Welfare Effects of Home Automation Technology with Dynamic Pricing*

Bryan Bollinger Wesley R. Hartmann
Fuqua School of Business Graduate School of Business
Duke University Stanford University
April 1, 2015

Abstract

A fixed cost investment in home automation technology can eliminate consumers’ marginal costs of responding to changing demand conditions. We estimate the welfare effects of a home automation technology using a field experiment run by a large electric utility that randomly assigned both a technology and price treatment. Average treatment effects reveal that the home automation technology reduces demand more than twice as much as an alternative technology that only informs consumers of price changes. Furthermore, the average demand reductions during critical price events provide sufficient supply-side welfare gains to fully offset the installation costs of the device. Finally, we estimate household-specific treatment effects by matching households on their pre-treatment policy functions. This demonstrates the additional surplus gained by the utility if it targeted these treatments to households with the largest estimated demand responses.

*Email addresses are bryan.bollinger@duke.edu and wesleyr@stanford.edu. The authors would like to thank Severin Borenstein, Mar Reguant, seminar participants at the Energy Institute at Haas and conference participants at the 2012 Marketing Science Conference, the 2014 Field Experimentation Conference at Rady School of Business, the 2014 Workshop on Economics of Advertising and Marketing, 2014 IOFest, and the 2015 Winter Marketing-Economics Summit. Rachana Gogate provided valuable research assistance on this project.
1 Introduction

Home automation technologies shift households’ marginal costs of adjusting their consumption to a single upfront installation cost of the device. The technology increases a consumer’s demand elasticity by automatically executing her response, potentially in real-time, to weather and other demand shocks, and possibly to price. As both electricity and water expenditures are small within short time windows, it might take a long horizon for such technology investments to payoff for consumers. On the other hand, automated response technologies may provide significant supply-side benefits as utilities struggle to manage scarce supply. In fact, coupling a price-responsive automation technology with dynamic pricing may provide sufficient consumer demand reductions when capacity is constrained to justify the utility subsidizing installations of the home automation technology. We consider the welfare effects of such a home automation technology by analyzing a field experiment in which an electric utility randomly assigned households to information technology, automated response technology or neither and to pricing plans that included varying levels of dynamics.

Electric utilities have increasingly become interested in demand response (DR) to better match supply and demand. The Federal Energy Regulatory Commission defines DR as: “Changes in electric usage by end-use customers from their normal consumption patterns in response to changes in the price of electricity over time, or to incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized.” Reasons for supply shocks in electricity include forced outages, transmission line failures, and sharp changes in generation. These factors coupled with the high costs of energy storage result in large changes in wholesale prices over time and vary by an order of magnitude over the course of a day. This problem is exacerbated with the increase in renewable energy generation, because renewable sources such as solar and wind can fluctuate greatly in their electricity generation. Borenstein and Holland (2005) demonstrates that real-time pricing, which passes wholesale price swings through to consumers, can yield long-term efficiency gains even with low demand elasticities. However, a major hurdle to implementing
real-time pricing is the risk transferred to consumers who are uninformed or unequipped to respond to price changes. Home automation technology can resolve this problem by enabling response.

To evaluate how combinations of technology and dynamic pricing could help manage peak load demand, an electric utility in the southern United States conducted a field experiment in the summer of 2011. Households interested in participating were randomly assigned to a control group or a price-technology paired treatment. Usage patterns for households in the control group are similar to those who did not opt in to the experiment. Treated households were placed in either a time of use (TOU) or variable peak price (VPP) plan. TOU households faced a single elevated price during peak hours, whereas the VPP households faced a peak period price that varied by day depending on the supply and demand conditions at the utility. Treated households were also assigned i) an in home display (IHD) to reveal price information, ii) a programmable communicating thermostat (PCT) that cooled the house based on consumers’ weighting of savings vs. comfort, iii) both technologies, or iv) neither. All households also had access to a web portal with price and consumption information. Average treatment effects reveal that adding the home automation PCT on top of the information only IHD technology nearly doubles the critical peak demand reduction for customers on a TOU plan and triples the critical peak demand reduction for customers with VPP.

To estimate the consumer welfare effects of the home automation technology, we compare the demand curves when consumers have both an information and automation device to the demand curve when consumers only have the information device. Our field experiment does not have random variation in price, but we can condition on the demand intensity of the randomly assigned control group, which faces a flat price. This focuses inference on price variation arising from factors outside the participants in the experiment, such as supply shocks and demand shocks exclusive to commercial customers or customers in other cities. The identifying assumption is an absence of price correlated demand shocks which only shift
the behavior of treated households. Using this approach we demonstrate that the home automation technology does result in more elastic demand with greater reductions at high prices and expanded demand when prices are low. This increases consumer welfare by just over ten percent of the $250 cost of the device, per year. These consumer welfare gains would likely be even greater in the presence of true real-time pricing which is more variable and lacks the price ceiling in our experiment.

The electric utility also increases its welfare because the demand reductions during critical periods defer investments in expanded capacity. Using their assumption that a permanent kWh decrease in critical peak demand yields a net present value of $700, the break-even demand reduction is much less than the average incremental demand reduction estimated for the home automation technology. The utility could however increase its surplus if it targeted the device to only those homes where the gains fully offset the installation costs. To evaluate this, we estimate household-level treatment effects that match households based on their pre-treatment consumption. A treatment effect that conditions on our potential target’s vector of past usage is unrealistic, but we argue that conditioning on the distribution of usage is both practical and intuitively appealing. The distance in usage distributions between our focal household and all other households in the data can be calculated with the Kolmogorov statistic (Kolmogorov, 1933), which provides a probability that two distributions are the same. Furthermore, we can calculate the distance between usage distributions conditional on obvious states in a consumer’s electricity consumption problem (e.g. hot vs cold, weekend vs weekday, high vs. low past usage). This effectively matches households on their pre-treatment policy functions which should be indicative of the unobserved ways in which households differ. Targeting households with a household-level treatment effect large enough for the utility to break even on the home automation technology increases the supply side surplus per installation by 30% to $425.

