NBER WORKING PAPER SERIES

BREAKING ROUTINE FOR ENERGY SAVINGS:
AN APPLIANCE-LEVEL ANALYSIS OF SMALL BUSINESS BEHAVIOR UNDER DYNAMIC PRICES

Jiyong Eom
Frank A. Wolak

Working Paper 27263
http://www.nber.org/papers/w27263

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2020

This work was supported by the Korea Electric Power Corporation (KEPCO) [grant number R16DA23]. We thank colleagues from the Korea Electric Power Research Institute (KEPRI), in particular, Dr. Dongsik Jang, for collaborating in designing and implementing the Smart Save Days Campaign and sharing data that made this research possible. We also thank participants at the Energy Institute at Haas (UC Berkeley) Summer Camp 2019 for helpful comments on a previous draft. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Breaking Routine for Energy Savings: An Appliance-level Analysis of Small Business Behavior under Dynamic Prices
Jiyong Eom and Frank A. Wolak
NBER Working Paper No. 27263
May 2020
JEL No. Q4,Q41

ABSTRACT

Small businesses are typically committed to providing a positive customer experience and therefore may exhibit a response to dynamic electricity prices different from residential or industrial customers. We conduct a field experiment to determine the extent to which small businesses respond through re-configuration of typical routines throughout the experiment period versus through adjustments to specific dynamic pricing events. Using a customer-level survey of appliance ownership, we estimate the hourly response patterns of individual appliances to participation in the experiment versus individual dynamic pricing events. Consistent with our re-configuration hypothesis, small businesses primarily curtail electricity usage throughout the experiment period, although we also find a small imprecisely estimated response to dynamic pricing events on top of the re-configuration effect. Appliances not critical to a positive customer experience such as dish dryers, food storage units, lights, electric motors & pumps, and industrial heaters are the major sources of the energy savings from the re-configuration actions of these small businesses.

Jiyong Eom
College of Business
KAIST
85 Hoegi-ro Dongdaemun-gu
Seoul
Republic of Korea
eomjiyong@kaist.ac.kr

Frank A. Wolak
Department of Economics
Stanford University
Stanford, CA 94305-6072
and NBER
wolak@zia.stanford.edu

"If the shop is hot, the guests simply leave, developing a bad reputation. Maintaining a pleasant environment is important for us."

– A barber shop owner

"It didn't take long before I got into energy saving lifestyle. I pulled out one or two light bulbs at my store during the campaign."

– A keysmith store owner

"Whenever out for home visiting services, I turned off nearly all appliances not in use and pulled their power cords. I didn't do any of these before joining the campaign"

– A small hardware store owner

1. Introduction

There is a widespread agreement that the efficiency and reliability of the electricity supply industries could be improved by making final demand more responsive to the hourly price of wholesale electricity. Although the economic literature has extensively investigated the responsiveness of residential electricity consumers to dynamic prices in various experimental settings, little progress has been made to understand how small commercial consumers or, equivalently, small businesses might respond to dynamic prices.[1] We report on the results of a dynamic pricing experiment involving small businesses in order to understand how their price responsiveness might be similar to and different from that of residential consumers. We also make use of detailed customer-level appliance surveys to unpack the hour-of-day level aggregate price responsiveness into a set of hour-of-day appliance-level price responses.

Consistent with the above quotations from participants in our experiment, we hypothesize that small businesses are committed to giving every customer a positive service experience through an established set of routines and this may cause them to respond different from residential consumers to facing dynamic prices. Responding to dynamic price signals typically only costs the household a temporary inconvenience, but the same temporary demand response could cost a small business both a temporary and permanent loss of customers and impose additional burdens on employees. This difference can create a different set of incentives for the small business owners in reacting to dynamic prices, which to our knowledge, has not been rigorously investigated.

We conjecture that the small business are likely to re-configure how they do business to be able to reduce their demand during dynamic pricing events periods, rather reduce demand through one-off

[1]For experimental evidence on the impact of dynamic pricing for residential consumers see Wolak (2007), Wolak (2010), Jessoe and Rapson (2014), Ito et al. (2018) and Burkhardt et al. (2019). For the case of industrial and large commercial customers see Patrick and Wolak (2002).

routine-breaking efforts during these time periods which could cause a temporary or permanent loss of revenues and impose increased burdens on employees. The goal of our study is to investigate the extent to which small businesses employ each of these two modes of demand response at the appliance level using hourly customer-level consumption data and detailed inventories of customer-level appliance holdings for our treatment and control groups.

We find that the primary mode of price-responsiveness by small businesses appears to be through re-configuration of their use of appliances after entering the dynamic pricing program. Customers change their use of appliances to reduce their electricity consumption during peak hours of day, whether or not there is a dynamic price incentive to do so. Electricity usage in peak hours (1-5 pm) decreased during our experiment period by, on average, about 8% (0.1 kWh per hour). This dynamic pricing *campaign effect* demand reduction consisted primarily of a non-dynamic response by the small business of making it a regular practice to curtail usage of non-customer-service critical appliances during peak hours on all days of the business participated in the dynamic pricing campaign.

The significantly higher price charged for electricity during peak hours on dynamic pricing event days and the service quality constraints and employee routine constraints on the ability of small businesses to reduce demand on dynamic pricing event days argue in favor of a *campaign effect* response to participating in a dynamic pricing experiment. This perspective is supported by our findings that the estimated mean bill savings during the entire two-month experiment period from this campaign effect added up only to $US 3.80. However, much higher price paid by the small businesses during the event hours on the ten event days during the campaign period returned about the same estimated mean additional savings (or, equivalently, rebates) of $US 3.70.

To capture the ten event days' worth of bill savings, these small businesses re-configured their operations for all peak hours throughout the experiment period. Re-configuring their operations throughout the entire dynamic pricing campaign allowed small businesses to capture these bill savings during the ten events days without risking the loss in revenues from decline in service quality or the cost of changing employee routines from a temporary reduction in energy demand. The estimated demand reduction during all peak hours during the campaign period is a non-trivial economic and environmental benefit from facing small business customers with dynamic prices. The customer-level mean energy savings from the event days (8.9 kWh) is dwarfed by total energy savings induced by the dynamic pricing campaign (32.9 kWh).

The second key finding from our experiment is that we are unable to detect any statistically meaningful incremental aggregate or appliance-level consumption change from individual dynamic pricing events beyond the ongoing demand reduction due to the *campaign effect*. This outcome is consistent with our *difficulty-of-breaking-the-routine* hypothesis for small businesses. The finding that the firms did not undertake any further curtailment of uses of their appliances on event days is likely the result of the high fixed, irreversible cost of setting up a new service routine for the small business and the likelihood that a temporary demand reduction would lose the small business more revenues than it would save on its electricity bill.

Additional evidence in favor of our *difficulty-of-breaking-the-routine* hypothesis is the breakdown of the total within-day *campaign effect* into response patterns of individual appliances belonging to the business. We extend the statistical procedure of estimating the baseline patterns of residential end-use demand, referred to as conditional demand analysis by Aigner et al. (1984) and Bartels and Fiebig (2000), allowing for weather-dependent baseline patterns of appliance use within the day and changes in these within-day patterns of consumption during the dynamic pricing campaign period and dynamic pricing event days. We estimate both unrestricted within day response patterns, as well as response patterns that follow high order polynomials in the hour of the day, to address the issue of the high dimension of our unrestricted parameter space.

Our appliance-level analysis consistently reveals that the direction, magnitude, and timing of within-day demand responses are very different across the appliances, making a convincing case for the re-configuration undertaken at the appliance-use level. Persistent peak hour consumption reduction was observed only for non-service critical appliances, such dish dryers, food storage units (refrigerator and Kimchi fridge), lights (incandescent, fluorescent, and LED), motors & pumps, and industrial heaters, with limited shifting of these appliance uses to off-peak hours of the day.

Our findings collectively suggest that although small businesses have little flexibility manually in responding to hourly electricity price signals, this does not mean that exposing them to dynamic prices will not lead to a demand response. Consistent with post-experiment comments from participants given in the Appendix A7, the opportunity to capture significant bill savings from rebate earned during dynamic pricing events led these small businesses to undertake a re-configuration of service routines during the entire dynamic pricing campaign. Our results also point to a critical role for automated demand response and timely information feedback in alleviating customer discomfort and limiting inefficiencies in business practices that can allow small business to respond to dynamic price signals.

The remainder of the paper is organized as follows. Section 2 describes the unique challenges facing small business customers in responding to dynamic prices and uses that as a basis for making our *difficulty-of-breaking-the-routine* demand response hypothesis. Section 3 describes the design of our field experiment and data collection process. Section 4 develops the econometric models used to estimate appliance-level weather-dependent within-day baseline consumption patterns, within-day pattern of consumption responses to the dynamic pricing campaign, and within-day pattern of consumption responses to a dynamic pricing event day. Section 5 discusses policy implications of our experimental results, and Section 6 concludes.

2. The Economics of Small Business Demand Response

Electricity consumers can be expected to evaluate the benefit versus cost of any potential demand response action in order to determine whether to undertake it. The benefits of such an action are primarily the savings in the consumer's monthly electricity bill and any additional financial compensation granted for that action. However, small business consumers face costs that are not relevant to residential consumers. Members of the household are typically the only individuals that experience the financial or inconvenience cost associated with a demand response action. For small business, many demand response actions also impose costs of both customers and employees. For example, pre-cooling a store or restaurant to reduce air conditioning use during the hottest hours of day, imposes costs on customers and employees both in terms of cooler than desired temperatures early in the day and hotter than desired temperatures later in the day.

Because financial viability is typically a necessary condition for continued existence of a small business, demand response actions with expected financial benefits greater than the expected costs are the only ones likely to be undertaken.[2] For the small businesses in our sample, the expected bill savings from responding to a dynamic pricing event is likely to be less than the cost of taking this action. For example, a customer in our sample consumes 5.6 kWh during a four-hour-long peak period (our sample average consumption during this time period) would gain only $US 0.44 from a one-time estimated reduction of 0.45 kWh (8%) during the four-hour-long peak-time-rebate dynamic pricing event (through current period bill savings and rebate payments). This level of bill savings is unlikely compensate for the likely negative impact on the daily average $US 220 sales revenue of the small business in our sample that reported sales revenue.

For this reason, we hypothesize that small businesses participating in a dynamic pricing experiment take an alternative approach to profiting from their participation. These small businesses take one-time re-configuration actions at the start of the dynamic pricing experiment that impose little, if any, costs on customers and employees in order to capture ongoing savings from reduced consumption

[2] This is different from residential consumers where the desire to feel environmentally conscious may tip the scales in favor of undertaking a demand response action.

during all peak hours of the experiment period *and* the additional savings during all dynamic pricing events that do require them to break this new routine. The roughly similar magnitude of average bill savings from the ten dynamic pricing events as the average bill savings these small business received for their re-configuration actions throughout the entire dynamic pricing campaign period also explains why the small businesses did not undertake these actions before participating in the dynamic pricing experiment.

Our experiment examines the extent to which small businesses undertake a *re-configuration* of electricity usage during the entire dynamic pricing campaign, which we refer to as the *campaign effect* versus respond to individual dynamic pricing events, which we call the *marginal event-day effect*. We define the *full event-day effect* as the sum of the *campaign effect* and *marginal event-day* effect. Our field experiment relies on a peak-time rebate (PTR) mechanism as the treatment that provided small businesses with the economic incentive to curtail electricity consumption during dynamic pricing events. Besides estimating a weather-dependent baseline within-day electricity consumption pattern for each appliance category, we also estimate within-day *campaign effects* and within-day *marginal event-day effects* for each appliance category.

A possible alternative interpretation of our *campaign effect* result might be that the treatment subjects altered their electricity consumption because they were aware of being observed due to the so-called Hawthorne effect (Monahan and Fisher, 2010). Our experiment controlled for this by alerting both the treated and control subjects that their electricity consumption would be monitored and used for research purposes and compensated both groups for this with an upfront financial payment. Therefore, we believe that our *campaign effect* is likely to be result of the expected bill savings from peak time rebates during dynamic price events for the treated group relative to the control group.

Further evidence for our *re-configuration* hypothesis is the specific changes in appliance use that make up our *campaign effect*. Because we administered an on-site survey to all small businesses in our treatment and control groups of their ownership of an exhaustive list of appliances, this information allows us to decompose both types of customer-level within-day demand response activities into changes in within-day use of each appliance the customer owns. We found *campaign effect* peak hour electricity consumption reductions for non-service critical appliances, such dish dryers, food storage (refrigerator and Kimchi fridge), light (incandescent, fluorescent, and LED), motor & pump, and industrial heater, with limited shifting of these appliance uses to off-peak hours of the day.

An appliance-level understanding of the choice and timing of demand response measures could have a number of follow-on benefits. Energy service providers, for example, can use this information to design plug-level appliance curtailment or load-shifting contracts or to predict the potential reconfiguration response of individual program participants based on their appliance holdings.

3. Research Design and Data Preparation

We partnered with the state-owned Korea Electric Power Corporation (KEPCO) to design and implement a field experiment that introduced a peak-time rebate (PTR) program called "Smart Save Days Campaign" to a sample of small commercial and industrial (C&I) electricity customers in Seoul. These C&I customers were mostly small commercial outlets and service businesses located across the city's 25 districts. We selected the initial list of target customers from a stratified random sample[3] from approximately 20,000 small C&I customers who had hourly interval meters as part of the nationwide advanced metering deployment plan to be achieved by 2020.[4] During sample recruitment, however, unanticipated changes were made to the original list of target subjects that would result in systematic difference between the recruited control and treatment groups which would keep the former from constituting an unbiased measure of baseline electricity usage for the latter and thereby from drawing causal inference. We thus attempted to resolve this issue after the experiment using the two approaches that produced quantitatively similar results. The problems associated with the original sample and recruitment, as well as methodologies we employed to address them are discussed in Appendix A1.

After obtaining on-site consent from the subjects for collecting their hourly electricity use for research purposes, the recruiters administered tablet-assisted personal interviews on various items including, business characteristics, operation hours, appliance holdings, and self-reported summer bills. Any incorrect information about business characteristics previously given by KEPCO was revised then. The treatment subjects were also given with a short on-site education on the PTR that they would be subject to in a couple of weeks. This educational experience was assisted by the leaflets shown in Fig. A1, which describe the duration and operation of the campaign and possible measures to reduce peak-time electricity usage on event days. It was not until both the treatment and control group subjects completed the on-site consent and personal interviews that they were paid the fixed participation incentive of $US 26 and allowed to be part of the experiment.

Upon completion of the recruitment process by mid-July of 2017, the experiment started in early

[3] The stratified random assignment was undertaken by first assigning 29 standard business classes to one of seven strata established based on their electricity usage in August and September of 2016, followed by random sampling of control and treatment subject pairs from the individual strata. The procedure is explained in Appendix A1.

[4] As KEPCO had been responsible for sequencing the rollout of interval meters to different service sites and C&I customers equipped with the meters continued to remain under seasonal flat tariffs like other customers, there is no particular reason to believe that our target customers were systematically different from the entire C&I population.

August and finished at the end of September 2017, broadly covering the latter part of the hot summer season in Korea. The "Smart Save Days Campaign" offered the treatment subjects a rebate of $US 0.90 per kWh reduced during peak hours (1-5 pm) on PTR event days relative to their counterfactual baseline.[5] Because both the treatment and control groups were interviewed and paid the fixed participation incentive of $26, and only the treatment group was allowed to earn the rebate, the Hawthorne effect should not be relevant to our treatment effect.

