0	1	59,140
1(dog)	0.22	0.42	0	1	59,140
1(renter)	0.08	0.27	0	1	59,139
1(democrat)	0.23	0.42	0	1	59,140
1(group sports)	0.46	0.46	0	1	59,140
1(group outdoors)	0.39	0.49	0	1	59,140
1(group travel)	0.50	0.50	0	1	59,140
1(group reading)	0.62	0.49	0	1	59,140
1(group cooking)	0.50	0.50	0	1	59,140
1(group exercise)	0.50	0.50	0	1	59,140
1(group music)	0.40	0.49	0	1	59,140
1(group electronics)	0.64	0.48	0	1	59,140
1(group home improvement)	0.57	0.50	0	1	59,140
1(group investing)	0.45	0.50	0	1	59,140
1(group antiques)	0.35	0.48	0	1	59,140
1(group SOHO)	0.09	0.28	0	1	59,140
distance to school (000s ft)	1.49	1.04	0	8.08	71,477
distance to park (000s ft)	2.73	1.92	0	10.96	71,477

Notes: Distance to school or park may be zero due to how ArcMap defines boundaries of polygons. Demographic variables only apply to remote sensing image data set, where an observation is a household parcel-year. The units of home value are hundreds of thousands of dollars.

B.3 Tables of Balance

The following three tables display comparisons of summary statistics of important observable household and demographic characteristics for key subsets of our data. Table A.4 compares the means of key variables between the water consumption data and the landscaping data. Table A.5 compares the means of key variables between the off and on-SRP households in the water consumption data while Table A.6 compares the means of key variables between the off and on-SRP households in the landscape data. We performed two-sided t-tests of differences in means for each variable and report them in parentheses below the differences in means in column 3 of each table. The off and on-SRP households are statistically different from one another in each of the data sets. Hence, in our primary analysis of off-SRP households (Table 4 column 5) we use a propensity score matching technique to control for potential selection bias.

Table A.4: Summary Statistics by Water and Landscape Data Sets

	Water Data	Landscape Data	Difference
Median Household Income	60,132	63,126	-2993
	(44.1)	(89.5)	(-29.0)
% White	70.3	77.8	-7.5
	(0.03)	(0.05)	(-104.7)
% Black	5.5	4.1	1.5
	(0.01)	(0.01)	(69.0)
% Latino	34.9	22.7	12.2
	(0.05)	(0.07)	(111.3)
# of Bath Fixtures	7.1	7.0	0.07
	(0.005)	(0.009)	(6.4)
House Size (ft^2)	1,795	1,838	-42
	(1.3)	(2.5)	(-14.0)
Construction Date	1979	1972	6.5
	(0.04)	(0.06)	(81.9)
Lot size (ft^2)	8,941	9,968	-1027
	(12.6)	(24.0)	(-35.5)
1(Pool)	0.33	0.43	-0.10
	(0.001)	(0.002)	(-51.1)
N	308,652	71,477	308,652

Notes: Column 1 reports means for households in the water consumption data with standard deviations in parentheses. Column 2 reports means for households in the landscape data with standard deviations in parentheses. Column 3 reports the difference in means with t-stats for a two-sided test of differences in means in parentheses.

Table A.5: Summary Statistics by On and Off-SRP Households: Water Data

	Off-SRP	On-SRP	Difference
Median Household Income	71,048	45,913	25,136
	(58)	(44)	(328)
% White	80.9	56.5	24.4
	(0.03)	(0.04)	(520.4)
% Black	3.74	7.89	-4.15
	(0.01)	(0.02)	(-226.7)
% Latino	18.0	56.8	-38.8
	(0.04)	(0.06)	(-558.5)
# of Bath Fixtures	7.88	6.08	1.80
	(0.006)	(0.006)	(201.9)
House Size (ft^2)	1,975	1,561	413
	(1.85)	(1.56)	(164.2)
Construction Date	1985	1971	14.1
	(0.034)	(0.059)	(217.9)
Lot size (ft^2)	9650.2	8048.8	1601.4
	(19.5)	(14.1)	(63.6)
1(Pool)	0.43	0.19	0.24
	(0.001)	(0.001)	(145.6)
N	174,840	133,812	308,652

Notes: Column 1 reports means for off-SRP households in the water consumption data with standard deviations in parentheses. Column 2 reports means for on-SRP households in the water consumption data with standard deviations in parentheses. Column 3 reports the difference in means with t-stats for a two-sided test of differences in means in parentheses.