Our analysis provides substantive, policy, and methodological contributions to economics. The economics of acquiring information has been extensively studied since Stigler (1961)’s
seminal paper. Buyers incur search costs to resolve their uncertainty about sellers prices at “any given time.” Search costs have drastically fallen as price and other information can more be easily found online or communicated by electronic mail, text messages or notifications on mobile phones or wearable devices. Abundant or cheap information does not however appear to return us to the perfect information world as the price dispersion Stigler observed is still great for apparently homogeneous goods and large costs for price search have been measured both online (Honka, 2014) and across the aisles or shelf facings in a grocery store (Seiler, 2013). Our information only treatment yields significant demand response that is consistent with these findings. Nevertheless, much larger gains arise with investments in technology that automates the response to information. Such technologies are increasingly important as they can reduce or eliminate buyers’ costs of acquiring, processing, and responding to the abundance of information inundating consumers today.

From the perspective of energy policy, our findings document the advantages of moving utilities to dynamic pricing plans if they can provide the technology for consumers to respond. As speculated in Allcott (2011), we find that a programmable communicating thermostat which can respond to price (in contrast to a traditional thermostat which responds only to temperature) provides significant improvements in demand response. The price treatments in our study involved day ahead price setting in all but the critical cases, but that timing of notification had no impact on the demand reduction of the automation technology. This suggests real-time pricing could instead be coupled with the automated response technology to provide even greater welfare gains. The presence of heterogeneity in demand response that we document is consistent with previous work, such as Reiss and White (2005). However, the authors find that nonlinear pricing in California exhibits a highly skewed distribution in DR effectiveness such that a small fraction of households accounts for the majority of the response. Their high responders were likely households with low marginal costs of adjusting

---

1Cappers, Goldman, and Kathan (2010) show that DR without home automation has increased by 10% since 2006 in reducing peak load and that existing DR resource potential ranges from 3 to 9% of a region’s summer peak demand in most regions.
demand across multiple appliances. In contrast, we find the home automation technology to reduce the marginal adjustment costs for the average customer such that our heterogeneous treatment effects reflect the heterogeneity among the larger group of customers that otherwise would have found it too costly to adapt behavior.

From a methodological perspective, we propose a household targeting strategy using an average treatment effect weighted by the distance between each household in the sample and the focal household. The program evaluation literature currently uses observed heterogeneity to shift between, for example, sample average, population average and treatment effects on the treated or controls (see for example Abadie, 2002 and Imbens, 2004). Our focus on weighting around a focal household arises from targeting and its supply-side surplus gains. We contrast with extant targeting approaches using Bayesian estimation and a hierarchical structure that “shrinks” inference to individual units based on the time dimension of the panel (Rossi, Allenby, and McCulloch, 2005). In most field experiments, the randomization is purely cross-sectional at a single point in time, such that the length of the panel provides no additional inference on the individual’s specific treatment effect. Rather, we use pre-treatment outcomes to draw inferences about the similarities and differences among the cross section of subjects in the experiment. Our weighted average treatment effects lack the full structure of the data generating process typically included in the Bayesian approach, but in our context and other field experiments, targeting decisions may be assessed by simply comparing the household weighted average treatment effect to a simple threshold.

Our matching estimator also flips the problems associated with too much pre-treatment information on each unit into a benefit when that information can be viewed as outcomes of economic agents’ policy functions. The matching literature has long struggled with excessive pre-treatment information. The propensity score estimator, for example, collapses a high dimensional set of observables into a single metric based on a first stage estimation of how observables relate to treatment assignment (Rosenbaum and Rubin (1983)). Proliferation of observables about subjects has even overtaxed this estimator as they may be greater than the
number of subjects whose treatment assignments we observe. Variable selection models have since been applied to systematically reduce the amount of information used in estimation. In contrast, the Kolmogorov distance measure in our matching estimator becomes more precise with more observed pre-treatment outcomes. While we apply this matching estimator in an experimental setting to recover the household weighted average effects, this matching approach could also be used to reduce bias in non-experimental settings in much the same way as the propensity score, but in a context where increasing information about the subject is a benefit as opposed to a problem.

The paper proceeds as follows. Section 2 describes our unique data set and experimental treatments. In section 3, we estimate average treatment effects that illustrate how a combination of technology and flexible pricing can be effective in reducing electricity usage during peak periods that tax the grid’s capacity and call into use less efficient and potentially more polluting production capacity. Section 4 illustrates the demand estimation with and without home automation to recover consumer welfare effects. In Section 5 we consider the supply side and use the estimation of household level treatment effects to illustrate the electric utility’s gain in surplus if it targets PCT devices as opposed to offering them to all households based on the average demand reductions. We provide concluding remarks in Section 6.
2 Data

2.1 The Experiment

The goal of providing the smartmeter enabled technologies to consumers was to help them monitor and reduce their consumption during critical pricing events. There were seven events in 2011 which all occurred sometime in the window between 1 PM and 7 PM. In general, the period between 2 PM and 7 PM on non-holiday week days is the peak demand period. Outside this window and on weekends and holidays are off-peak periods.

Households who volunteered for a demand-reduction program were randomly assigned to one of nine conditions. Aside from the control condition, the treatment conditions combine one of four technology treatments with one of two pricing plans. The four technologies treatments are: i) a computer portal to monitor usage and price, ii) an in-home display that removes the need to be online to assess usage and price, iii) a programmable thermostat which can be set to turn off air conditioning based on the time, in-home temperature, or price, and iv) a combination of all three technologies. See Figures 1 and 2 for pictures of the IHD and PCT technologies.