Throughout the experiment period, a total of 10 event days were declared for the peak hours (1-5 pm) for the treatment subjects. Each event notice was sent at 3 pm day before via text messages. There were about 30 subjects that dropped out of the experiment because of inconveniences experienced or a business relocation that was not anticipated at the time of recruitment, as well as nearly 90 subjects that later had to be excluded due to malfunction of their interval meters. In response to the ongoing attrition, we undertook the parallel recruitment of additional 120 control subjects, eventually securing a total of 1,517 sample subjects—902 for the treatment and 615 for the control group.

After the experiment period, we also attempted to interview all treatment subjects on their frequency and means of undertaking demand response and the self-reported frequency of checking text messages. We conducted two sets of post-experiment interviews, each with seven individuals from the treatment group based on their demographic profile, business type, and level of electricity consumption. During the interviews, we asked a series of questions on the participant's experiences during the campaign. Selected results of the post-campaign survey and interviews are summarized in Appendix A7.

Because of the above mentioned difference between the original sample assignment and the finally enrolled participants, as well as the sample attrition during the experiment, we attempted improve the integrity of randomization *ex post* by performing a random re-assignment from the eventually secured 1,517 subjects now with correct business information.[6] The re-assignment was done based on the sampling frequencies used at the time of designing our experiment, with which we believe the finally secured subjects should have been stratified. This procedure is described in detail in Appendix A1. This procedure returned a total of 800 subjects—400 for the treatment and 400 for the control group—which would be representative of small C&I customer population in Seoul. We

[5] The determination of customer's baseline relative to which rebates were issues takes the average peak-time electricity consumption from the four highest consumption days of the last five non-event business days.

[6] The two experimental groups before the re-assignment exhibited statistically significant differences at the p <0.01 level in terms of average hourly electricity consumption in the pre-campaign period. Specifically, the treatment subjects had a 10% higher average hourly consumption than the control subjects.

use this sample for our empirical analysis that follows. As a robustness check, we also performed a coarsened exact matching to create statistically equivalent groups of treatment and control subjects (Iacus et al., 2011, 2012). The details of this procedure is also explained in Appendix A1. The appliance-level demand response effects estimated from the matched treatment-control pairs and key insights thereof remained almost the same as those from our sample re-assignment, so that we do not report the these results separately.

Table 1 presents descriptive statistics of pre-treatment observables and surveyed characteristics of the two re-assigned experimental groups and their differences. A comparison across the treatment and control groups confirms statistical balance in pre-treatment observables collected from interval meters, such as daily, peak usage, and average off-peak usage. The sample balance in terms of average daily usage, peak usage, and off-peak usage in the pre-experimental period is supported even more strongly by the two-sample Kolmogorov-Smirnov test (Smirnov, 1939), which is shown in Fig. 1. Table 1 also indicates that self-stated measures for electricity use behavior in the summer (summer bill and set temperature), measures for the scale and pattern of business operation (floor area, operation hours, and number of employees), and proxies for overall thermal insulation level and motivation for energy efficiency investments (years built and store ownership) are all similar between the two groups. The small number of estimates differences between the means in treatment and control columns in Table 1 that are substantially larger than their standard error is not inconsistent with the validity of the null hypothesis that the treatment and control groups have the same population mean vectors.[7]

4. Estimation Procedure and Results

Three key features of our data need to be taken into account in order to estimate the baseline, campaign and event day appliance-level electricity demand response to our peak-time rebate (PTR) program. First, the model should be able to leverage both individual-level electricity consumption data and individual-level appliance ownership information.[8]

Second, the model should be capable of identifying hourly PTR treatment effects for individual appliances while controlling for each appliance's baseline weather-dependent electricity consumption. We do not believe it is reasonable to estimate a customer's demand response based on the assump-

[7] A $\alpha = 0.05$ test of a valid null hypothesis rejects with 0.05 probability.

[8] The appliance ownership information received 100% survey compliance as it was in essential part of the survey. However, whether or not all of the respondents correctly filled out the checklist of appliance ownership might be called into question. For example, those who run system heaters might have claimed air conditioning (AC) ownership because many of the system heater models have AC functionalities. Also, those who own drink showcases might have checked the kimchi fridge category instead of fridge because the showcases were used to store kimchi.

tion that one appliance category, such as air conditioners, shares the same sensitivity of baseline electricity consumption to weather disturbances as the others, such as television sets. Our baseline specification should allow for a different weather-dependent baseline within-day hourly consumption for each appliance.

The third modeling consideration is specifying a model that is capable of recovering within-day hourly appliance-level treatment effects separately for the *campaign* period and *event days*. Fig. 2 compares the within-day distribution of electricity usage of the treatment and control groups for three different time frames relevant to our analysis: 71 weekdays before the campaign, 27 non-event weekdays during the campaign, and 10 event weekdays during the campaign. It presents the median and inter-quartile range (IQR) of hourly usage distribution for each time frame with solid lines and shaded areas around them, respectively. Here, the campaign period is defined as the time frame during which the treatment subjects were made aware that PTR event could be called at anytime, which ranges from Aug. 8—the day campaign started *and* the first day-ahead event alert was sent—through Sept. 27—the day after the last event *and* when the completion of the campaign was announced. Fig. 2 indicates, first, that the distribution of electricity usage is right-skewed (the shaded area above the hourly medians much larger than the shaded area below them), which makes its log-transformation of hourly consumption a suitable specification for the dependent variable. Second, as anticipated, the treatment and control groups are well balanced in terms of hourly medians of electricity usage during the pre-campaign period. Once the campaign began the treatment group seems to curtail usage in and around the hours for which PTR events were called (1-5 pm). Interestingly, a similar curtailment response is found also on non-event days, which supports our *re-configuration* interpretation.

4.1. The Main (Unrestricted) Model

We present a model to quantify appliance-level hourly treatment effects that the campaign triggers on an ongoing basis, as well as marginal effects that dynamic price events produce. The model specification is given by:

$$\begin{aligned} Y_{itd} &= \sum_{j=1}^{J}\sum_{h=1}^{24}\left[\beta_{jh}^{base} + \mathbf{\Omega}_{td}\boldsymbol{\beta}_{jh}^{weather}\right] D_i^j H_t^h \\ &+ \sum_{j=1}^{J}\sum_{h=1}^{24}\left[\beta_{jh}^{cmpgn} D_i^j H_t^h I_{id}^{cmpgn} + \beta_{jh}^{event} D_i^j H_t^h I_{id}^{event}\right] + \gamma_i + \epsilon_{itd} \end{aligned} \tag{1}$$

$\mathbf{\Omega}_{td}$ weather variables matrix, consisting of columns for cooling degree hours (CDHs), humidity, and their interactions in hour t on day d

γ_i	individual fixed effect for subject i
D_i^j	indicator of 1 if subject i owns appliance j during the sample period, and 0 if otherwise
H_t^h	indicator of 1 if hour t equals a given hour of the day, $h \in \{1, 2, \cdots, 24\}$, and 0 if otherwise
I_{id}^{cmpgn}	indicator of 1 if subject i is assigned to the treatment group *and* consuming on day d on which the campaign is in progress (from Aug. 8 through Sep. 27), and 0 if otherwise
I_{id}^{event}	indicator of 1 if subject i is assigned to the treatment group *and* consuming on day d for which PTR event is called during the campaign, and 0 if otherwise[9]
Y_{itd}	logarithm of electricity usage of subject i in hour t on day d of the entire sample period (from Jul. 1 through Sept. 27)

Besides the usual individual fixed effect, γ_i, which controls for time-independent heterogeneity in consumption patterns across the sample, our model consists of two major parts. The first part represents counterfactual hourly baseline consumption for each appliance as predicted by the subject's ownership of appliance j in hour h, $D_i^j H_t^h$, as well as the appliance's unique hourly sensitivity to weather conditions Ω_{td}. This flexible baseline specification allows for different hourly impacts that a given change in weather can bring to the employment of the individual appliances. For example, a high temperature day would imply a higher baseline usage throughout the day of AC and other cooling appliances. The second part of the model captures the average treatment effects at the appliance level, which are the main subject of our analysis. To distinguish the ongoing campaign effect from the marginal event-day effect that the dynamic price events provide, hourly appliance ownership variable, $D_i^j H_t^h$, is interacted with two treatment indicators, that is, campaign treatment, I_{id}^{cmpgn}, and marginal event-day treatment, I_{id}^{event}.

The coefficient of interest, β_{jh}^{ϕ}, represents the individual effects of contribution type $\phi \in \{base, weather, cmpgn, event\}$ by appliance j in hour h. In our case, this specification presents a total of 3,517 variables for about 1.2 million observations—2,717 predictors (the number of appliance $\times$ hours $\times$ contribution types) and 800 individual fixed effects. Given the computational challenge posed by the large number of covariates, we employ the two-step estimation algorithm utilizing the Frisch-Waugh-Lovell Theorem of Somaini and Wolak (2016). The first step is to recover residuals from the projection of Y_{itd},

[9] The superscript '*event*' stands for the presence of marginal treatment effect provided by PTR events on top of the running effect of the campaign itself.

$D_i^j H_t^h$, $\Omega_{mtd} D_i^j H_t^h$, $D_i^j H_t^h I_{id}^{cmpgn}$, $D_i^j H_t^h I_{id}^{event}$ on fixed effects for every j and h:

$$\begin{aligned}
Y_{itd} &= \gamma_i + e_{itd}^{Y} \\
D_i^j H_t^h &= \gamma_i + e_{jh,itd}^{base} \\
\Omega_{mtd} D_i^j H_t^h &= \gamma_i + e_{jh,itd}^{\chi} \\
D_i^j H_t^h I_{id}^{cmpgn} &= \gamma_i + e_{jh,itd}^{cmpgn} \\
D_i^j H_t^h I_{id}^{event} &= \gamma_i + e_{jh,itd}^{event} \\
& \quad \text{for all } j,\ h,\ \text{and } m
\end{aligned}$$

where Ω_{mtd} is the m-th vector element of weather variables matrix $\mathbf{\Omega}_{td}$. The second step is to estimate the coefficients of interest based on the set of residuals computed as described above, which is given by

$$\widehat{e_{itd}^{Y}} = \sum_{j=1}^{J} \sum_{h=1}^{24} \left[\beta_{jh}^{base} \widehat{e_{jh,itd}^{base}} + \sum_{m} \beta_{jh}^{m} \widehat{e_{jh,itd}^{m}} + \beta_{jh}^{cmpgn} \widehat{e_{jh,itd}^{cmpgn}} + \beta_{jh}^{event} \widehat{e_{jh,itd}^{event}} \right] + \epsilon_{itd}. \tag{2}$$

Now with much reduced computational costs, this process allows us to estimate β_{jh}^{ϕ} and their variances, $\text{var}(\beta_{jh}^{\phi})$, for every j, h, and ϕ. Having identified the model, we compute the change in the average kWh amount of electricity usage of appliance j in hour h during the campaign, Δ_{jh}^{cmpgn}, the marginal change that dynamic price events additionally introduces on event days, Δ_{jh}^{event}, and the resulting full treatment effect on event days, Δ_{jh}^{full}, by

$$\begin{aligned}
\Delta_{jh}^{cmpgn} &= \left[1 - \exp(-\widehat{\beta_{jh}^{cmpgn}})\right] \bar{Y}_h \\
\Delta_{jh}^{event} &= \left[1 - \exp(-\widehat{\beta_{jh}^{event}})\right] \bar{Y}_h \\
\Delta_{jh}^{full} &= \left[1 - \exp(-\widehat{\beta_{jh}^{cmpgn}} - \widehat{\beta_{jh}^{event}})\right] \bar{Y}_h
\end{aligned}$$

where $\bar{Y}_h$ is the average hourly electricity consumption of the treatment subjects in hour h during the campaign. The confidence intervals of these treatment effects are obtained from parametric bootstrapping with all standard errors of the coefficients clustered at the level of individual *and* day-of-the-sample. Also, the weather-responsive hourly baseline electricity consumption for appliance j in hour h is given by

$$B_{jh} = \exp\left(\widehat{\beta_{jh}^{base}} + \sum_{m} \widehat{\beta_{jh}^{m}} \bar{\Omega}_{mtd}\right).$$

where $\bar{\Omega}_{mtd}$ is the assumed hourly vector of the m-th element of weather variables matrix Ω_{td}. To illustrate the importance of weather on hourly baseline consumption, we choose three sets of weather scenarios: (i) highest hourly cooling degree hours *and* highest hourly humidity, (ii) mean hourly cooling degree hours *and* mean hourly humidity, and (iii) lowest hourly cooling degree hours *and* lowest hourly humidity all during the sample period. Fig. 3 plots hourly baseline electricity consumption in the three weather scenarios for each of the 19 appliances categories surveyed in the on-site personal interviews. The names of the appliances and their abbreviations are listed in Table 3.

Several findings emerge from these results. First, some appliances exhibit much higher sensitivity to temperature and humidity than the others, which justifies our assumption of different weather-dependent baselines for each appliance category. Appliances for space cooling (AC and FN), lighting (FL and LD), and copy machine (CM) are the examples. Not only the first but also the following two services seem to be employed more intensively on hotter and more humid days because such weather is likely to bring more customers into air-conditioned stores. By contrast, dish dryers (DD) and motors & pumps (MP) are used less on hotter and more humid days presumably to reduce internal heat gains from the appliances during service hours. Second, most appliances are used more intensively in typical business hours than in the other hours, which serves our intuition. Such appliances include AC, FN, WM, FL, LD, CM, MP, IH, and MI. Third, several appliances are used in a relatively large proportion during non-business hours, that is, in the morning and the late afternoon. The examples include DD, RC, and OR, which are run or loaded in preparing for or closing daily business operations. Such behavior also makes sense because our baseline consumption is for the summer season, in which internal heat gains from running the kitchen appliances during daytime would cause significant discomfort. Fourth, appliances that require some warming-up lead time with high wattage power, such as OR and MP, tend to be used more intensively in the morning, so that the small businesses can serve their customers for the rest of the day. Fifth, interior or exterior lighting devices, such as IN, FL, and LD, exhibit steadily increasing electricity usage from around noon through the night.