Table A.6: Summary Statistics by On and Off-SRP Households: Landscape Data

	Off-SRP	On-SRP	Difference
Median Household Income	70,460	50,694	19,767
	(107.7)	(122.5)	(117.0)
% White	83.1	68.9	14.3
	(0.03)	(0.08)	(177.5)
% Black	3.25	5.44	-2.20
	(0.01)	(0.02)	(-115.6)
% Latino	14.9	35.9	-21.0
	(0.04)	(0.13)	(-180.3)
# of Bath Fixtures	7.53	6.19	1.3
	(0.01)	(0.01)	(74.1)
House Size (ft^2)	1,925	1,689	236
	(3.12)	(4.10)	(46.0)
Construction Date	1979	1959	20.1
	(0.05)	(0.09)	(209.9)
Lot Size (ft^2)	10,422	9,199	1223
	(33.8)	(29.6)	(24.7)
1(Pool)	0.51	0.30	0.21
	(0.002)	(0.003)	(53.6)
N	42,203	29,274	71,477

Notes: Column 1 reports means for off-SRP households in the landscape data with standard deviations in parentheses. Column 2 reports means for on-SRP households in the landscape data with standard deviations in parentheses. Column 3 reports the difference in means with t-stats for a two-sided test of differences in means in parentheses.

C First Stage Results and Robustness Checks

This appendix provides the results from the first stage of the IV specifications in our primary results, as well as an additional set of robustness checks and further results. We begin by exploring how our primary water consumption results change if we restrict the sample to the subsample that has the landscape greenness coded.

Table A.7: Water Regressions On Remote Sensing Image Subsample

	(1) OLS	(2) OLS	(3) IV	(4) IV
log(consumption within 500′)	0.14***	0.08***	0.36***	0.24***
	(0.01)	(0.01)	(0.07)	(0.09)
1(housing transaction)	-0.18***	-0.18***	-0.18***	-0.18***
	(0.006)	(0.006)	(0.006)	(0.006)
Household Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y		Y	
Year x Zip Code Fixed Effects		Y		Y
R-squared	0.80	0.80	0.80	0.80
N	533,426	533,426	533,426	533,426

Notes: Specifications are identical to those in Table 2 but estimated on the remote sensing image subsample. Standard errors are clustered at the subdivision level. *** denotes significance at the 1% level, ** at the 5% level, * at the 10% level.

Table A.8: First Stage Preferred IV for Water and Landscape Data

	(1) log peer consumption	(2) log peer landscape
mean sold houses radius 500	-0.2029***	-0.0096***
	(0.0072)	(0.0012)
move this year	-0.0031***	-0.0000
	(0.0006)	(0.0001)
Household Fixed Effects	Y	Y
Year by Zip Fixed Effects	Y	Y
R-squared	0.947	0.421
N	2533715	538963

Notes: Dependent variable in column 1 is the log of average water consumption for $i's$ peers within a 500 foot radius. Dependent variable in column 2 is log of average green landscaping for $i's$ peers within a 500 foot radius. Mean sold houses radius 500 is the average number of sold houses with $i's$ peer group. Standard errors are clustered at the subdivision level. *** denotes significance at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