The pricing treatments are i) time of use pricing (TOU) which sets different prices for the peak and off-peak hours on weekdays and ii) variable peak pricing (VPP) in which the peak price can be varied by the utility depending on its aggregate demand. We have hourly electricity usage for 4,443 residential households in 2011; 2,589 households appear in both the 2010 and 2011 data, with 5,491,443 hours of electricity usage data. We exclude from
the analysis households who have the low-income price rate (all assigned non-randomly to the control group) and those without a treatment start date recorded, as well as accounts with multiple meter ids. We also exclude 2011 consumption data for treated households before their treatment. Since the vast majority of homes have air conditioning, we exclude those without. Finally, we exclude households in the extreme tails of the distribution of 2010 usage, 2011 peak usage, and 2011 off-peak usage i.e. those with more than 7 kW or less than 0.5 kW of average usage. The final estimation sample includes 2,178 households with 6,159,548 hours of electricity usage data in 2010 and 4,627,987 hours of electricity usage data in 2011. 294 households are in the control condition, with the remaining spread across the various treatments, as shown in Table 1.
Figure 4: Average Hourly Electricity Consumption by Treatment
Figure 3 shows the household average hourly usage by treatment – it is clear that all of the technology treatments (with the new price treatments) do lead to an aggregate reduction in usage. The timing of the reduction in usage can be seen in Figure 4, which plots average hourly usage by treatment, on both peak and off-peak days. It is clear that all treatment conditions lead to a reduction in electricity consumption during the peak period of peak days. Consumption is then slightly higher both during the off-peak hours of peak days and on off-peak days, indicating that there might be inter-temporal substitution in electricity consumption. The clear superiority of the PCT device is evident with the sudden reduction in electricity consumption at the beginning of peak hours (14:00). Note however that a wider rollout of these devices would call for a smoothing of the start times to avoid the sudden demand shock this would create.

Figure 5 shows kernel density estimates of the average hourly usage, where the averages are taken by household over the peak and off-peak periods of each day. There is a drop for all technology treatments during the peak hours but a small increase in usage during the off-peak hours. The PCT and All 3 technologies reduce peak demand much more than the portal and IHD.

We have not said anything about the statistical significance of these findings. While the effects are apparent in both the the usage and kernel density graphs, usage varies greatly across both time and households, leading to large variance in usage under all treatments in the kernel density graphs. As such, these graphs are not in themselves sufficient in determining whether the treatments have statistically significant mean treatment effects. To test the significance of the average treatment effects, we use a regression analysis described in the next section to not only assesses the impact of the technology treatments but also their interaction with the two price treatments.
Figure 5: Average Hourly Electricity Consumption by Treatment, On-Peak Days
3 Average Treatment Effects

To estimate the average treatment effects for our panel of electricity consumption in the summer following treatment, we use the following simple regression:

\[ y_i = \alpha_0 + \alpha_1 \tilde{A}_i + \epsilon_{ia} \]  

where \( y_i \) is household \( i \)'s average hourly electricity consumption in either critical, peak, or off-peak periods – we run separate regressions for each. We collapse all of the data across time within these three periods. \( \tilde{A}_i \) is a vector of dummies corresponding to each of eight treatments (the four technology treatments interacted with the two price treatments), with zeros for all but the household’s treatment \( A_i \). \( \alpha_0 \) captures the average usage of the control group, while \( \alpha_1 \) measures the change in usage attributable to each treatment; \( \epsilon_{ia} \) is the unobservable consumption shock that is uncorrelated with treatment because of the randomization. Results are shown in Tables 2. We report robust standard errors.

We see significant effects of all technology treatments for critical and peak periods, leading to large reductions in usage between 0.347 and 1.303 kW for critical periods and 0.200 and
Table 2: Average Treatment Effect Regression Results

<table>
<thead>
<tr>
<th>Treatments</th>
<th>Critical</th>
<th>Peak</th>
<th>Off-Peak</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TOU</td>
<td>VPP</td>
<td>TOU</td>
</tr>
<tr>
<td>Portal</td>
<td>-0.437</td>
<td>-0.281</td>
<td>-0.324</td>
</tr>
<tr>
<td></td>
<td>(0.137)</td>
<td>(0.136)</td>
<td>(0.123)</td>
</tr>
<tr>
<td>IHD</td>
<td>-0.671</td>
<td>-0.414</td>
<td>-0.491</td>
</tr>
<tr>
<td></td>
<td>(0.141)</td>
<td>(0.137)</td>
<td>(0.127)</td>
</tr>
<tr>
<td>PCT</td>
<td>-1.190</td>
<td>-1.219</td>
<td>-0.942</td>
</tr>
<tr>
<td></td>
<td>(0.141)</td>
<td>(0.141)</td>
<td>(0.127)</td>
</tr>
<tr>
<td>All three</td>
<td>-1.082</td>
<td>-1.231</td>
<td>-0.830</td>
</tr>
<tr>
<td></td>
<td>(0.134)</td>
<td>(0.149)</td>
<td>(0.131)</td>
</tr>
</tbody>
</table>

N=2,141 \hspace{1cm} N=2,178 \hspace{1cm} N=2,178

Standard errors in parentheses

0.918 kW for peak. It is clear that the PCT technology (used in the PCT and All 3 treatments) is the most effective of the three alternatives for lowering consumption during peak periods for households with AC. Furthermore, although the impact of the technologies are similar across price treatments for critical events, we find evidence that there is a main effect of the TOU pricing schedule in comparison to the VPP pricing schedule in lowering peak consumption. This could be due to average VPP prices being lower than TOU. There is no clear effect of technology treatment or price treatment in off-peak consumption behavior, although this is likely due to the aggregation over all off-peak periods in the case of the PCT and All 3 treatments since the hourly usage graph in Figure 4 clearly showed positive consumption effects immediately following the peak period.