Fig. 4 displays the hourly estimates of the appliance-level treatment effects, separately for the campaign effect and the marginal event-day effect and their pointwise 95% confidence intervals indicated in shaded error bands. As anticipated, the hourly coefficient estimates are less precise during business hours than the other hours, in which the small businesses have little discretion in the choice and use of appliances; and the estimates for event days are less precise than those for the encompassing campaign days, in part due to the number of observations being smaller in the former than in the latter. To put the significance of these hourly treatment effects into another perspective, we performed simultaneous exclusion tests for the peak-hour coefficients on the null hypothesis, $\beta^{\phi}_{j14} = \beta^{\phi}_{j15} = \beta^{\phi}_{j16} = \beta^{\phi}_{j17} = 0$, as well as for its off-peak hour counterparts, $\beta^{\phi}_{j1} = \beta^{\phi}_{j2} = \cdots = \beta^{\phi}_{j13} = \beta^{\phi}_{j18} = \cdots = \beta^{\phi}_{j24} = 0$, separately for the campaign effect ($\phi = cmpgn$) and the marginal event-day effect ($\phi = event$). We also performed the simultaneous exclusion tests for the full event-day treatment effect on the hypotheses, $\beta_{j14}^{cmpgn} + \beta_{j14}^{event} = \cdots = \beta_{j17}^{cmpgn} + \beta_{j17}^{event} = 0$ for peak hours and $\beta_{j1}^{cmpgn} + \beta_{j1}^{event} = \cdots = \beta_{j13}^{cmpgn} + \beta_{j13}^{event} = \beta_{j18}^{cmpgn} + \beta_{j18}^{event} = \cdots = \beta_{j24}^{cmpgn} + \beta_{j24}^{event} = 0$ for off-peak hours. Here we allow for a cluster-robust inference from the tests by employing the Wald statistic, $W = (\mathbf{R}_j^{\phi}\widehat{\boldsymbol{\beta}})'(\mathbf{R}_j^{\phi}\widehat{\mathbf{V}}_{\boldsymbol{\beta}}\mathbf{R}_j'^{\phi})^{-1}(\mathbf{R}_j^{\phi}\widehat{\boldsymbol{\beta}})$, where $\widehat{\boldsymbol{\beta}}$ is the consistent unrestricted estimator for the full set of the coefficients, $\widehat{\mathbf{V}}_{\boldsymbol{\beta}}$ is their covariance matrix estimated with error terms clustered by the-day-of-the-sample-of-the-individual, and $\mathbf{R}_j^{\phi}$ is the restriction matrix assumed for the coefficients of appliance j by contribution type ϕ. These statistics results are shown in Table 4.

Four key findings emerge from the appliance-level hourly demand response patterns (Fig. 4). First of all, many of the appliances have precisely estimated *campaign effect* demand responses, but none of them exhibit *marginal event-day* response that is statistically different from zero.[10] Our results suggest that the treatment subjects responded by undertaking a non-dynamic response in the form of a one-time investment early on in a new set of service routines, instead a dynamic response whenever an PTR event day was called. Such re-configuration may take various forms, including saving previously wasted appliance uses, curtailing usage of non-service critical appliances, and turning them on later in the day or operating in shorter periods.[11] The minimal treatment differential on event days indicates that the treatment subjects did not introduce any significant measures besides what they had already implemented upon participating in the campaign. Although the ongoing non-dynamic response, which is not well synchronized with peak hours (1-5 pm), is compensated directly by bill savings, it would have reduced or even avoided the cost of paying immediate attention to the dynamic events when the curtailment is paid at a much higher rate. In this sense, the ongoing response made throughout the campaign can be considered as a fixed cost to deliver a sizeable curtailment on event days to come anytime soon. This finding is also supported by our test results in Table 4—no appliance rejects $\alpha = 0.05$ test of the joint exclusion hypothesis for the marginal event-day effects for both peak and off-peak hours.

[10] Largely similar aggregate results were found from a simple aggregate difference-in-difference model with the customer fixed effects and weather dependent baselines. Specifically, the campaign effect comprises most of the within-day curtailment with the marginal event-day effect not being statistically different from zero. The hourly treatment effects and hypothesis testing results are shown in Appendix A3.

[11] Our post-experiment focused group interviews of a subset of treatment participants also support this up-front re-configuration and continued use during the campaign. One participant with a golf shop stated "I did not check SMS messages except the first two. I just continued to save energy during the summer, while taking such behavior for granted. It was not until I joined the campaign before I got into the habit of pulling power cords." One participant running a hardware store said "The event notices were almost useless to me as I just continued to save electricity use since I joined the campaign. Whenever out for home visiting services, I turned off nearly all appliances not in use and pulled their power cords. I didn't do any of these before joining the campaign." One interviewee with a keysmith store added "It did not take long before I got used to the energy-saving lifestyle. I did not do as such before joining the campaign. To reduce electricity usage, for example, I pulled out one or two light bulbs at my store during the program." A herbal medicine shop owner stated "I recall that when I received event notices, I tended to turn on the air conditioner a little later in the morning of the event days. Or sometimes I just left the door open turning it off until customers entered in."

The second finding is that, all appliances that generate a precisely estimated campaign demand response also produce a precisely estimated *full* event-day demand response as well. As shown in Table 4, the simultaneous exclusion of the peak hour coefficients is rejected at the 10% significance level in both the campaign and full event-day cases for all appliances except MW, WM, and TV, and the same test of the off-peak hour coefficients is strongly rejected for all studied appliances. No appliance category fails to reject the exclusion of both peak and off-peak hour coefficients at the 1% significance level in the campaign or the full event-day treatment case. The robust full event-day demand response that is very generously compensated under the dynamic pricing PTR mechanism justifies why the non-dynamic response was undertaken during the campaign.[12]

Third, as far as the direction of demand response is concerned, desired (positive) peak-hour curtailment is observed only for a subset of appliances used for housekeeping (DD and CP), food storage (FS), illumination (IN, FL, LD), and machine drive and heating (MP and IH), most of which are not critical for customer service during summer business hours (Fig. 4). It seems that the difficulty of breaking service routines in response to PTR event notices made the businesses instead alter their regular operation of these appliances during the campaign. This re-configuration would have been relatively easy as the use of these appliances is a "you-schedule" service, and less prone to to lead to customer complaints or increased employee burdens than other appliance services.

However, the re-configuration argument worked in an unanticipated direction for several appliances, such as AC, OR, and VC, which exhibit a sharp curtailment in the morning followed by apparent energy-intensification in and around peak hours during the campaign (Fig. 4). Note that those appliances are mostly service-critical, delivering "customer-command" services, such as comfort, cleanness, and appetite, which customers care about the most. It seems that the treatment subjects attempted to limit the overall use of these appliances during the campaign without a compromise in customer comfort. One possible way is to adjust hours of operation (e.g., opening late or closing early) or to leave the appliances shut down in less-visited morning hours until significant customer visits occur in later working hours, in which their operation had to ramp up quickly to the full capacity.[13]

[12] The same test for the hourly coefficients estimated from a simple aggregate difference-in-difference model provides similar results. See Table A1

[13] In this regard, the field study conducted by Herter (2009) for 78 small commercial consumers in Sacramento shows that the small restaurants failed to precool and reduce AC service during event hours because their AC units were undersized. Our post-experiment focus group interviews confirmed that the participants indeed made a serious effort in limiting the use of appliances especially from morning hours during the campaign. For instance, one interviewee running a herbal medicine shop said "I recall that when I received day-ahead event notices, I tended to turn on the air conditioner a little later in the morning of the event days. Or sometimes I just left the door open turning the AC off until customers entered in." Importantly, the interviewees all agree that curtailing electricity usage during business hours was much more difficult than doing it in the morning hours before they opened for customers. One mobile phone

Last, significant within-day load shifting from peak to off-peak hours is hardly observed in the estimated *campaign effects*. One exception is WM, which presents a load shifting into early morning and late-night hours. Nevertheless, the paucity of evidence for load shifting suggests that, in most appliances, their peak-time reductions are less than offset by comparable usage increases in neighboring off-peak hours or on non-event days, even though the treatment subjects had a reasonably long time (22 hours) to undertake meaningful load shifting. Instead, in our case, peak-time curtailment entails usage reductions in neighboring hours before and after peak hours, for which usage curtailment is not supposed to be rewarded other than through usual bill savings. This preparatory and prolonged demand response behavior is significant for DD, FS, IN, FL, LD, CP, MP, IH, and MI, which are not critical components of customer experience in summer business hours. Limited load shifting and the longer-run curtailment behavior under dynamic prices are also reported in the residential setting by Jessoe and Rapson (2014), Allcott (2011), and Burkhardt et al. (2019).

Our finding of precisely estimated appliance-level *campaign effects* and statistically zero *marginal event-day* effects is robust to a number model specifications. We also tried additional control variables, such as day-of-sample fixed effects with and without hour-of-the-day fixed effects. These models produced essentially the same results. As another set of control variables, we also employed the business type *and* day-of-the-week fixed effects to capture common weekday-specific business operation patterns that vary with business type. A total of 15 standard business categories were used—wholesale, retail, restaurant & bar, repair service, retail & rental, personal service, organization, manufacturing, IT service, health service, engineering service, education service, construction, business support, art & sports. Again, the model presented very similar coefficient estimates. Thus, for the rest of the paper, we present results estimated from the form shown in Eq. (1).

4.2. The Polynomial Interpolation Model

While we have so far maintained the assumption that the baseline, demand and campaign and event day demand responses of a given appliance in one hour of the day is unrestricted relative to that in other hours of day, it is not unreasonable to expect that hourly the demand responses would occur smoothly between hours of the day. This could also be one reason why we do not find very precisely estimated *marginal event-day* effects. To reduce the number parameters we estimate and still allow for hour-of-day treatement effects, we consider a specification that allows for polynomial parameter

shop owner said "Although my store wasn't particularly sweltering, when visiting customers felt uncomfortable, that was time to turn on the air conditioner." One participant running a flour mill said "It was hard to curtail electricity usage because customers visited almost anytime. Then I had to start the machine right away without exception." Another one with a BBQ restaurant said "As our restaurant is open 24 hours, we leave the air conditioners on all day in the summer. During the campaign, however, I left one or two air conditioners turned off, and whenever customers requested, I simply turned them on."

dependency between the hourly coefficients of the baseline and treatment effects for each appliance. The approach implements the interpolated polynomial specification used to model distributed lag between investment appropriations and expenditures in Almon (1965). In our case, coefficient β^{ϕ}_{jh} for any ϕ and j is set to be related across hours of the day, such that

$$\beta^{\phi}_{jh} \equiv \sum_{k=1}^{K} b^{\phi}_{jk} h^k = b^{\phi}_{j1} h + b^{\phi}_{j2} h^2 + \cdots + b^{\phi}_{jK} h^K \tag{3}$$

where K is the degree of polynomial sequence chosen to interpolate the sets of hourly coefficients. It is straightforward to show $\sum_{h=1}^{24} \beta^{\phi}_{jh} H^h_t = \sum_{k=1}^{K} b^{\phi}_{jk} t^k$ for all ϕ and j, so that our main model in Eq. (1) boils down to

$$\begin{aligned} Y_{itd} &= \sum_{j=1}^{J}\sum_{k=1}^{K} \left[b^{cmpgn}_{jk} I^{cmpgn}_{id} + b^{event}_{jk} I^{event}_{id} \right] t^k D^j_i \\ &\quad + \sum_{j=1}^{J}\sum_{k=1}^{K} \left[b^{base}_{jh} + \mathbf{\Omega}_{td} \mathbf{b}^{weather}_{jh} \right] t^k D^j_i + \gamma_i + \epsilon_{itd} \end{aligned} \tag{4}$$

where, thanks to reduced dimensionality of the polynomial model, weather matrix $\mathbf{\Omega}_{td}$ now has five covariate columns, allowing for appliance-level quadratic effects of changing weather—that is, cooling degree hours, humidity, and their squares and interaction terms. As this specification results in near singular model matrices because of high degree of correlation between the polynomial sequence columns, we change basis in the interpolation by generating an orthogonal polynomial sequence of degree K for each j sequence of $\sum_k \{tD^j_i I^{cmpgn}_{id}\}^k$, $\sum_k \{tD^j_i I^{event}_{id}\}^k$, and $\sum_k \{tD^j_i\}^k$. This process ensures that any two different polynomials in any given j sequence remain orthogonal with each other. We applied the three-term recursion given by Kennedy and Gentle (1980) for this process. Let us denote the orthogonalization matrix as Γ^{ϕ}_j with its elements denoted by $\gamma^{\phi}_{j,mn}$ where $m, n \in \{0, ...K\}$. Then, the original sequence matrix, $T^{\phi}_j = [1; \tau_j; \tau^2_j; \cdots; \tau^K_j] \in \mathbb{R}^N \mathbb{R}^{K+1}$, and the new sequence matrix, $Z^{\phi}_j = [1; z_{j1}; z_{j2}; \cdots; z_{jK}] \in \mathbb{R}^N \mathbb{R}^{K+1}$, have the relationship of $Z^{\phi}_j = T^{\phi}_j \Gamma^{\phi}_j$, where $\Gamma^{\phi}_j \in \mathbb{R}^{K+1}\mathbb{R}^{K+1}$ is the upper triangular matrix defined by the recursion relationship. Then Eq. (4) is orthogonalized to give

$$\begin{aligned} Y_{itd} &= \sum_{j=1}^{J}\sum_{k=1}^{K} \left[\theta^{cmpgn}_{jk} z^{cmpgn}_{kj} + \theta^{mtrt}_{jk} z^{event}_{kj} \right] \\ &\quad + \sum_{j=1}^{J}\sum_{k=1}^{K} \left[\theta^{base}_{jk} z^{base}_{kj} + \sum_{m} \theta^{m}_{jk} \Omega_{mtd} z^{base}_{kj} \right] + \gamma_i + \epsilon_{itd} \end{aligned} \tag{5}$$

where θ^{ϕ}_{jk} is the coefficients to estimate, an element of the vector with the size of $k+1$, $\Theta^{\phi}_j \equiv \{0, \theta^{\phi}_{j1}, \theta^{\phi}_{j2}, \cdots, \theta^{\phi}_{jK}\}$. Note that as long as K is much less than 24, the two-step algorithm used for the main model may be no longer required to estimate Θ_j^ϕ, and the covariance matrix, $\widehat{V}(\Theta_j^\phi)$, may also be estimated from the direct one-step regression, instead of the above two-step approach.[14] The coefficients of the original polynomial model, $\widehat{\mathbf{b}}_j^\phi \equiv \{b_{j0}^\phi, b_{j1}^\phi, \cdots, b_{jK}^\phi\}$, as well as its covariance matrix, $\widehat{V}(\mathbf{b}_j^\phi)$, are recovered as follows:

$$\begin{aligned} \widehat{\mathbf{b}}_j^\phi &= \Gamma_j^\phi \widehat{\Theta}_j^\phi \\ \widehat{V}(\mathbf{b}_j^\phi) &= \Gamma_j^\phi \, \widehat{V}(\Theta_j^\phi) \, \Gamma_j^{\phi'} \end{aligned}$$

The final step is to calculate the hourly coefficient estimates and their variances, $\widetilde{\beta}_{jh}^\phi$ and $\widetilde{\mathrm{var}}(\beta_{jh}^\phi)$, of the campaign and marginal event-day treatment effects for all j and h, which can be computed by

$$\begin{aligned} \widetilde{\beta}_{jh}^\phi &= \sum_{k=0}^{K} \widehat{b_{jk}^\phi} h^k \\ \widetilde{\mathrm{var}}(\beta_{jh}^\phi) &= \sum_{m=0}^{K}\sum_{n=0}^{K} h^{m+n} \, \widehat{\mathrm{cov}}(b_{jm}^\phi, b_{jn}^\phi). \end{aligned}$$

For the rest of the paper, we present the hourly coefficients estimated from the 10^{th}- and 5^{th}-degree polynomial interpolation models. Fig. 6 and Fig. 7 list the hourly demand response for the entire set of appliances in the two cases. The overall pattern of hourly coefficients from either of the two models is consistent with those from the unrestricted model shown in Fig. 4. Statistically significant demand response during the campaign, desired peak-time curtailment made by a subset of the appliances with preparatory and prolonged responses shown around the peak hours, fairly limited load shifting, and a sharp morning curtailment followed by energy intensification for several appliances are all observed more evidently and consistently.