Table A.9: Regressing Landscape Fixed Effects on Demographics

	(1) Off-SRP	(2) On-SRP	(3) Full Sample
distance to school (1,000 ft)	-0.0005***	0.0032***	-0.0006***
	(0.0002)	(0.0003)	(0.0002)
distance to park (1,000 ft)	-0.0011***	0.0014***	-0.0003***
	(0.0001)	(0.0002)	(0.0001)
lot size (1,000 sq ft)	1.4679***	7.9256***	2.1484***
	(0.0340)	(0.0879)	(0.0317)
income bracket	0.0012***	0.0018***	0.0016***
	(0.0001)	(0.0001)	(0.0001)
age	-0.0001***	0.0001***	0.0000
	(0.0000)	(0.0000)	(0.0000)
home value	0.0019***	0.0034***	0.0034***
	(0.0001)	(0.0001)	(0.0001)
number of kids	-0.0003	-0.0021***	-0.0009***
	(0.0002)	(0.0003)	(0.0002)
house size (1,000 sq ft)	2.7294***	-21.0662***	-0.3607
	(0.4288)	(0.6629)	(0.3580)
construct	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)
has pool	-0.0141***	-0.0234***	-0.0175***
	(0.0004)	(0.0006)	(0.0003)
has cat	0.0065***	0.0053***	0.0061***
	(0.0005)	(0.0008)	(0.0005)
has dog	0.0032***	0.0002	0.0020***
	(0.0005)	(0.0008)	(0.0004)
renter	0.0053***	-0.0041**	0.0022*
	(0.0016)	(0.0019)	(0.0012)
democrat	0.0017***	0.0014**	0.0020***
	(0.0004)	(0.0006)	(0.0003)
group sports	-0.0013**	0.0049***	0.0008*
	(0.0005)	(0.0007)	(0.0004)
group outdoors	0.0020***	0.0002	0.0009**
	(0.0005)	(0.0007)	(0.0004)
group travel	0.0032***	0.0009	0.0025***
	(0.0005)	(0.0007)	(0.0004)
group reading	-0.0002	0.0077***	0.0033***
	(0.0007)	(0.0009)	(0.0006)
group cooking	0.0079***	0.0072***	0.0077***
	(0.0006)	(0.0008)	(0.0005)
group exercise	-0.0034***	-0.0019**	-0.0032***
	(0.0006)	(0.0008)	(0.0005)
group music	-0.0034***	-0.0050***	-0.0042***
	(0.0005)	(0.0008)	(0.0004)
group electronics	0.0034***	0.0019**	0.0029***
	(0.0007)	(0.0008)	(0.0005)
group home improvement	0.0031***	0.0069***	0.0048***
	(0.0006)	(0.0009)	(0.0005)
group investing	-0.0013***	-0.0055***	-0.0034***
	(0.0005)	(0.0007)	(0.0004)
group antiques	0.0006	0.0009	0.0009**
	(0.0004)	(0.0007)	(0.0004)
group SOHO	0.0029***	-0.0008	0.0013**
	(0.0006)	(0.0009)	(0.0005)
R-squared	0.043	0.185	0.097
N	215557	123684	339241

Notes: Results of regressing post-estimation fixed effects from Table 4 column 4 on demographics. Note we do not have demographic information and household characteristics for all households. Hence the number of observations in Column 3 is fewer than the number of observations in our main landscape specification. *** denotes significance at the 1% level, ** at the 5% level, * at the 10% level.

Table A.10: Additional Robustness Checks (IV Specification)

	(1)	(2)	(3)	(4)
	Water Consumption		Landscape	
log(consumption within radius)		0.1215***		0.2027***
		(-0.0336)		(-0.0875)
lag of log(consumption within radius)	0.0718*		0.096	
	(-0.0122)		(0.073)	
lag of 1(housing transaction)	-0.012***		-0.001***	
	(-0.0016)		(-0.0002)	
Household Fixed Effects	Y	Y	Y	Y
Year x Zip Code Fixed Effects	Y	Y	Y	Y
R-squared	0.732	0.781	0.0355	0.0126
N	2,515,689	2,418,080	472,574	510,227

Notes: In the first two columns, the dependent variable is water consumption. In the last two, the dependent variable is the % landscape greenness. Columns 1 and 3 estimate the preferred model using lagged peer consumption/landscaping in place of contemporaneous peer consumption/landscaping instrumented by contemporaneous and lagged peer housing transactions. Columns 2 and 4 estimate the preferred specification limiting the sample to households that were not sold in year t. Standard errors are clustered at the subdivision level. *** denotes significance at 1% level, ** at 5% level, * at 10% level.

Table A.11: Propensity Score Matching Comparison

	(1) IV Off-SRP	(2) IV Matched
log(consumption within radius)	0.288***	0.489**
	(0.104)	(0.187)
1(housing transaction)	-0.007***	-0.01***
	(0.0004)	(-0.002)
Household Fixed Effects	Y	Y
Year x Zip Code Fixed Effects	Y	Y
R-squared	0.03	0.0092
N	328,620	230,372

Notes: Column 1 is the primary landscaping regression for off-SRP households only. Column 2 is identical to column 5 in Table 4, which implements a propensity score matching routine to control for selection bias between on and off-SRP households. Standard errors are clustered at the subdivision level. *** denotes significance at 1% level, ** at 5% level, * at 10% level.