4 Consumer Welfare Effects from Home Automation Technology

Home automation technology can reduce or eliminate consumers’ costs of adjusting demand to changing conditions. The technology therefore shifts consumers to a more elastic demand curve. To evaluate the consumer welfare effect of automated response technology, we compare
demand under the All 3 treatment with demand under the IHD treatment. The All 3 treatment is the best estimate of the “true” demand curve in that they can rely on the PCT algorithm to respond to price, but also override based on their enhanced knowledge of price delivered via the IHD. We compare this to the IHD treatment because it measures the gains of automation technology relative to merely informing consumers.

4.1 Demand Estimation

There are two potential sources of price variation we can use to estimate demand: i) across pricing plans (e.g. TOU vs. VPP) and ii) within peak variation in the VPP price. Variation across plans is based on the random assignment, but on any given day, there are at most two different peak prices necessitating strong assumptions about the shape of demand between and outside of these price points. Variation within the VPP price ranges the entire span from less than 5c to a high of 46c, yet much of this variation likely reflects different demand conditions. To use this variation for identification, we must therefore isolate the variation arising from either supply-side factors or demand shocks specific to customers outside the experiment. While we do not observe such instruments for price, we can however condition on the demand from the randomly assigned control group. Any demand factors affecting VPP assigned households should also affect control households. Thus conditioning on the demand of the control households, allows us to isolate price variation from the external factors mentioned above.

Consider the following demand and supply functions: $q(p_t, \xi_{It}, \epsilon_{it}, \theta_i)$ and $p(q(\xi_{It}, \xi_{-It}, \theta), Z)$, where $\xi_{It}$ represents demand factors common to all $i \in I$ households in the experiment and $\xi_{-It}$ represents demand factors common to all the utility’s customers outside the experiment (e.g. households and businesses in other cities). $\epsilon_{it}$ is an idiosyncratic shock to consumer $i$ at $t$. $\theta_i$ is $i$’s time invariant preferences and $\theta$ is the vector of all customers’ (in and out of the experiments) time invariant preferences. $Z$ represents supply shifters, excluded from demand. Without $Z$ and $\xi_{-It}$ the VPP price should not vary, conditional on $\xi_{It}$. Thus
variation we observe conditional on \( \xi_I \) must be arising from these exogenous sources. The only possible violation of our identification assumption might occur if \( \xi_I \) is changing but households assigned to the control condition are unable to respond because of their lack of portal access, IHD or PCT. However, all three of these technologies are oriented around price movements, which we observe. There are no obvious non-price factors that would alter these households’ demands and not control households’. Weather is the primary demand determinant for these households with air conditioning in a hot southern summer.

Figure 7 depicts the All 3 and IHD only (dotted) estimated demand curves for the 10th, median and 90th percentiles of control demand during peak periods. Confidence intervals for each point on the demand curves are depicted with error bars. As noted earlier, the home automation included in the All 3 demand curves clearly makes it more elastic. In fact, the IHD demand curve is effectively perfectly inelastic with insignificant movements in demand creating the illusion of a subtle upward slope at times. When aggregate (control) demand is low, the 10th percentile demand to the left illustrates the additional PCT device helps consumers expand demand when price reaches its floor of 4.5c. The largest price observed at this low demand is 23c. At larger aggregate demands, we observe the maximum 46c price as well. In these cases, the PCT device enables an additional demand reduction of 34 percent and 23 percent respectively for the median and 90th percentiles of control demand. As depicted in Figure 8, the addition of the PCT reduces expenditures in these cases, while expanding it when the low price is offered in a low demand state. We consider the lowest and highest prices for the 10th and 90th percentiles, while choosing the 23c price for the median. 23c is actually above the median for VPP customers, but is the fixed peak price for TOU customers.

Finally, the welfare in each case is depicted in Figure 9 by considering the surplus changes. The dark red shaded area in the figure to the left depicts the additional consumer surplus from expanding consumption at the 4.5c price point. In the middle and right figures, the dark-blue shaded area to the upper-right represents the extra expenditures with the IHD
Figure 7: All 3 vs. IHD Demand: 10th, 50th and 90th Percentiles of Control Demand

Figure 8: All 3 vs. IHD Expenditure Differences
device that are not offset by consumer utility.

4.2 Aggregating the Surplus Changes Across States

The figures above depict the surplus changes for particular demand states as represented by the 10th, median and 90th percentiles of demand. To estimate the total welfare effects, we must integrate over all such states. We do this by considering each tenth percentile of the control demand, solving for the welfare effects at each price observed in that percentile and then calculating the weighted average based on how frequently that price was charged within that percentile of control demand.

Across an entire summer, we estimate the welfare effects to be between $27 and $30. The variation is based on lower and upper bounds of the shape of the demand curve below the 4.5c price point. Elsewhere, we assume linearity between price points. If we relax the linearity assumptions and create lower and upper bounds based on forcing the demand curve to immediately drop either just before or after each price point, we obtain a range of $21 to $49. If we consider the net present value at a rate of 10%, these bounds are $213 to $486 with the welfare under linearity between price points between $273 and $296. In other words, it appears the consumer welfare might be enough to justify a household making the $250 investment to acquire the technology. But, the cost of advertising to communicate the benefits could make this unprofitable. Nevertheless, in this case the utility subsidized the cost
of the device and based on the demand reduction number reported above that investment paid off.