Fig. 5 takes the example of dish dryers (DD) to demonstrate how the polynomial models fit these baseline, campaign, and marginal event-day effects estimated from the main unrestricted model. The overall pattern of the estimates from the two polynomial models is remarkably consistent with those from the unrestricted model. It appears that the 10^{th}-degree polynomial model is flexible enough to allow the identification of the necessary demand response behavior at the appliance level while at the same time providing the above-mentioned parameter-space-dimension-reduction benefits.

[14] We employed a commercial cloud computing service featuring the memory size of 64.4 GB along with 36 vCPUs. Although the two-step Frisch-Waugh-Lovell process was required for the above unrestricted model, it was not required for the polynomial interpolation model but it was used to reduce computing time. As we ran models with up to a 10^{th}-degree polynomial dependency complemented by two additional sets of quadratic weather variables, the entire calculation finished within a couple of hours for the two-step process but more than ten hours for the direct one-step regression.

We test the equality between the unrestricted model and polynomial models, using the Wald statistic, $W = (\mathbf{R}_j^{\phi}\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})'(\mathbf{R}_j^{\phi}\widehat{\mathbf{V}}_{\boldsymbol{\beta}}\mathbf{R}_j^{'\phi})^{-1}(\mathbf{R}_j^{\phi}\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})$, where $\widehat{\boldsymbol{\beta}}$ is the consistent estimator of the main model's coefficients, $\widetilde{\boldsymbol{\beta}}$ is the consistent estimator of the polynomial model's coefficients, $\widehat{\mathbf{V}}_{\boldsymbol{\beta}}$ is the cluster-robust estimator of the covariance matrix for $\widehat{\boldsymbol{\beta}}$, and $\mathbf{R}_j^{\phi}$ is the restriction matrix assumed for appliance j for contribution type ϕ. As shown in Table 5. the higher the degree of the polynomial model, the smaller the statistical difference in parameter estimates between the main unrestricted model and either of the polynomial models. Compared to the case of the 5^{th} degree polynomial model, the 10^{th} degree case shrinks nearly all of the Chi-square statistic for all 24 hours, peak hours, and off-peak hours, with a greater number of tests failing to reject at the 1% significance level. Note that, even with the 5^{th} degree polynomial, some of the Wald equality tests for the campaign effect fail to reject at the same significance level in peak hours, and nearly all of the tests for the marginal event-day effect are not rejected both in peak and off-peak hours. This suggests that the polynomial interpolation model can reasonably well represent hourly demand response that the unrestricted model would identify, possibly enhancing understanding of appliance-level demand response within days with much reduced computational costs.

Our modeling exercise has so far assessed the effect of the PTR program for all appliance categories reported (j) in all hours-of-the-day (h) of all contribution types (ϕ). Below, we employ the dense parameter estimates to calculate period- and individual-level demand response, providing detailed insights into the development of a behavioral demand response program.[15]

5. Prediction of Demand Response

In this section, we use our estimated appliance-level hourly parameters to develop measures that are easier to interpret and perhaps more relevant to the decision making of the stakeholders.

5.1. Period-wise Demand Response

Using the estimated hourly parameters, we construct aggregate demand response measures during peak period (1-5 pm) and off-peak period (all other hours) in terms of the campaign effect and the full event-day effect, the latter of which is the summation of the campaign effect and the marginal event-day effect. The calculation of the period-wise demand response and their confidence intervals are described in Appendix A4.

[15] We also performed the group lasso (Meier et al., 2008, Yuan and Lin, 2006), grouping at the appliance and contribution level. The non-zero *campaign effect* results continue to hold for this modeling approach. The estimation process and results of the group lasso are discussed in Appendix A6.

Fig. 8 summarizes the aggregated demand response in peak and off-peak periods during campaign days (*campaign.effect*) and on event days (*full.eventday.effect*) and their 95% confidence intervals from the 10^{th} degree polynomial model. The figure nicely encapsulates the above findings from the analyses of hourly coefficients and their simultaneous exclusion tests. First of all, the treated businesses do alter electricity demand patterns for many of the appliances they own in both peak and off-peak periods, but they do so not only on event days but also during the experiment period, supporting the argument that they alter daily service routines in anticipation of dynamic price events.

Second, non-service critical appliances exhibit desired peak curtailment (DD, FS, IN, FL, LD, CP, MP, and IH), but the others do not. The peak-time energy intensification shown in the hour-level analysis is also evident in the period-wise analysis for service-critical appliances, such as AC, OR, and VC. The demand increase by these appliances more than offsets any attendant early morning conservation, resulting in net increases in electricity consumption on both event and non-event days.

Third, except WM, which shows positive load shifting toward the off-peak period, all of the peak-curtailing appliances entail net reductions in off-peak consumption as well. This response arises because the load shedding of the appliances often goes beyond the peak period due to the preparatory and prolonged response, as shown in the above hour-level analysis. We suggest that significant peak-time curtailment *and* strong (positive) load shifting into the off-peak period are hard to come by from a behavioral demand response program like this study, unless the demand response is automated or guided by timely information feedback.

5.2. Individual-level Demand Response

Our estimates of appliance-level hourly demand response and appliance ownership information for a small business allows us to predict of within-day hourly pattern of demand response. Fig. 9 presents the distribution of the individuals' demand response effects in peak and off-peak periods as predicted from the main unrestricted model (upper panels) and those from the 10^{th}-degree polynomial model (lower panels).[16]

Fig. 9 yields two key findings. First, there is substantial heterogeneity in the demand response of the small businesses to the PTR program. Although a large portion of the subjects delivers intended demand response (i.e., positive curtailment) in peak hours both on campaign days (left panels)

[16]The distribution of the individuals' demand response effects by business type is shown in Appendix A5, which indicates that within-sector variance of demand response is greater than between-sector variance.

and event days (right panels), a small but non-trivial portion ends up generating reverse demand response (i.e., increased usage). Most of the demand increase patterns are exhibited by businesses running several appliances in combination that exhibit the peak-time energy intensification (i.e., AC, VC, and OR). Note that the similarity in the level of demand response in peak and off-peak periods does not mean that they also exhibit a similar hourly demand response. As the peak period response is summed over the four hours (1-5 pm) while the off-peak period response over the remaining twenty hours, the hourly demand response is much higher for the peak period than for the off-peak period.

Second, consistent with our main findings, a large part of the demand response also involves reductions in both peak and off-peak usage during the campaign. That is, the individual businesses did alter their routine service operation during the campaign for the prospect of receiving generous rebates from dynamic pricing events but did little additional on event days. In addition, the polynomial interpolation approach shifts to the left the mode of the distribution of the unrestricted demand response. This shift seems to be an artifact of the restriction imposed by the polynomial model, which gives less coefficient variability between adjacent hours than the unrestricted one, yielding more enhanced demand response patterns for all periods. Also, such restrictions produce a more significant separation of distribution between peak and off-peak hours than the unrestricted case.

The validity of our individual-level prediction of demand response based on the appliance-level hourly treatment effect modeling is demonstrated in Fig. 10. It depicts mean hourly electricity usage and its demand response impacts all fitted by the estimated appliance-level coefficients in the three weather scenarios.[17] As shown, the higher the levels of temperature and humidity, the higher the electricity consumption during business hours, and the higher the magnitude of demand response. Our mean weather case indicates a reduction in peak-hour electricity usage during the campaign period by, on average, about 8% (0.1 kWh per hour). The fitted demand response pattern seems largely compatible with the case from the usual difference-in-difference model discussed Appendix A3.

The capability to predict potential demand response of individual electricity consumers and its distribution for the entire market base, as demonstrated above, is of significant importance to utility planners and energy service providers in improving the cost-effectiveness of a behavioral demand response program. For example, in recruiting program participants, a utility firm can have each of

[17] The mean usage plot is done by calculating the hourly demand responses of the individual customers in the three different scenarios and summing them up across the entire sample.

the applicants complete a simple checklist of appliance holdings to produce the first-order estimate of demand response. This way, the firm can segment the market into groups based on the predicted performance and thus price-discriminate across them by offering different dynamic incentive rates. In case where truthful revelation of private information from the applicant side is called into question, random auditing by utility personnel can be arranged as part of the contract between the firm and its customers. As long as the costs of recruitment and program administration are not exorbitantly high, the individual-level resource assessment and pricing approach would constitute the first-best strategy for the utility firm.

6. Conclusion

This paper investigates the demand response of small business consumers to a dynamic pricing tariff. Our hypothesis was that small businesses would re-configure how they do business to be able to reduce their demand during high-priced periods, rather than to reduce demand through one-off efforts during these time periods, which may cause a temporary or permanent loss of customers. Using a field experiment of the dynamic pricing campaign complemented with on-site appliance surveys, we determined weather-dependent within-day baseline patterns of appliance use and changes in these within-day patterns during the dynamic pricing campaign period and on dynamic pricing event days, without having to perform costly plug-load monitoring. We estimated both unrestricted within-day hourly response patterns, as well as response patterns with polynomial parameter dependency between hourly coefficients.

Our results demonstrate that the primary mode of response to facing a dynamic pricing tariff by small businesses is non-dynamic, consisting of a regular practice to curtail usage of appliances not critical to a positive customer and employee experience during all days of participation in the campaign. In contrast, we were unable to detect a precisely estimated incremental appliance-level consumption change from individual dynamic pricing events beyond these ongoing demand reductions. These findings are consistent with these businesses undertaking a re-configuration of service routines during the campaign because they have little flexibility in responding to day-to-day price signals because of our difficulty-of-breaking-the-routine hypothesis.

The analysis consistently reveals that the direction, magnitude, and timing of demand response are very different across the appliances in a manner that is consistent re-configuration of routines at the appliance level. Persistent peak hour consumption is observed only for non-service critical appliances, such as dish dryers, food storage units, lights, motors & pumps, and industrial heaters, with significant load shifting to off-peak hours hardly observed.

Our study contributes to ongoing policy discussion about the demand response potential of small commercial & industrial electricity consumers. The evidence of non-dynamic response dominating dynamic response implies that significant peak-time curtailment is hard to come by from a dynamic pricing program for small businesses, although these customers do significantly reduce their peak period consumption because they face dynamic prices. Our difficulty-of-breaking-routine result points to the critical role that automated demand response and timely information feedback may play to alleviate customer discomfort or limiting inefficiencies in business service practices, as a means to promote dynamic response to dynamic price signals. In addition, the appliance-level determination of electricity demand response demonstrated in our study should be of use to energy service providers in developing individual-level marketing strategies and thereby improving the overall cost-effectiveness of the demand response program.

References

Aigner, D. J., Sorooshian, C., and Kerwin, P. (1984). Conditional demand analysis for estimating residential end-use load profiles. *The Energy Journal*, 5(3).

Allcott, H. (2011). Rethinking real-time electricity pricing. *Resource and energy economics*, 33(4):820–842.

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica: Journal of the Econometric Society*, pages 178–196.

Bartels, R. and Fiebig, D. G. (2000). Residential end-use electricity demand: results from a designed experiment. *The Energy Journal*, 21(2).

Breiman, L. (2017). *Classification and regression trees*. Routledge.

Burkhardt, J., Gillingham, K., and Kopalle, P. K. (2019). Experimental evidence on the effect of information and pricing on residential electricity consumption. Technical report, National Bureau of Economic Research.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). The elements of statistical learning: prediction, inference and data mining. *Springer-Verlag, New York*.

Herter, K. (2009). Small business demand response with communicating thermostats: Smud's summer solutions research pilot.

Iacus, S. M., King, G., and Porro, G. (2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, 106(493):345–361.

Iacus, S. M., King, G., and Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political analysis*, 20(1):1–24.

Ito, K., Ida, T., and Tanaka, M. (2018). Moral suasion and economic incentives: Field experimental evidence from energy demand. *American Economic Journal: Economic Policy*, 10(1):240–67.

Jessoe, K. and Rapson, D. (2014). Knowledge is (less) power: Experimental evidence from residential energy use. *American Economic Review*, 104(4):1417–38.

Kennedy, W. J. and Gentle, J. E. (1980). *Statistical computing*. Routledge.

Meier, L., Van De Geer, S., and Bühlmann, P. (2008). The group lasso for logistic regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):53–71.

Monahan, T. and Fisher, J. A. (2010). Benefits of 'observer effects': lessons from the field. *Qualitative research*, 10(3):357–376.

Patrick, R. H. and Wolak, F. A. (2002). Real-time pricing and demand side participation in restructured electricity markets. In *Electricity Pricing in Transition*, pages 345–360. Springer.

Smirnov, N. V. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bull. Math. Univ. Moscou*, 2(2):3–14.

Somaini, P. and Wolak, F. A. (2016). An algorithm to estimate the two-way fixed effects model. *Journal of Econometric Methods*, 5(1):143–152.

Tibshirani, R., Wainwright, M., and Hastie, T. (2015). *Statistical learning with sparsity: the lasso and generalizations*. Chapman and Hall/CRC.

Wolak, F. A. (2007). Residential customer response to real-time pricing: The anaheim critical peak pricing experiment. Available at `http://www.stanford.edu/~wolak`.

Wolak, F. A. (2010). An experimental comparison of critical peak and hourly pricing: The *PowerCentsDC* program. Available at `http://www.stanford.edu/~wolak`.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.

Tables and Figures

Table 1: Sample summary statistics by the control and treatment group§

	Treatment	Control	T-C		Treatment	Control	T-C
Daily usage†	1.405	1.518	-0.113	Art & Sports	0.015	0.018	-0.003
[kWh/h]	(0.073)	(0.083)	(1.03)	[pct]	(0.006)	(0.007)	(0.28)
Peak usage†	2.171	2.203	-0.032	Biz Support	0.028	0.04	-0.013
[kWh/h]	(0.098)	(0.098)	(0.23)	[pct]	(0.008)	(0.01)	(0.98)
Off-peak usage†	1.253	1.381	-0.128	Construction	0.018	0.013	0.005
[kWh/h]	(0.07)	(0.081)	(1.2)	[pct]	(0.007)	(0.006)	(0.58)
Summer bill	206.932	201.978	4.955	Education Service	0.08	0.015	0.065
[10^3 KRW]	(9.952)	(10.175)	(0.35)	[pct]	(0.014)	(0.006)	(4.37)
AC installation	0.91	0.895	0.015	Engineering Service	0.04	0.018	0.023
[pct]	(0.014)	(0.015)	(0.72)	[pct]	(0.01)	(0.007)	(1.91)
Set temperature‡	24.4	24.4	0	Health Service	0.008	0.028	-0.02
[^{o}C]	(0.126)	(0.11)	(0)	[pct]	(0.004)	(0.008)	(2.16)
Floor area	19.13	19.63	-0.5	IT Service	0.015	0.018	-0.003
[pyeong]	(1.146)	(1.131)	(0.31)	[pct]	(0.006)	(0.007)	(0.28)
Operation hours	11.47	11.4	0.07	Manufacturing	0.025	0.033	-0.008
[hours]	(0.149)	(0.152)	(0.33)	[pct]	(0.008)	(0.009)	(0.63)
Employees no.	2.355	2.423	-0.068	Organization	0.018	0.02	-0.003
[person]	(0.115)	(0.112)	(0.42)	[pct]	(0.007)	(0.007)	(0.26)
Built before 1980	0.098	0.059	0.039	Personal Service	0.173	0.17	0.002
[pct]	(0.015)	(0.013)	(1.95)	[pct]	(0.019)	(0.019)	(0.09)
Built in 1980s	0.403	0.402	0	Restaurant & Bar	0.165	0.17	-0.005
[pct]	(0.025)	(0.027)	(0)	[pct]	(0.019)	(0.019)	(0.19)
Built after 1990s	0.5	0.539	-0.039	Realtor & Rental	0.068	0.098	-0.03
[pct]	(0.025)	(0.028)	(1.04)	[pct]	(0.013)	(0.015)	(1.54)
Owner-run	0.07	0.079	-0.009	Repair Service	0.028	0.025	0.003
[pct]	(0.013)	(0.015)	(0.47)	[pct]	(0.008)	(0.008)	(0.22)
Monthly rent	0.917	0.902	0.015	Retail	0.293	0.28	0.013
[pct]	(0.014)	(0.017)	(0.71)	[pct]	(0.023)	(0.022)	(0.39)
Long-term deposit	0.013	0.019	-0.006	Wholesale	0.03	0.058	-0.028
[pct]	(0.006)	(0.008)	(0.63)	[pct]	(0.009)	(0.012)	(1.91)

Note: §Standard errors reported in parenthesis; All variables except the first three metered observables are self-reported items from the on-site survey conducted at the time of sample recruitment; †These variables are hourly weekday averages in June and July of 2017, which are immediate two months before the start of the experiment; and ‡The set temperature is what the respondents stated about their set temperature of AC on typical summer weekdays.