5 Firm Welfare Effects from Home Automation Technology

In addition to the consumer welfare effects of the PCT, the firm can benefit from reducing demand during periods of critical peak demand. The utility we are working with estimates the net present value of 1 kW of demand reduction in critical periods to be $700 due to the deferred investment in generation capital. The costs of supplying a PCT is $250. This means that the PCT needs to decrease critical period usage by \(-\frac{250}{700} = -0.357\) kW in order for the utility to want to subsidize the installation of a PCT in a consumers household. Based on the average treatment effects reported above, the utility is certainly incentivized to do this as the difference between the PCT and portal demand reductions are nearly a full kWh. In this section, we explore whether heterogeneity in these treatment effects creates an incentive for the utility to select (target) particular households for the installation of the PCT. To do this we first document the limited value of conditional treatment effects for uncovering heterogenous treatment effects, then propose a methodology for estimating household specific treatment effects. We then compare the welfare difference from installing a PCT in all households vs. restricting it to households where the expected treatment effect suggests it would reduce critical demand by more than .357 kW.

5.1 Treatment Effects Conditional on Observed Types

The most common form of targeted marketing is to condition on observable characteristics such as demographics. For example, the age of the household head or the size of the family could relate to the ability to use technology based treatments or to differences in the ability to time shift consumption of electricity. We therefore seek an expected difference in outcomes
that is conditional on demographics:

\[
\hat{\alpha}_{ax} = E [Y|A = a, X = x] - E [Y|A = 0, X = x]
\]

To determine heterogeneity in the treatment effects as a function of demographic variables, we regress critical usage on the eight treatment dummy variables (four technology treatments interacted with the two price treatments) interacted with the demographic variables, age and income. The estimated regression coefficients for the treatment and demographic interactions are shown in Figure 10.

While in general, the effects appear to be slightly larger for higher income households and those with a younger head-of-household, it may be the case that demographic variables are not the most effective information in explaining responsiveness to these different treatment effects. And in fact, we do have much more data for all of these households, namely their usage patterns in previous years. The next section discusses our methodology for defining consumer types as a function of their past usage and we demonstrate the usefulness in using this information in targeting households.

### 5.2 Individual Treatment Effects

Our goal is to develop an estimation approach that complements the randomized treatments by utilizing extensive pre-treatment consumption histories to estimate individual (household)-level treatment effects. A straightforward application of non-parametrically estimated conditional treatment effects would take expectations over all households \( j \), while conditioning on their vector of (pre-treatment) usage history, \( H_j \) being equal to a focal household \( i \)'s usage history, \( H_i \):

\[
\hat{\alpha}_{ai} = E [Y_j|A_j = a, H_j = H_i, X_j = X_i] - E [Y_j|A_j = 0, H_j = H_i, X_j = X_i]
\]
Figure 10: Treatment effects conditional on demographics

N=2,177. Includes controls for age, income, 2010 average hourly usage
Bootstrap standard errors shown by error bars
In this true non-parametric approach, $H$ includes each hourly interval over the summer of 2010, while $X$ includes demographics as in the previous section. The non-parametric estimator would weight each household $j$ based on its distance, $D_{ij}$ from the focal household $i$:

$$D_{ij} = \sum_t |H_{jt} - H_{it}|$$

Taking the number of households to infinity, we would converge to an estimate based on essentially “matching” households. The challenge is that this is a very high-dimensional problem. Furthermore, it is not efficient. For example, if two individuals roll identical dice each day to determine their usage, this distance will be quite large even though they are the same. Despite having identical “policy functions,” these two individuals would “match” with a very low probability that decreases quickly in the number of times we observe them. On the other hand, their distributions of state-dependent usage would converge the more we observe them.

We propose estimating individual-level treatment effects using non-parametric regressions that weight each household based on its Kolmogorov distance between its pre-treatment usage and that of the focal household. This distance has well documented converge properties for testing between differences in distributions such that it suggests itself as a valuable distance statistic for this type of non-parametric regression. In our application, we will have a balanced panel, but the Kolmogorov distance has the advantage that it can also be used in cases where the number of pre-treatment outcomes vary across units.

In addition to taking a simple Kolmogorov distance between the past usage of households, we also consider a measure that weights the distribution of usage across “states” in which a household’s policy function might differ. Examples include outside temperature, weekends vs. weekdays or past usage. The idea is that changes in the distribution across these conditions could uncover other important primitives of the household’s preferences for electricity. For example, if the distributions change dramatically with time of day (holding temperature fixed), this may indicate someone who may get less value from a device that
adapts their electricity usage, because they already adapt well.

Although we do not attempt to recover the policy functions and underlying structural parameters at each state, the motivation for this approach is based on recent development in the estimation of dynamic decisions using a two-step approach. Bajari, Benkard, and Levin (2007) follow Hotz and Miller (1993) in highlighting the importance of estimating policy functions in a first stage. More recently, Kasahara and Shimotsu (2009) and Arcidiacono and Miller (2011) have considered the identification and estimation of unobserved heterogeneity in these contexts. Extending our analysis to recover household-level policy functions and structural parameters would be necessary if the firm were to consider treatments other than the ones we observe, such as alternative price schedules.

5.2.1 The matching procedure

The household-level treatment effect, conditional on the distribution of past usage is:

\[ \hat{\alpha}_{ai} = E \left[ Y_j | A_j = a, F_{j,T_j,s} (h|s) = F_{i,T_i,s} (h|s) \right] \]

\[ - E \left[ Y_j | A_j = 0, F_{j,T_j,s} (h|s) = F_{i,T_i,s} (h|s) \right] \]

where the expectation is taken across households indexed by \( j \neq i \). \( F_{j,T_j,s} (h|s) \) is the empirical distribution function for \( T_{j,s} \) iid observations of \( j \)'s historical usage, \( H_{j,t} \), when \( j \) is in state \( s \). Note that the above estimator does not require that we know the distribution of \( h \) for \( i \), but that the expectations of post treatment outcomes \( Y \) are taken conditional on the empirical distribution of pre-treatment outcomes \( H \) coming from the same distribution as \( i \)'s. We therefore evaluate the equivalence of the pre-treatment outcome distributions in state \( s \) for \( i \) and any other candidate household \( j \) using the Kolmogorov-Smirnov statistic for two samples:

\[ D_{ij,s} = \sup_h \left| F_{j,T_j,s} (h|s) - F_{i,T_i,s} (h|s) \right| \]
Figure 11: Histogram of Kolmogorov-Smirnov statistics, with and without states

In our application, we consider the following state variables in the pre-treatment period (the summer of 2010): time of day, temperature, weekend vs. weekday, and recent electricity consumption. All state variables are discretized. Time of day is segmented into the following blocks of time: midnight to 6am, 6am to noon, noon to 6pm, 6pm to midnight. Temperature is categorized as cold, average, or hot using the 25th and 75th percentiles of the temperature distribution from the summer of 2010 as the boundaries. Finally, we summarize past electricity consumption as low or high consumption on the preceding day based on the median of the household’s daily consumption over the summer. This accommodates state-dependent consumption in which heavy laundry one day may alleviate the need to consume as much electricity the following day. Such effects could persist for more than one day, but the effect on the following day should be sufficient to separate households that do, or do not, exhibit state-dependent consumption; \( x \) therefore is a state-point that combines all three factors.

To create a distance measure that integrates over these states, we next derive the unconditional distance between \( i \) and \( j \) by taking a weighted average over state points:

\[
D_{ij} = \frac{\sum_s p_{i,s} D_{ij,s}}{\sum_s p_{i,s}}
\]  

(4)

where \( p_{i,s} \) is the fraction of the time household \( i \) is in state \( s \). Given we do not know the
true distribution $p_{i,s}$, we use the empirical distribution in its place. $D_{ij}$ is then plugged into a kernel to form a weight in a non-parametric regression. Weights are assumed to be zero if $A_j \neq a$. Thus some $j$ households will contribute to the first expectation in Equation 2 (if they were in fact in the focal treatment), while others will contribute to the latter expectation (i.e. if in they were in fact in the control condition).

The estimator will seek to reduce the influence of households with distances close to one, while heavily weighting those households close to zero. We consider two Kolmogorov based estimates. Our first estimator uses the Kolmogorov-Smirnov statistics in which we calculate the distance without conditioning on the states. In terms of the preceding equations, this is equivalent to assuming there is a single state. A histogram of these between households distances is depicted in the first graph in Figure 11. The second estimator uses the weighted average Kolmogorov distance across all of the state-specific distances. A histogram of these distances is shown in the second graph. The distribution here is shifted further to the right. One obvious reason for this is that each state-specific distance contains only a subset of the observations. Just as distances go to zero with additional observations, fewer observations tends to increase distances.

We estimate the treatment effects (and control variable coefficients) using a local constant estimator for each household $i$, where $K$ is the Epanechnikov kernel with optimal bandwidth $h_i$. As a reminder, $A_{i-}$ is the treatment dummy matrix and $X_{i-}$ are the demographic variables for all other households $j \neq i$ (indicated by $i-$), and $D_{ii-}$ and $Y_{i-}$ are the distance matrix to household $i$ (which depends on the estimator being used) and usage for all other households:

$$\hat{\theta}_i = [A_{i-}, X_{i-}]' \left( [A_{i-}, X_{i-}] \ast \frac{1}{h_i} K \left( \frac{D_{ii-}}{h_i} \right) \right)^{-1} [A_{i-}, X_{i-}]' \left( Y_{i-} \ast \frac{1}{h_i} K \left( \frac{D_{ii-}}{h_i} \right) \right).$$

Without $X$, this provides an average treatment effect on the individual or household. Adding
Table 3: Average SSE of estimators

<table>
<thead>
<tr>
<th></th>
<th>Average Usage</th>
<th>KS statistics</th>
<th>KS statistics as function of state</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>Fraction&lt;.5</td>
<td>Fraction&lt;1</td>
</tr>
<tr>
<td>Average Usage Critical</td>
<td>1.029</td>
<td>0</td>
<td>0.204</td>
</tr>
<tr>
<td>Average Usage Peak</td>
<td>0.663</td>
<td>0.044</td>
<td>1.000</td>
</tr>
<tr>
<td>Average Usage Off-Peak</td>
<td>0.184</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>KS statistics Critical</td>
<td>0.903</td>
<td>0</td>
<td>0.636</td>
</tr>
<tr>
<td>KS statistics Peak</td>
<td>0.575</td>
<td>0.004</td>
<td>1.000</td>
</tr>
<tr>
<td>KS statistics Off-Peak</td>
<td>0.150</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

\[ WSSE_i = \frac{\sum_{i=1}^{I} K \left(\frac{D_{ii}}{h_i}\right) \left(y_{-i} - X_{-i} \hat{\theta}_i\right)^2}{\sum_{i=1}^{I} K \left(\frac{D_{ii}}{h_i}\right) - L/2} \]

We see a reduction in the WSSE for the critical period regressions of 12.1% and 11.5% when using the KS matching and KS matching by state, respectively, compared to just matching on 2010 mean usage. The use of KS matching has great benefits in reducing the error, but the benefit of using KS matching by state appears limited (it is actually a little higher) when

\( X \) allows us to also include common observable heterogeneity in demographics.

For each household, we optimize the bandwidth by minimizing the weighted sum of squared errors (WSSE) where \( L \) is the length of \( \theta_i \):

We again estimate the treatment effect for critical pricing events, peak usage not during critical pricing events, and off-peak usage.