Table 2: Equality tests for mean hourly usage between the treatment and control groups in different periods

	Before campaign		Non-event days		Event days	
	χ^2	p-value	χ^2	p-value	χ^2	p-value
Hourly usage [kWh]	23.9	0.467	24.9	0.409	23.5	0.488
Hourly usage [logkWh]	36.4	0.050	37.4	0.040	33.1	0.102

Note: Two-sample χ^2 tests were performed for the vectors of 24 hourly means of electricity usage of the treatment and control groups.

Table 3: Names, abbreviations, and counts

Appppliance†	Definition	Counts
AC	AC or cooling system	722
FN	electric fan	713
DD	dish dryer	24
RC	electric rice cooker	186
MW	microwave	216
OR	electric oven or range	51
FS	refrigerator or kimchi fridge	707
WD	electric water dispenser	521
WM	washing machine	106
VC	vacuum cleaner	121
IN	incandescent light	60
FL	fluorescent light	512
LD	LED light	379
CM	copy machine	188
CP	computer or laptop	503
TV	TV set	525
MP	electric motor or pump	20
IH	industrial heater or electric furnace	13
MI	electric heater, pad, or blanket	187
Business type	**Definition**	**Counts**
PSS	personal service	137
RPS	repair service	21
ORG	various organization	15
AAS	arts & sports service	13
HLS	healthcare service	14
EDS	education service	38
BNS	business support	27
ENG	engineering service	23
RAR	realtor & rental service	66
ITS	IT service	13
RAB	restaurant & bar	34
RTL	retail store	229
WSE	wholesale store	35
CST	construction service	12
MFG	small manufacturing	23

Note: †While our on-site interviews surveyed the ownership of each of 28 appliances, we re-categorized them into a total of 19 appliance types, according to their significance in constituting summer baseline and servicing purposes for small C&I customers.

Table 4: Wald tests for linear coefficient restrictions in the main model

	$\widehat{\beta_{jh}^{cmpgn}} = 0$		$\widehat{\beta_{jh}^{event}} = 0$		$\widehat{\beta_{jh}^{cmpgn}} + \widehat{\beta_{jh}^{event}} = 0$	
	peak $h^{\dagger}$	off-peak $h^{\ddagger}$	peak $h^{\dagger}$	off-peak $h^{\ddagger}$	peak $h^{\dagger}$	off-peak $h^{\ddagger}$
AC	73.1	266.0	2.5	28.7	31.9	136.7
FN	8.4	122.0	1.1	8.3	3.4	59.3
DD	30.6	208.0	0.9	13.2	10.0	135.7
RC	10.5	230.0	1.4	8.0	6.8	126.4
MW	2.3	165.0	1.2	17.6	2.6	61.8
OR	24.9	143.0	0.9	8.8	12.4	79.2
FS	9.8	181.0	0.7	16.9	2.9	105.9
WD	19.3	126.0	1.0	15.1	14.3	70.9
WM	6.5	217.0	0.7	25.3	3.4	131.0
VC	47.4	192.0	1.9	6.0	28.8	95.6
IN	24.8	216.0	3.0	20.3	24.5	135.6
FL	15.8	174.0	2.4	20.0	9.3	86.8
LD	16.9	199.0	1.2	15.0	8.6	130.4
CM	16.4	157.0	4.2	12.8	12.1	93.8
CP	15.3	143.0	2.6	5.1	14.3	69.5
TV	6.2	162.0	0.3	11.9	3.7	73.3
MP	15.8	335.0	3.4	12.2	8.7	208.2
IH	10.3	124.0	5.2	11.6	13.9	87.4
MI	76.9	132.0	3.8	13.3	21.0	39.2

Note: Wald tests were performed based on the OLS estimator of covariance matrix clustered at the customer-day level; †Critical values of chi-square distribution with 4 degrees of freedom at the significance level of 10%, 5%, and 1% are 7.8, 9.5, and 13.3, respectively; and ‡Critical values of chi-square distribution with 20 degrees of freedom at the significance level of 10%, 5%, and 1% are 28.4, 31.4, and 37.6, respectively.

Table 5: Wald tests for parameters from the main unrestricted model equal those from the polynomial model of degree 10 (upper) and 5 (lower)

	$\widehat{\beta_{jh}^{cmpgn}} = \widetilde{\beta_{jh}^{cmpgn}}$			$\widehat{\beta_{jh}^{mtrt}} = \widetilde{\beta_{jh}^{mtrt}}$		
10^{th} degree	all $h^{\dagger}$	peak $h^{\dagger\dagger}$	off-peak $h^{\dagger\dagger\dagger}$	all $h^{\dagger}$	peak $h^{\dagger\dagger}$	off-peak $h^{\dagger\dagger\dagger}$
AC	110.2	15.2	85.3	18.3	2.5	12.0
FN	80.8	4.2	72.5	6.0	1.0	5.3
DD	114.5	6.6	103.2	8.9	0.2	8.7
RC	179.2	9.3	156.5	8.4	1.9	6.3
MW	74.9	1.6	68.1	18.9	0.2	16.4
OR	45.6	5.8	35.2	10.0	1.2	6.9
FS	128.6	2.6	125.7	17.2	0.7	16.4
WD	84.2	8.0	68.0	15.2	0.6	12.8
WM	177.1	4.6	169.1	26.7	0.7	21.0
VC	140.2	7.6	134.7	4.1	1.7	2.3
IN	70.1	7.3	54.7	17.7	0.5	14.9
FL	92.5	3.5	80.9	14.2	1.6	12.3
LD	77.0	2.4	74.0	11.4	2.1	8.3
CM	109.6	7.3	93.4	11.5	1.8	8.6
CP	93.9	4.1	85.5	6.1	2.4	4.5
TV	91.3	33.0	54.6	12.8	1.0	10.6
MP	302.3	3.9	301.1	9.3	1.1	8.2
IH	112.6	11.6	105.0	19.7	3.6	12.0
MI	103.1	43.8	43.0	14.4	3.9	10.5
5^{th} degree	all $h^{\S}$	peak $h^{\S\S}$	off-peak $h^{\S\S\S}$	all $h^{\S}$	peak $h^{\S\S}$	off-peak $h^{\S\S\S}$
AC	303.0	18.8	283.9	33.1	3.5	26.8
FN	91.8	3.6	80.4	8.3	1.4	7.7
DD	218.4	19.8	210.9	14.5	0.5	14.1
RC	218.0	1.7	199.8	10.1	1.4	8.1
MW	117.7	4.6	105.4	20.8	1.2	14.9
OR	138.5	2.9	130.9	12.3	0.5	8.1
FS	166.2	8.0	145.2	15.8	0.5	14.7
WD	187.6	43.2	142.0	16.8	0.9	14.9
WM	276.3	11.7	254.2	34.5	0.3	27.2
VC	235.1	30.4	215.5	8.2	2.2	5.6
IN	184.0	8.2	152.4	19.9	1.5	17.7
FL	145.9	9.2	133.8	21.5	3.1	17.6
LD	150.7	9.3	138.2	18.1	2.0	15.9
CM	142.2	14.4	118.4	12.1	1.5	9.5
CP	130.7	16.5	126.7	6.2	1.3	4.7
TV	208.0	13.4	177.8	12.5	0.6	10.2
MP	397.7	3.0	388.2	11.9	3.6	9.4
IH	133.0	14.2	125.2	18.7	4.5	10.4
MI	154.4	51.5	88.6	17.3	3.6	14.2

Note: Wald tests were performed for the OLS estimator of covariance matrix clustered at customer-day level; †Critical values of chi-square distribution with 14 degrees of freedom at the significance level of 10%, 5%, and 1% are 21.1, 23.7, 29.1, respectively; ††Critical values with 6 *d.o.f.* for the significance level are 10.6, 12.6, and 16.8; †††Critical values with 10 *d.o.f.* for the significance level are 16.0, 18.3, and 23.2; §Critical values with 19 *d.o.f.* for the significance level are 27.2, 30.1, and 36.2; §§Critical values with 1 *d.o.f.* for the significance level are 2.7, 3.8, and 6.6; and §§§Critical values with 15 *d.o.f.* for the significance level are 22.3, 25.0, and 30.6.

Fig. 1. Two-sample Kolmogorov-Smirnov tests for the treatment and control groups in average daily usage (left), average peak usage (middle), and average off-peak usage (right) in June and July of 2017

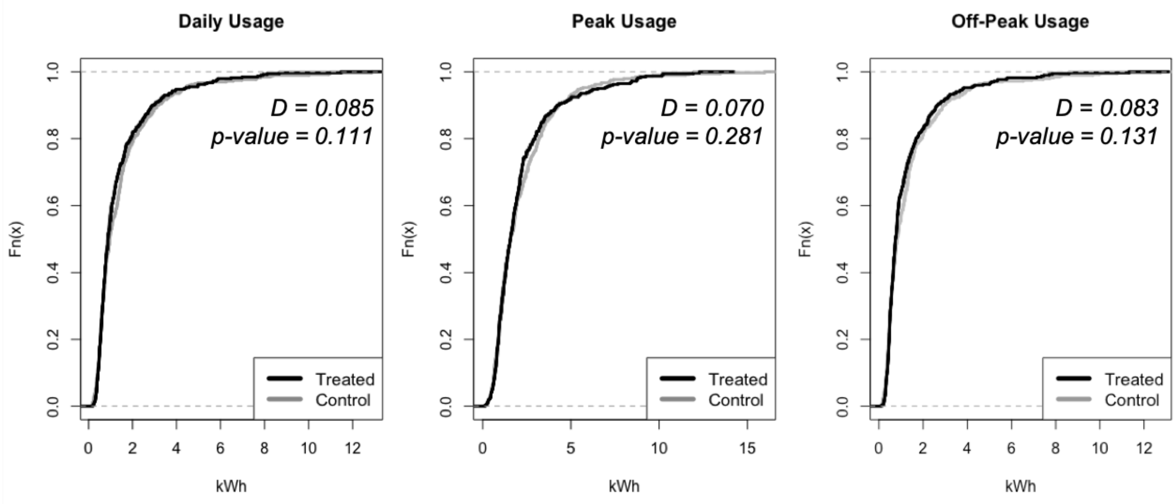


Fig. 2. Distribution of hourly electricity consumption of the treatment and control groups before the campaign (left), for non-event days during the campaign (middle), and for event days during the campaign (right)

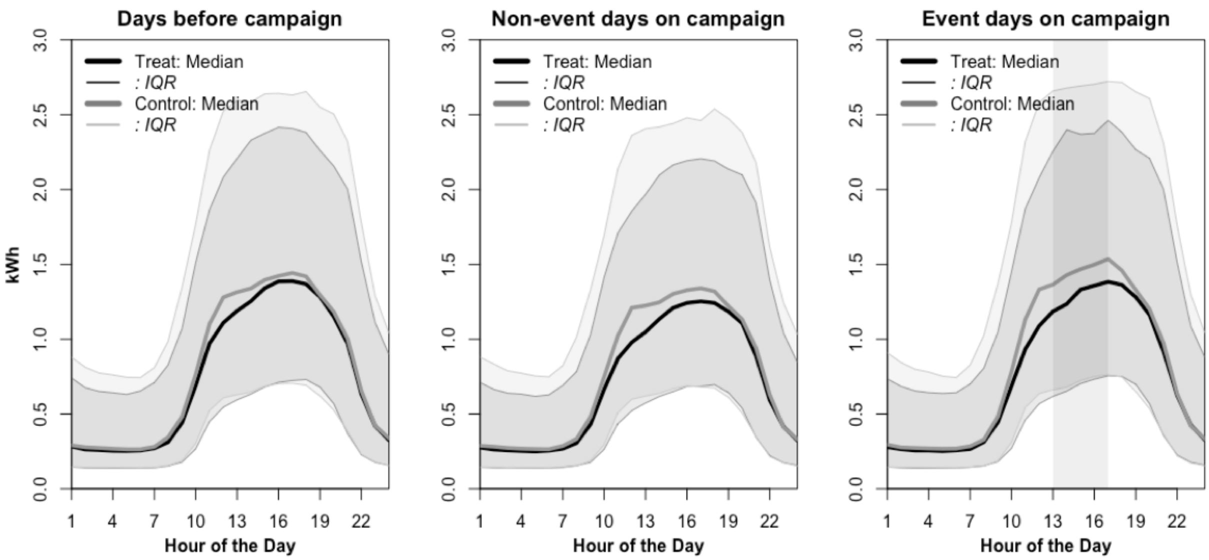


Fig. 3. Appliance-level estimates of hourly baseline consumption obtained from the main unrestricted model in the high, medium, and low temperature and humidity scenarios

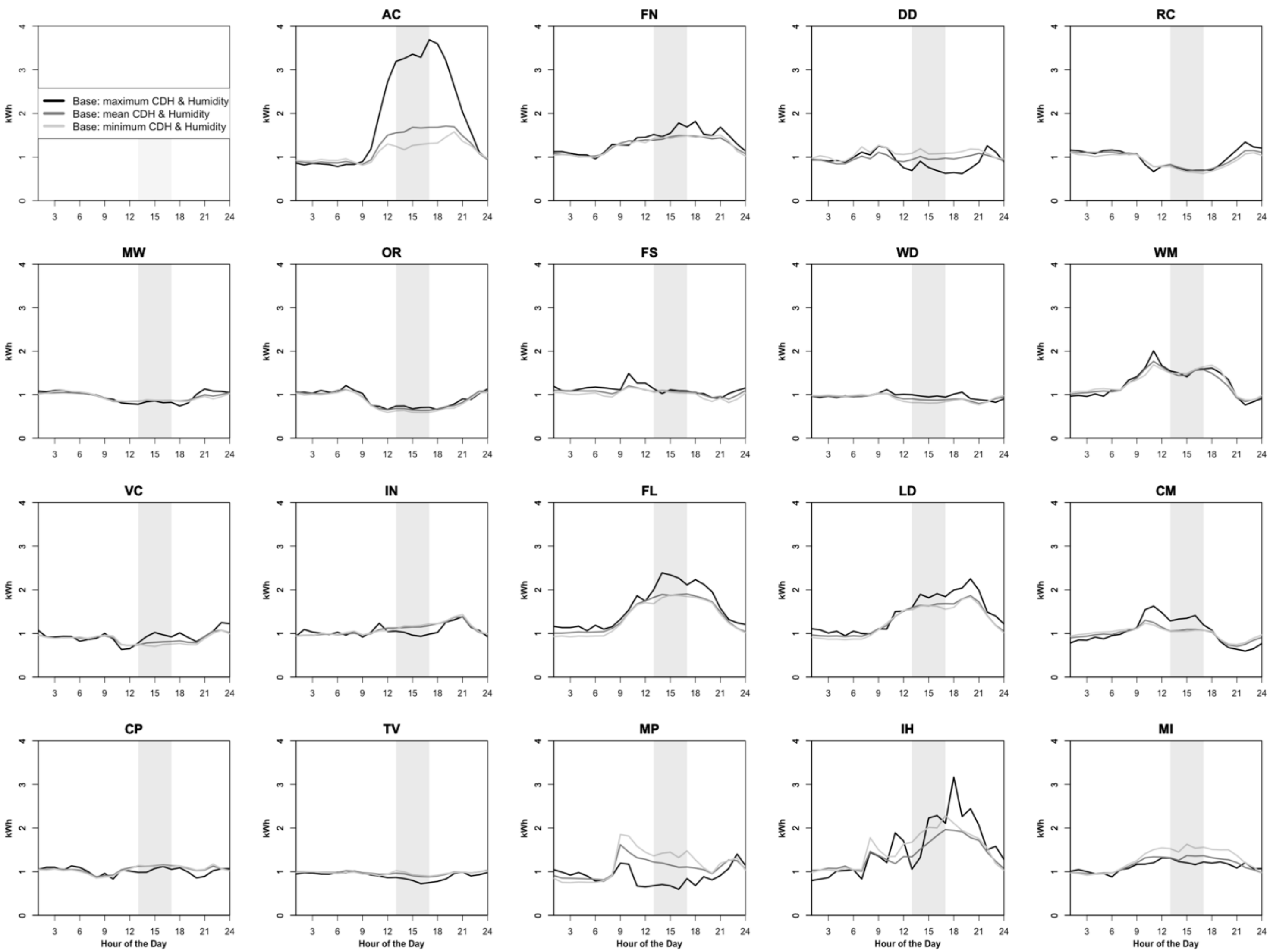


Fig. 4. Appliance-level estimates of the campaign treatment effect and the marginal event-day treatment effect obtained from the main unrestricted model

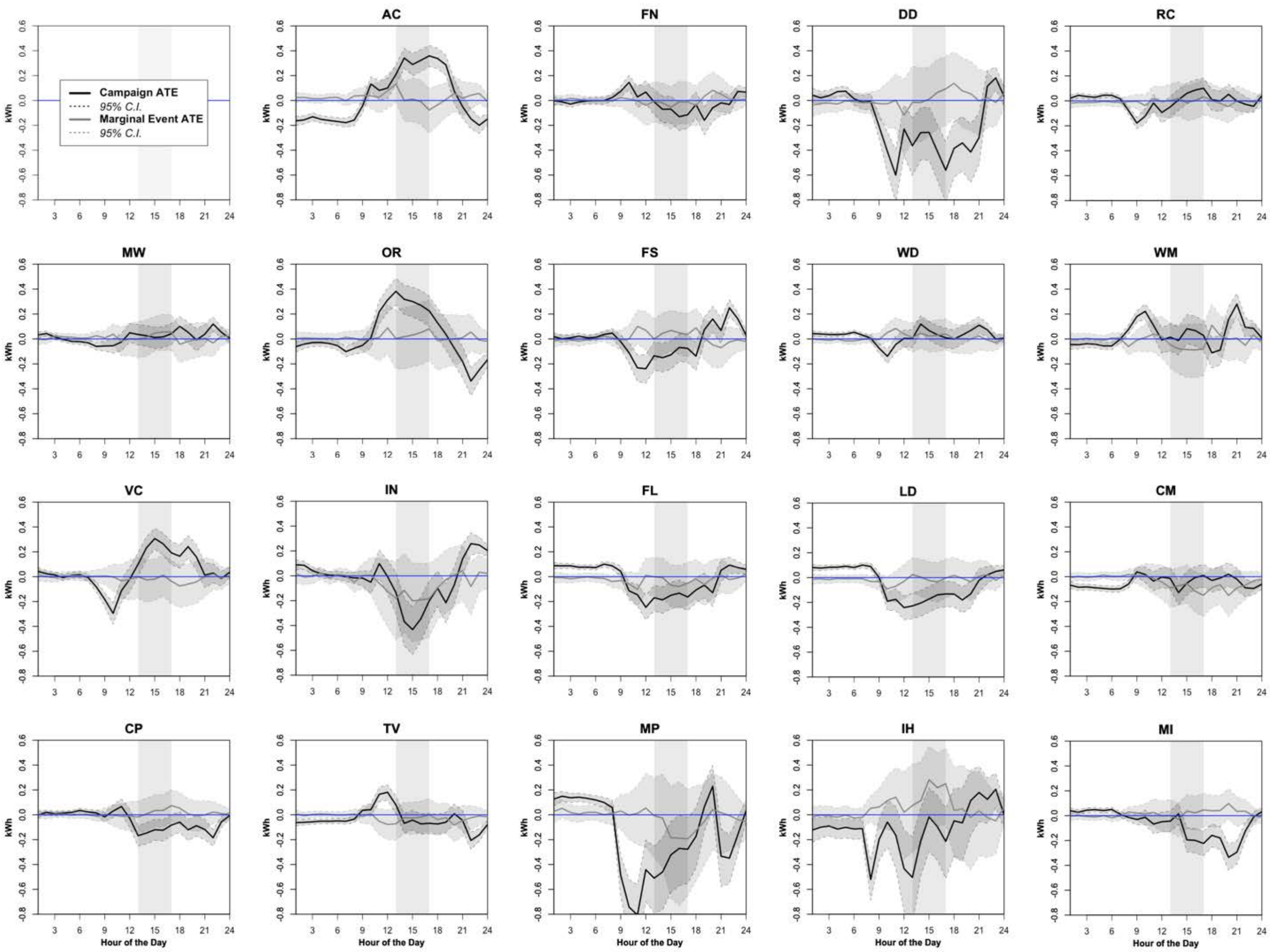


Fig. 5. The baseline effect (left), the campaign effect (middle), and the marginal event-day effect (right) of dish dryers (DD) estimated from the main unrestricted, the 10^{th} degree, and the 5^{th} degree polynomial models

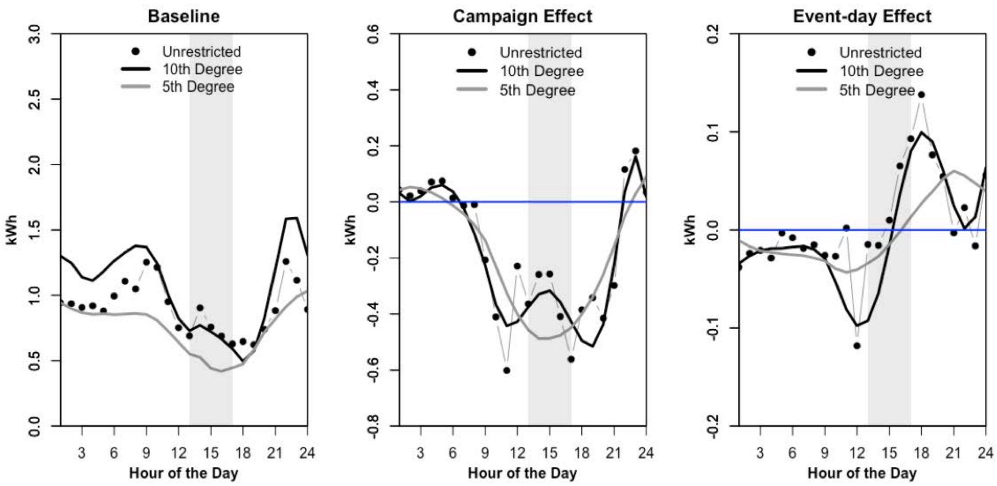


Fig. 6. Appliance-level estimates of the campaign treatment effect and the marginal event-day treatment effect obtained from the 10^{th} degree polynomial model

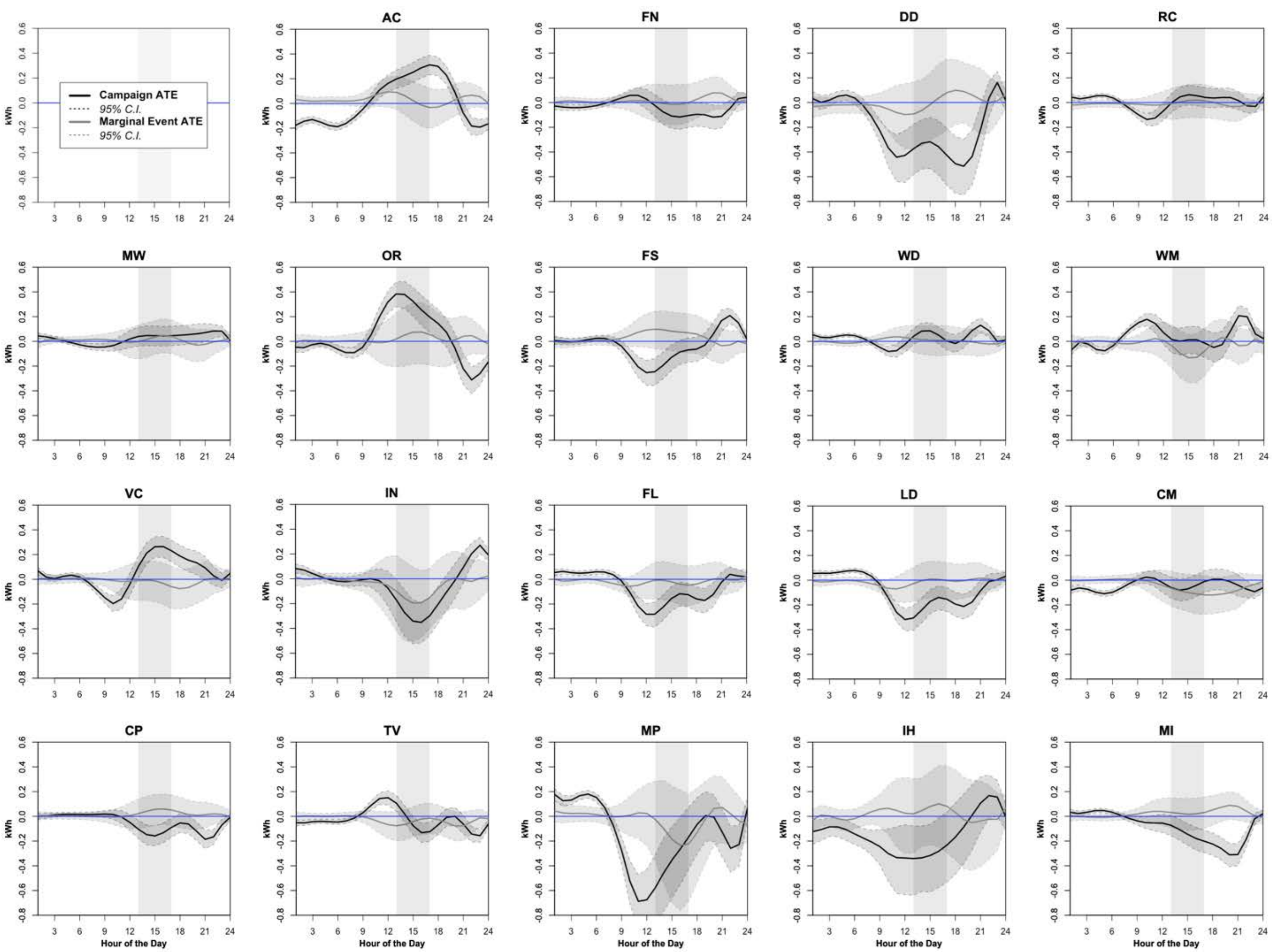


Fig. 7. Appliance-level estimates of the campaign treatment effect and the marginal event-day treatment effect obtained from the 5^{th} degree polynomial model

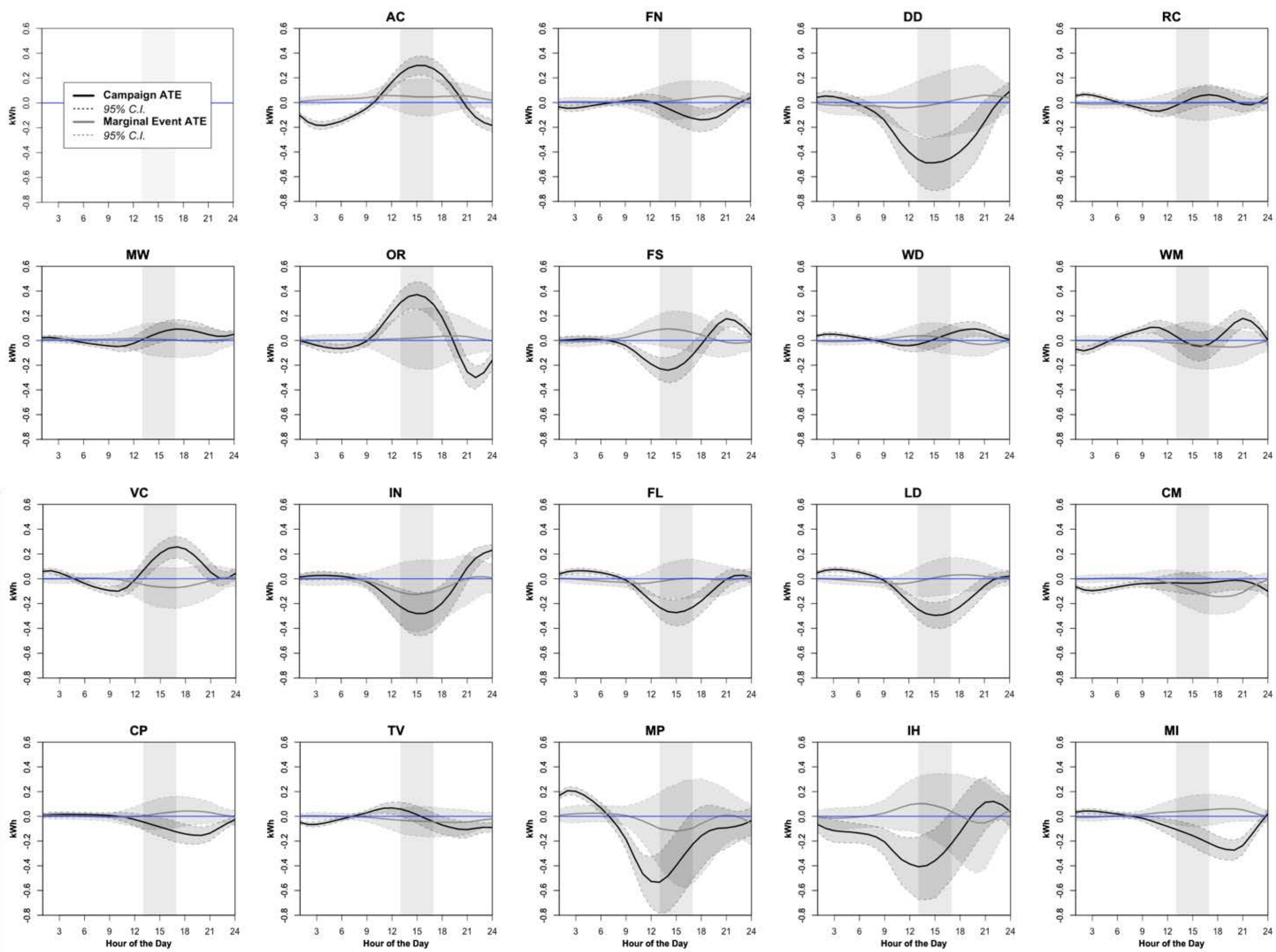
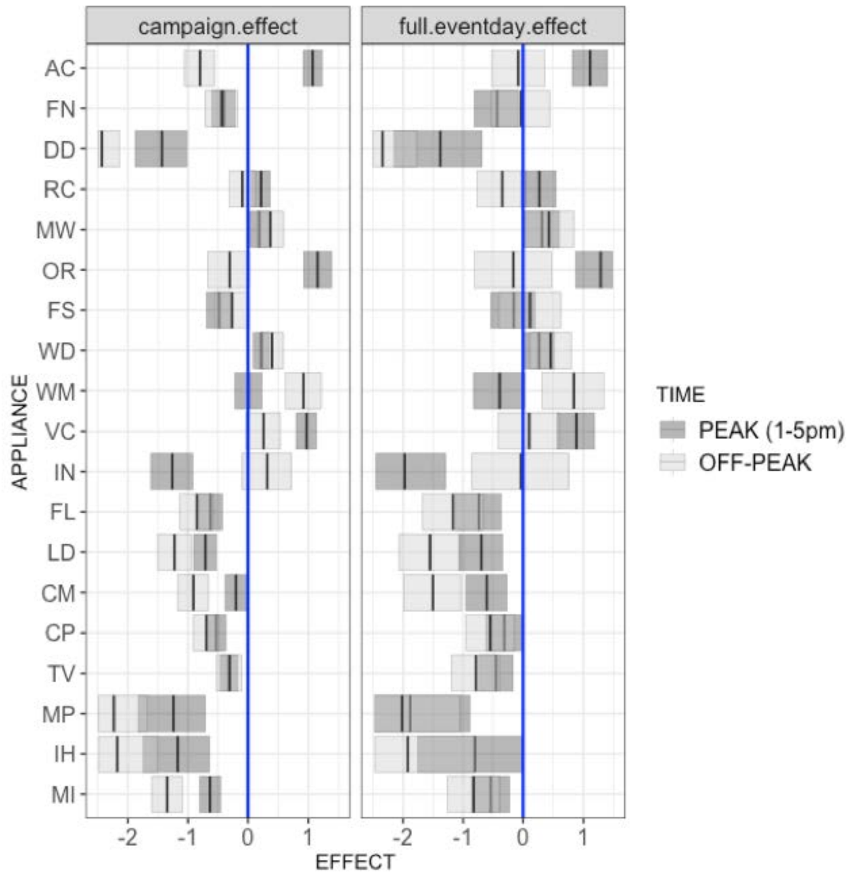