To assess the value of the rich histories of pre-treatment electricity usage, we also consider a benchmark model that calculates distance based only on average electricity usage in 2010. We also could include additional moments but then we lose the unidimensionality of the Kolmogorov statistic (KS). We compare this baseline model with matching on average usage to our KS based models by calculating the average sum of squared residuals across households for equation (5.2.1) for each. Table 3 shows the average SSE for each of the three methods.
Figure 12: Histogram of optimal bandwidths and WSSE focusing only on the mean. However, when looking at the number of households with WSSE less than 0.5 kW in the peak regressions, matching by state has a huge advantage since 38% of households are under this threshold compared to only 0.4 when ignoring state. We plot histograms of the bandwidth and resultant WSSE for the KS matching by state in Figure 12.

One final challenge that remains is the estimation of the standard errors. To do this we use bootstrapping, with replacement. Also, although we selected the bandwidth to minimize the WSSE, our estimator is biased in finite samples. In the results that we present in the next section, we report bias-corrected estimates, $\hat{\theta}_i^* = 2\hat{\theta}_i - \bar{\theta}_i$, where $\bar{\theta}_i$ is the average of the point estimates $\hat{\theta}_i^B$ over $B$ bootstrap samples, with $B = 100$. We report the bootstrapped standard errors, $\sqrt{\frac{1}{B} \sum_{b=1}^B (\hat{\theta}_i^B - \bar{\theta}_i)^2}$.

Since our method involves estimating the treatment effects for each household without using that household’s post-treatment usage data, one nice test of the benefit of using KS matching (with states) is to compare the prediction errors for the focal households to the
prediction error when matching based on 2010 average usage. We find that the prediction error for critical periods declines in magnitude from 0.824 to 0.802, peak declines from 0.663 to 0.622 and off-peak from 0.330 to 0.313.

5.2.2 Matching Results

We were able to estimate household-level treatment effects using state-specific distributions of past usage for 2146 out of the 2334 households who were not in the extreme tail of the usage distribution. The mean and standard deviation (with the average standard error) of the household-level treatment effects are shown in Table 4. These estimates of the average effects are similar to those reported in Tables 2. The important feature of this analysis is the standard deviation of the household effects. In many cases the standard deviation is greater than the average effect, indicating significant household heterogeneity in treatment effects. The most heterogeneity is observed for the most effective treatments: PCT and All Three. Notable between these is that the heterogeneity is greater for the All Three treatment despite its average effect being smaller than the PCT treatment. This may be due to heterogeneity in how the presence of an IHD in combination with the PCT altered adaptation of electricity consumption. Patterns across TOU and VPP pricing are similar, with the TOU generally performing better as reported in the average treatment effects.

To better illustrate the heterogeneity in the treatment effects, we plot the distribution of the effects in Figures 13 and 14 for TOU and VPP, respectively. We show the CDF of the treatment effects for both the peak and off-peak periods, relative to the control treatment. We also plot the CDF of the lower and upper bounds on the 95% confidence interval of the household-level estimates. We do this to give the reader a visual cue to the relative size of the household-level confidence interval – the confidence interval on the distribution itself is much tighter. We plot a vertical line at $x = 0$ to show the normalized usage for the control condition, and a horizontal line at $y = 0.5$ to assist the reader in assessing whether
Table 4: Summary of Household Treatment Effect Regression Results

<table>
<thead>
<tr>
<th>Treatments</th>
<th>Critical TOU Average</th>
<th>Critical TOU SD</th>
<th>Critical VPP Average</th>
<th>Critical VPP SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portal</td>
<td>-0.271 (0.164)</td>
<td>-0.351 (0.159)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IHD</td>
<td>-0.418 (0.178)</td>
<td>-0.224 (0.168)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCT</td>
<td>-1.079 (0.203)</td>
<td>-1.083 (0.185)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All three</td>
<td>-0.870 (0.197)</td>
<td>-1.078 (0.211)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatments</th>
<th>Peak TOU Average</th>
<th>Peak TOU SD</th>
<th>Peak VPP Average</th>
<th>Peak VPP SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portal</td>
<td>-0.195 (0.131)</td>
<td>-0.216 (0.126)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IHD</td>
<td>-0.327 (0.143)</td>
<td>-0.111 (0.134)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCT</td>
<td>-0.855 (0.156)</td>
<td>-0.645 (0.150)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All three</td>
<td>-0.712 (0.161)</td>
<td>-0.534 (0.169)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatments</th>
<th>Off-Peak TOU Average</th>
<th>Off-Peak TOU SD</th>
<th>Off-Peak VPP Average</th>
<th>Off-Peak VPP SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portal</td>
<td>-0.0431 (0.068)</td>
<td>-0.023 (0.074)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IHD</td>
<td>-0.060 (0.078)</td>
<td>0.064 (0.076)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCT</td>
<td>0.092 (0.074)</td>
<td>0.054 (0.083)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All three</td>
<td>0.046 (0.086)</td>
<td>0.078 (0.078)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Average standard error in parenthesis.
Figure 13: On and Off-Peak Treatment Effects, TOU
Figure 14: On and Off-Peak Treatment Effects, VPP
the median household consumes more or less electricity than in the control condition. In all technology treatments for TOU we find at least 50% of households significantly reduce peak consumption for the Portal and IHD, and these numbers are over 80% for the PCT alone. We also find that more than 50% of households reduce consumption at statistically significant levels for all technologies using VPP pricing with the exception of the IHD. It is clear from visual inspection that all treatments decrease usage during critical and non-critical peak hours.

5.2.3 Can demographics explain the heterogeneity?

By using rich pre-treatment data, we can estimate household treatment effects in order to best target them. To show the advantage of this methodology over using just demographic variables, in Figure 15 we plot separate CDFs for the different demographic bins for the effect of the PCT on critical consumption, for each of the different demographic variables. There does appear to be some evidence that family households are more affected by the PCT relative to the portal, with little difference between young and old households, and that the effectiveness of the PCT relative to portal increases with income. Nevertheless, through all of these plots, it is clear that demographic differences are small and there is substantial heterogeneity, conditional on demographics, explained by the past usage data used to create the household level estimates.

5.2.4 Who do we target?

Because we are interested in which technology to use when targeting specific households when giving them a smartmeter, in Table 5 we show counts of the sign and significance of the relative impacts of the IHD, PCT and All 3 treatments relative to the portal treatment. For the PCT and All 3 technology treatments, we see that the estimated increase in reduction of usage during critical and peak periods is significant in over half of households in

---

2It should be noted that these effects are in absolute changes in consumption - families and higher income households also consumer more electricity.
Figure 15: Difference between Peak and Off-Peak Treatment Effects
both pricing conditions. The IHD is significantly more effective than just the portal in 636 homes under TOU pricing, but it is only significantly more effective under VPP pricing for 10 households. IHD with VPP is less effective at reducing consumption for 287 households—it could be that the increased accessibility of information coupled with the VPP pricing actually encouraged consumers to use more electricity at the lower price point which occurred frequently in the observation window.

A substantial number of households exhibit significant increases in usage in the off-peak hours when using the programmable thermostats relative to the portal—970 under TOU pricing. We also see that the All 3 conditions leads to fewer households significantly reducing consumption during critical and peak periods or increasing consumption during off-peak periods, relative to the portal. Therefore the portal treatment appears to dominate the All 3 condition. However, despite its effectiveness on average, when targeting a household with the PCT, we have to recognize that the approximate cost of supplying a PCT is $250 and so the cost effectiveness of supplying consumers with the PCT will depend on the size of the expected critical reduction in usage relative to an estimate of the resultant deferred costs to the utility.

5.2.5 Cost effectiveness

In what follows, we focus on the portal vs. PCT technologies since the PCT dominates the other treatments in terms of peak load reduction. In Figure 16 we show scatter plots of the PCT versus portal coefficients for the two price treatments. These results have clear implications for firm behavior. In Figure 16, we also show the break-even line for the utility which represents the additional 0.357 kW reduction of the PCT over the portal. Any household below the line should receive the PCT. There is a clear incentive to target the PCT. Doing so avoids the $250 installation cost in those homes that would not reduce demand by the break-even amount. This increases the utilities welfare by 29% under VPP and 20% under TOU. The average of the household treatment effects reported in Table 4 implies a
Table 5: Significance in Household Differences between Non-Portal Technologies and Portal

<table>
<thead>
<tr>
<th></th>
<th>Critical</th>
<th>TOU</th>
<th>VPP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IHD</td>
<td>PCT</td>
<td>All 3</td>
</tr>
<tr>
<td>Sig. Negative</td>
<td>636</td>
<td>1436</td>
<td>1104</td>
</tr>
<tr>
<td>Negative</td>
<td>729</td>
<td>209</td>
<td>450</td>
</tr>
<tr>
<td>Positive</td>
<td>408</td>
<td>102</td>
<td>204</td>
</tr>
<tr>
<td>Sig. Positive</td>
<td>25</td>
<td>51</td>
<td>40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Peak</th>
<th>TOU</th>
<th>VPP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IHD</td>
<td>PCT</td>
<td>All 3</td>
</tr>
<tr>
<td>Sig. Negative</td>
<td>717</td>
<td>1447</td>
<td>1105</td>
</tr>
<tr>
<td>Negative</td>
<td>574</td>
<td>238</td>
<td>435</td>
</tr>
<tr>
<td>Positive</td>
<td>455</td>
<td>84</td>
<td>225</td>
</tr>
<tr>
<td>Sig. Positive</td>
<td>52</td>
<td>29</td>
<td>33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Off-Peak</th>
<th>TOU</th>
<th>VPP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IHD</td>
<td>PCT</td>
<td>All 3</td>
</tr>
<tr>
<td>Sig. Negative</td>
<td>81</td>
<td>17</td>
<td>11</td>
</tr>
<tr>
<td>Negative</td>
<td>1107</td>
<td>90</td>
<td>146</td>
</tr>
<tr>
<td>Positive</td>
<td>541</td>
<td>721</td>
<td>1211</td>
</tr>
<tr>
<td>Sig. Positive</td>
<td>69</td>
<td>970</td>
<td>430</td>
</tr>
</tbody>
</table>
surplus gain of $247 and $323 over the $250 cost for VPP and TOU respectively. However, the targeting of PCTs to those households with sufficient demand reduction increases the surplus gains to $319 and $386 respectively.

6 Conclusion

We have demonstrated that home automation technology can create surplus increases on both the consumer and firm side. The firm side surplus is substantially larger suggesting they should be willing to subsidize the installation of technology. We also illustrate how large historical databases of behavior can be used to identify individual-specific treatment effects. In the context of our electricity demand experiment, we show that the utility can increase its surplus by selectively subsidizing programmable communicating thermostats to a subset of households. Our approach is designed to work in instances where the historical data can be treated as a series of repeated draws from a process. While electricity usage conditional on weather and other states fits this description, there are many more potential applications. Technological innovations in communicating and/or wearable technology are enabling the collection of even more extensive behavioral data. By integrating these data into experiments, as we have done here, there is hope to extend varying treatment prescriptions beyond simple demographics which, in many applications such as marketing, explain very little heterogeneity across individuals.

References


Figure 16: PCT and Portal Peak Treatment Effects