Fig. 8. Appliance-level estimates of peak and off-peak period demand responses on campaign days (*campaign.treat*) and full demand response on event days (*full.treat*) obtained from the 10^{th} degree polynomial model)



Note: The estimated effect is indicated by the band inside each bar, and the estimate's 95% confidence interval is shown by the bar's right and left ends.

Fig. 9. Predicted distribution of the individual subjects' peak and off-peak period responses during the campaign (*campaign effect*), marginal responses on event days (*marginal eventday effect*), and full response on event days (*full eventday effect*) obtained from the main unrestricted model (upper panels) and the 10-th degree polynomial model (lower panels)

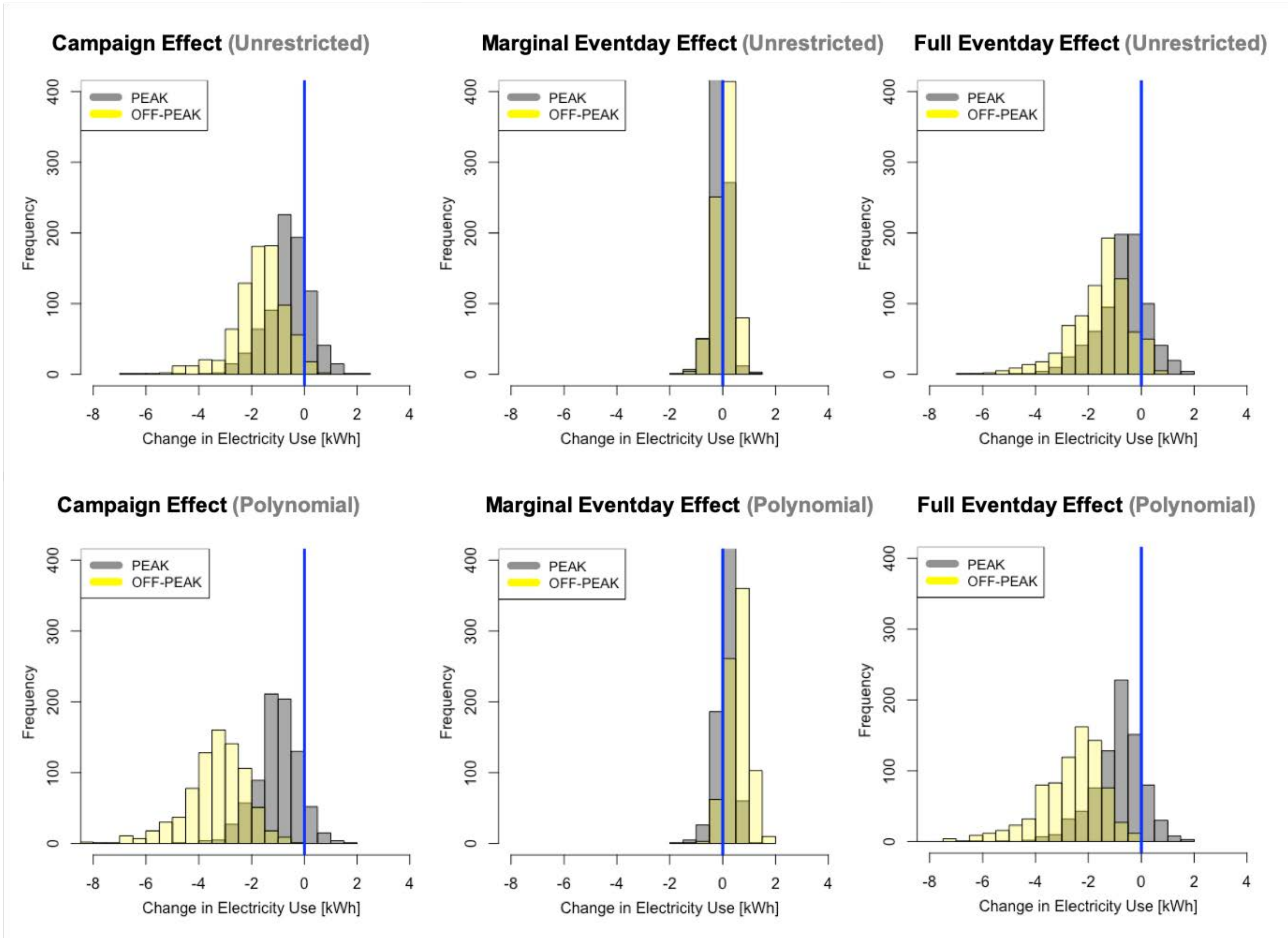
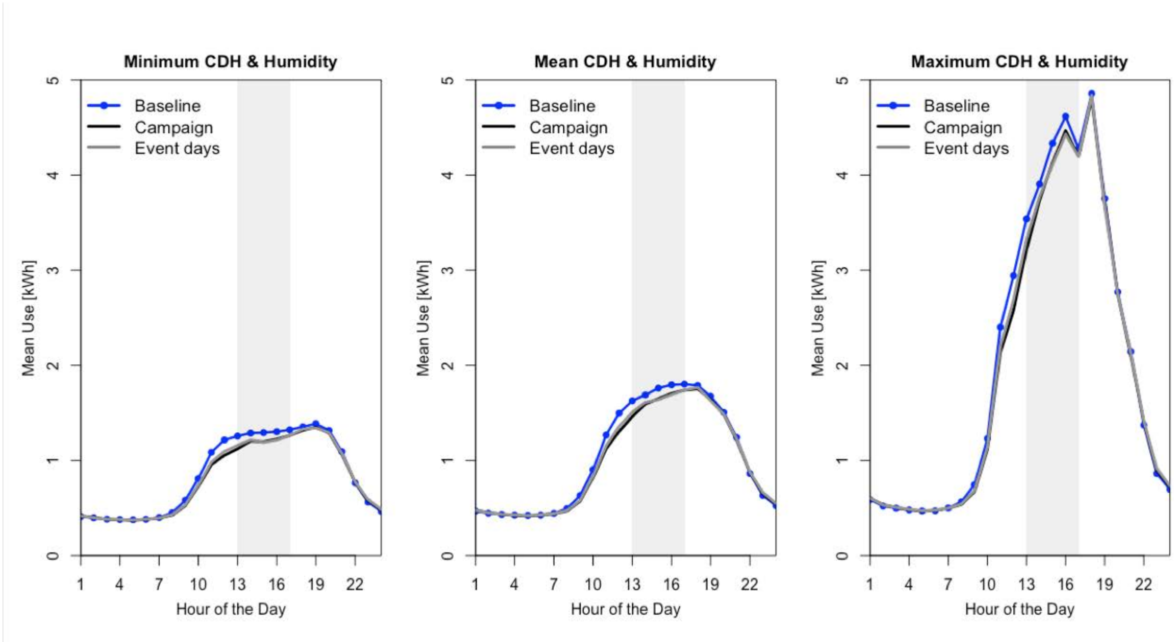


Fig. 10. Mean hourly electricity usage without the campaign (*baseline*), for non-event days during the campaign (*campaign*), and for event days (*event days*) obtained from the main unrestricted model in the high, medium, and low temperature and humidity scenarios



Appendix

A1. Sample Preparation and Assignment

The stratified sampling was originally conducted as follows. We first received a list of 33,000 small C&I customers who had hourly interval meters as part of the nationwide advanced metering deployment plan. These customers were assigned into one of 29 industry type clusters that KEPCO previously defined based on their business registration characteristics. The project team chose to group the industry type clusters for these customers into seven strata based on the mean and standard deviation of customer's electricity consumption in August and September of 2016. The first stratum (population frequency of 46.2%) includes small manufacturing, offices, non-store retailing, warehousing, non-retail shop, educational institutes, gyms, religious organizations, auto repair shops, electronics repair shops, and beauty salons. The second stratum (12.9%) includes bakeries, hotels, and restaurants, the third stratum (0.9%) has convenience stores, and the fourth stratum (2.8%) has snack bars, fast-food restaurants, sports facilities, and amusement centers. The fifth stratum (35.2%) includes wholesale and general retail businesses, general retailers, food retailers, refrigerated goods retailers, non-alcoholic beverage stores, nurseries and kindergartens, clinics, welfare facilities, and laundries. The sixth stratum (0.5%) has internet cafes, and the seventh stratum (1.4%) has grocery stores. Then a stratified random sample of control and treatment subject pairs was drawn from each strata to produce a list of target subjects to be used by the recruitment contractor.

During sample recruitment, however, there were unanticipated changes and additions to the original list of target subjects. When our recruiters approached the target subjects, many of them had incorrect names and business classifications, and a significant fraction of those with correct information declined to participate in the experiment. Given the tight project schedule, the team continued to recruit with a backup list of KEPCO small business customers. The recruitment process was repeated from May through July of 2017 until the target number of subjects of 1,515 were enrolled—1,000 for the treatment and 515 for the control group. The "Smart Save Days Campaign" started in early August and finished at the end of September 2017. Throughout the campaign, a total of 10 event days were declared for the peak hours for the treatment subjects. There were about 30 subjects that dropped out of the experiment because of inconvenience experienced or business relocation unplanned at the time of recruitment, as well as nearly 90 subjects that later had to be excluded due to malfunction of their interval meters. In response to the ongoing attrition, we undertook the parallel recruitment of additional 120 control subjects, eventually securing a total of 1,517 sample subjects—902 for the treatment and 615 for the control group.

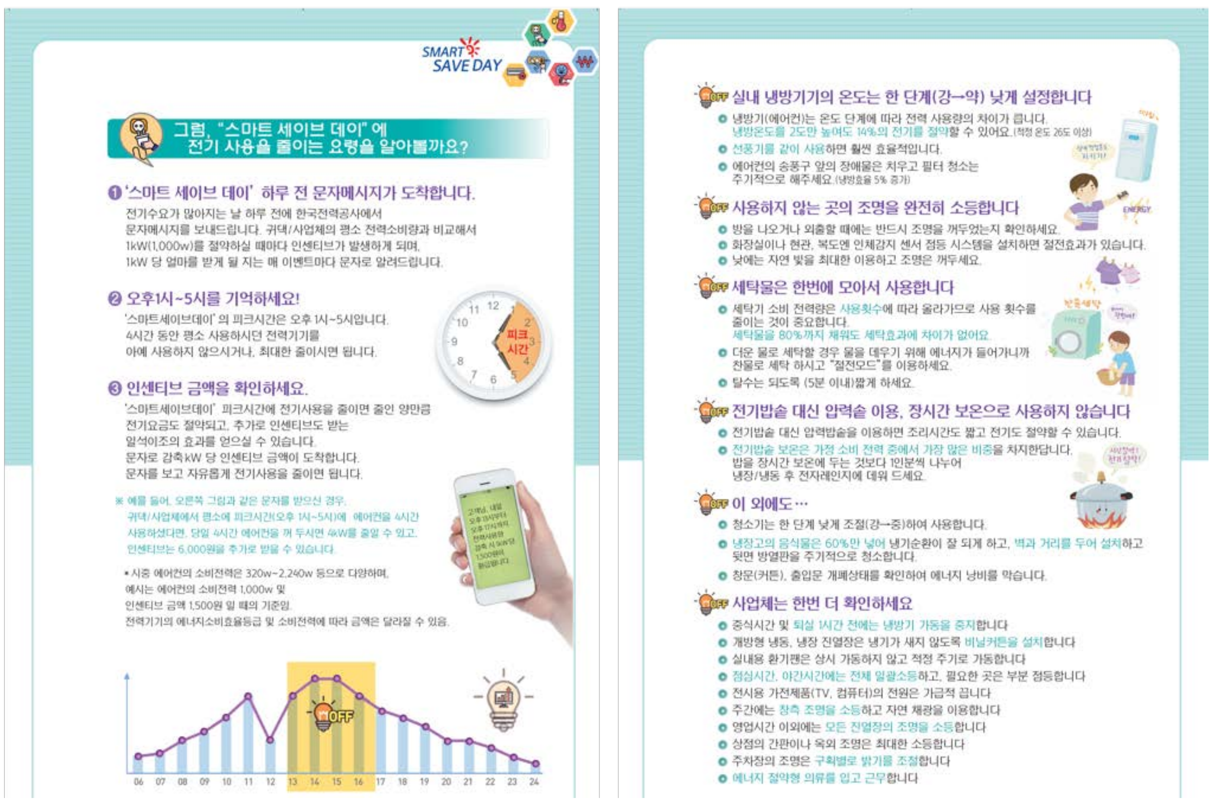
Because of the difference between the original sample assignment and the finally enrolled participants, as well as the sample attrition during the campaign, we attempted improve the integrity of randomization by performing a random re-assignment from the eventually secured 1,517 subjects now with correct business information and the full records of electricity consumption. The re-assignment was done based on the stratified sampling frequencies used at the time of designing our experiment, but with the correct business characteristics information. We attempted to use as many observations in our sample as possible, while maintaining the original sampling frequencies across the strata keeping an equal number of the treatment and control subjects. This procedure sampled 372 subjects from 664 subjects belonging to the 1^{st} stratum (frequency of 46%), 104 from 213 in the 2^{nd} stratum (13%), 10 from 29 in the 3^{rd} stratum (1%), 22 from 43 in the 4^{th} stratum (3%), 282 from 520 in the 5^{th} stratum (35%), and 10 from 28 in the 7^{th} stratum (1%)—the 6^{th} stratum was collapsed with the 3^{rd} stratum due to small size of this strata. In total, the re-assigned sample consisted of 800 subjects—400 for the treatment and 400 for the control group—which would be representative of small C&I customer population in Seoul. We used this re-assigned sample for all of the analyses presented in the paper.

As a robustness check, we also performed a coarsened exact matching (CEM) process which created statistically equivalent groups of treatment and control subjects, in an attempt to control for the potentially confounding influence of pre-treatment control variables as described in Iacus et al. (2011) and Iacus et al. (2012). Variables used for coarsening encompass nearly all of the covariates that might influence the demand response behavior of the subjects. They include observables such as average peak hour (1-5 pm) and off-peak hour usage and average morning (10 am to 1 pm) and evening work-hour (5-8 pm) usage on pre-experiment weekdays, as well as self-reported survey items, such as the last summer's electricity bill, occupied floor area, built year of the buildings, usual hours of operation, number of employees, and dummies for being classified as small manufacturing business and being gas-heated in the winter. To coarsen the continuous variables, we employed the standard automated univariate histogram method for each of the variables to create eleven equally spaced bins, except the four pre-experiment period consumption variables, for which the number of bins decreased to six so that the size of matched subjects can be increased at a small expense of sample balance. Then all observations were sorted into a total of 534 strata each with the identical values for the coarsened variables and removing all observations in any stratum without at least one treated and one control unit. This CEM procedure recovered a total of 1,012 subjects—588 for the treatment and 424 for the control group. The appliance-level demand response effects estimated from the matched treatment-control pairs and key insights remained almost the same as those from our sample re-assignment described above, although the former approach produced a sample with a slightly lower average electricity consumption than the latter sample.

A2. Other Details of the Field Experiment

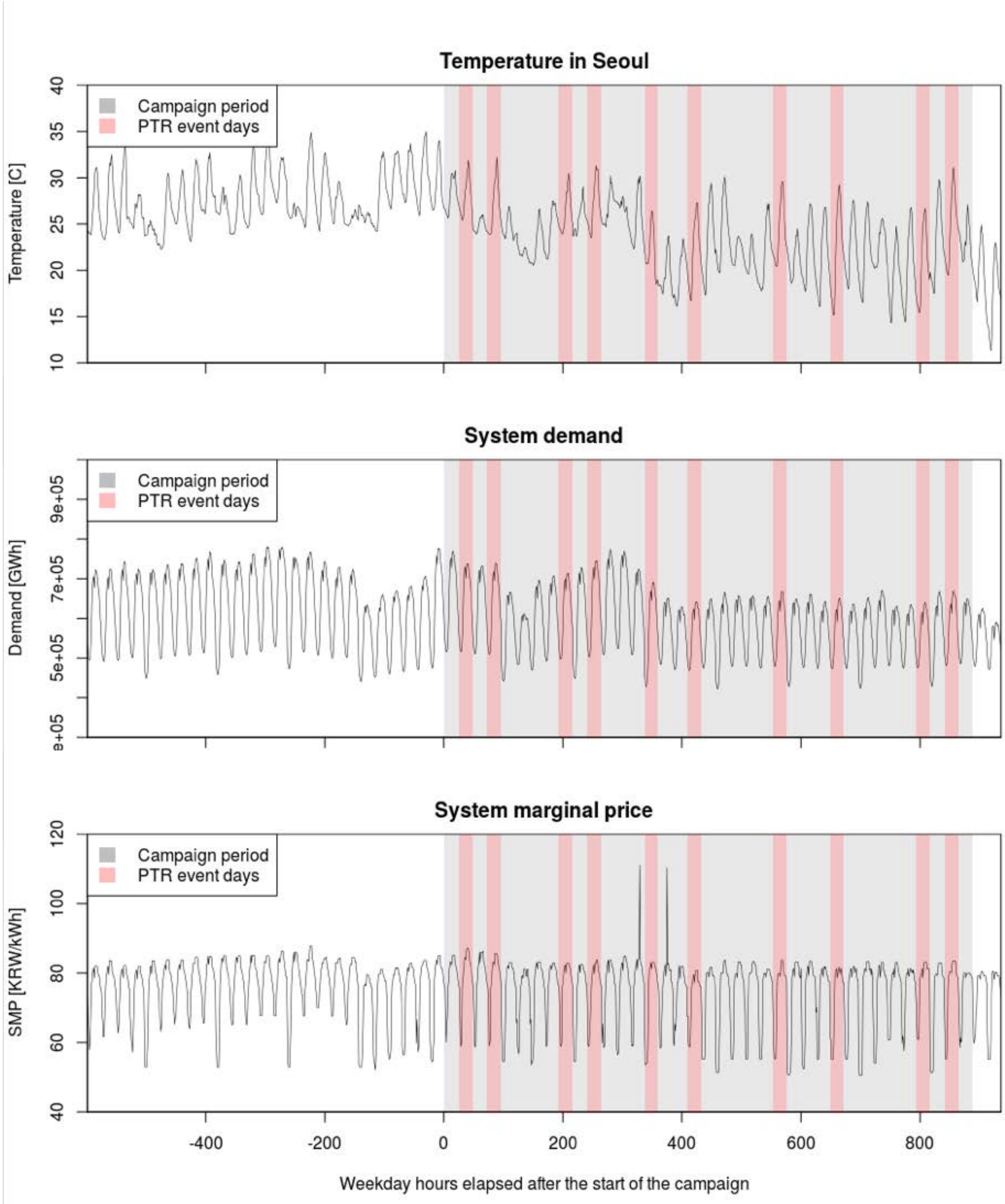
In our experiment, the treatment subjects were given with a short on-site education on the PTR that they would be subject to in a couple of weeks. The education was assisted by the leaflets shown below, which covers the duration and operation of the campaign and possible measures to reduce peak-time electricity usage on event days.

Fig. A1. KEPCO's educational leaflets on the "Smart Save Day" campaign



Note:

Fig. A2. Temperature in Seoul, system hourly demand, and system marginal price cleared at KPX during the sample period



A3. Supplementary Analyses and Figures on Treatment Effects

In this section, we estimate the aggregate average treatment effect (ATE) using a usual difference-in-difference specification, which is given by:

$$\begin{aligned} Y_{itd} &= \sum_{h=1}^{24} \beta_h^{cmpgn} H_t^h I_{id}^{cmpgn} + \sum_{h=1}^{24} \beta_h^{event} H_t^h I_{id}^{event} \\ &\quad + \sum_{h=1}^{24} \beta_h^{base} H_t^h + \boldsymbol{\Omega}\boldsymbol{\beta}_h^{weather} + \gamma_i + \epsilon_{itd} \end{aligned} \tag{6}$$

The coefficient of interest, β_h^{base}, β_h^{campgn}, and β_h^{event}, represents the baseline effect, the campaign treatment effect, and the marginal event-day treatment effect all in hour h, respectively. Note that this specification is very different from our main specification shown in Section 4.2., in that it does neither consider contributions from various appliances nor account for their different weather sensitivities.

The left panel of Figure A3 presents the estimation results.[18] The right panel shows the coefficient estimates for the campaign effect and the marginal event-day effect with their pointwise 95 % confidence intervals. Relative to the baseline case, the very presence of the campaign indeed reduces hourly electricity usage by, on average, about 0.1 kWh (8%) in and around the peak hours, and the effect remains statistically significant at the 5% level or lower. The campaign treatment effect begins in the morning, steeply increasing until noon before it tapers off and vanishes by the early evening. However, the marginal event-day treatment effect that comes into play only on event days seems to be not as precisely estimated, except during the hours adjacent to the event peak hours. Relative to the campaign-only case, the treatment subjects seem to use more electricity (or, equivalently, reduce less) for several hours before the event peak hours and use slightly less (or reduce more) several hours after. Note that the insignificant marginal event-day effect does not at all mean that the full event-day treatment effect is similarly small. The full event-day effect, which is the direct summation of the campaign effect and the marginal event-day effect, remains about the same magnitude as the campaign effect and precisely estimated as well.

[18] The mean usage profiles are fitted values after conditioning on all control variables except for the customer's fixed effects

Table A1: Wald tests for linear coefficient restrictions in the aggregate DID model

	$\widehat{\beta_h^{cmpgn}} = 0$	$\widehat{\beta_h^{event}} = 0$	$\widehat{\beta_h^{cmpgn}} + \widehat{\beta_h^{event}} = 0$
All hour coefficients†	457	35.4	145
Peak hour coefficients††	75.1	4.99	26.9
Off-peak hour coefficients†††	451	30.7	134

Note: Wald tests were performed based on the OLS estimator of covariance matrix clustered at the customer-day level; † critical values of chi-square distribution with 24 degrees of freedom at the significance level of 10%, 5%, and 1% are 33.2, 36.4, and 43.0, respectively; †† critical values of chi-square distribution with 4 degrees of freedom at the significance level of 10%, 5%, and 1% are 7.8, 9.5, and 13.2, respectively; and † † † critical values of chi-square distribution with 20 degrees of freedom at the significance level of 10%, 5%, and 1% are 28.4, 31.4, and 37.6, respectively.

Fig. A3. Mean hourly electricity usage without the campaign (*Baseline*), for non-event days during the campaign (*Campaign*), and for event days during the campaign (*Event days*) obtained from the main unrestricted model after conditioning on all control variables other than individuals' fixed effects (LEFT); and average treatment effect during the campaign (*Campaign Effect*) and marginal event-day treatment effect (*Marginal Event Effect*) (RIGHT)

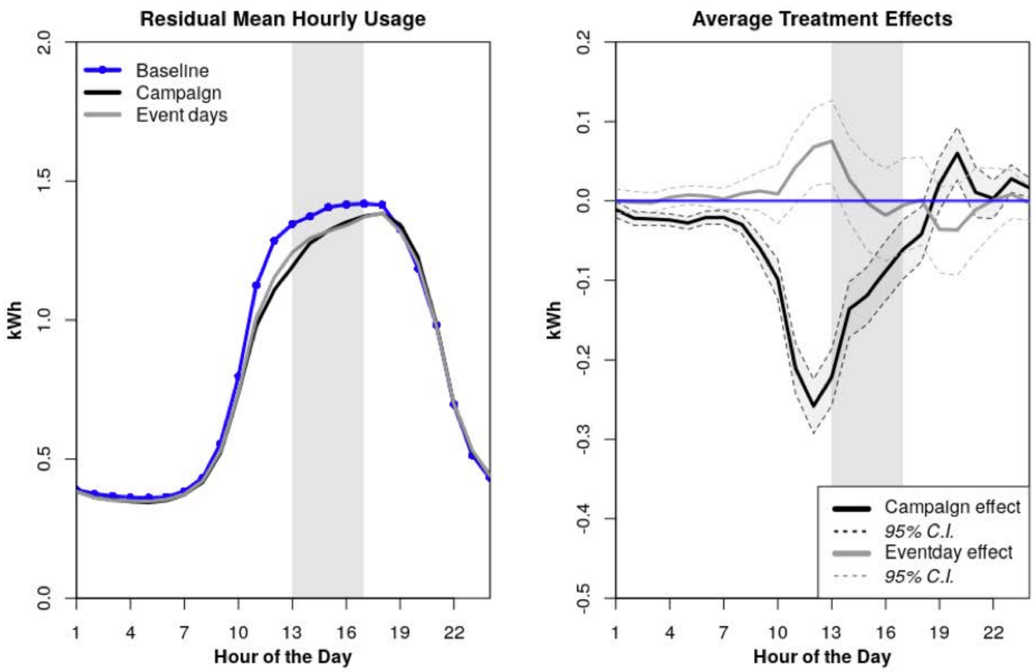


Fig. A4. The campaign treatment effects and the marginal event-day treatment effects for treatment subjects who had checked 7-10 event notices (black) versus 0-6 event notices (gray)

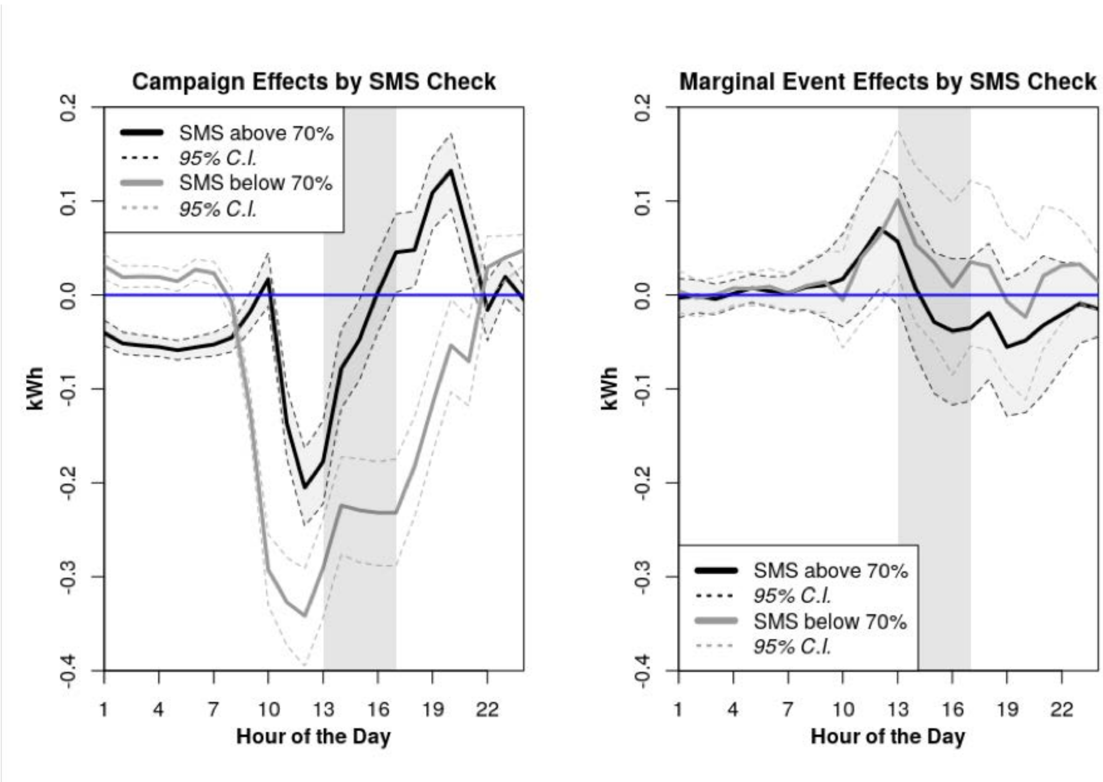


Fig. A5. Appliance-level estimates of hourly baseline consumption obtained from the 10^{th} degree polynomial model in the high, medium, and low temperature and humidity scenarios

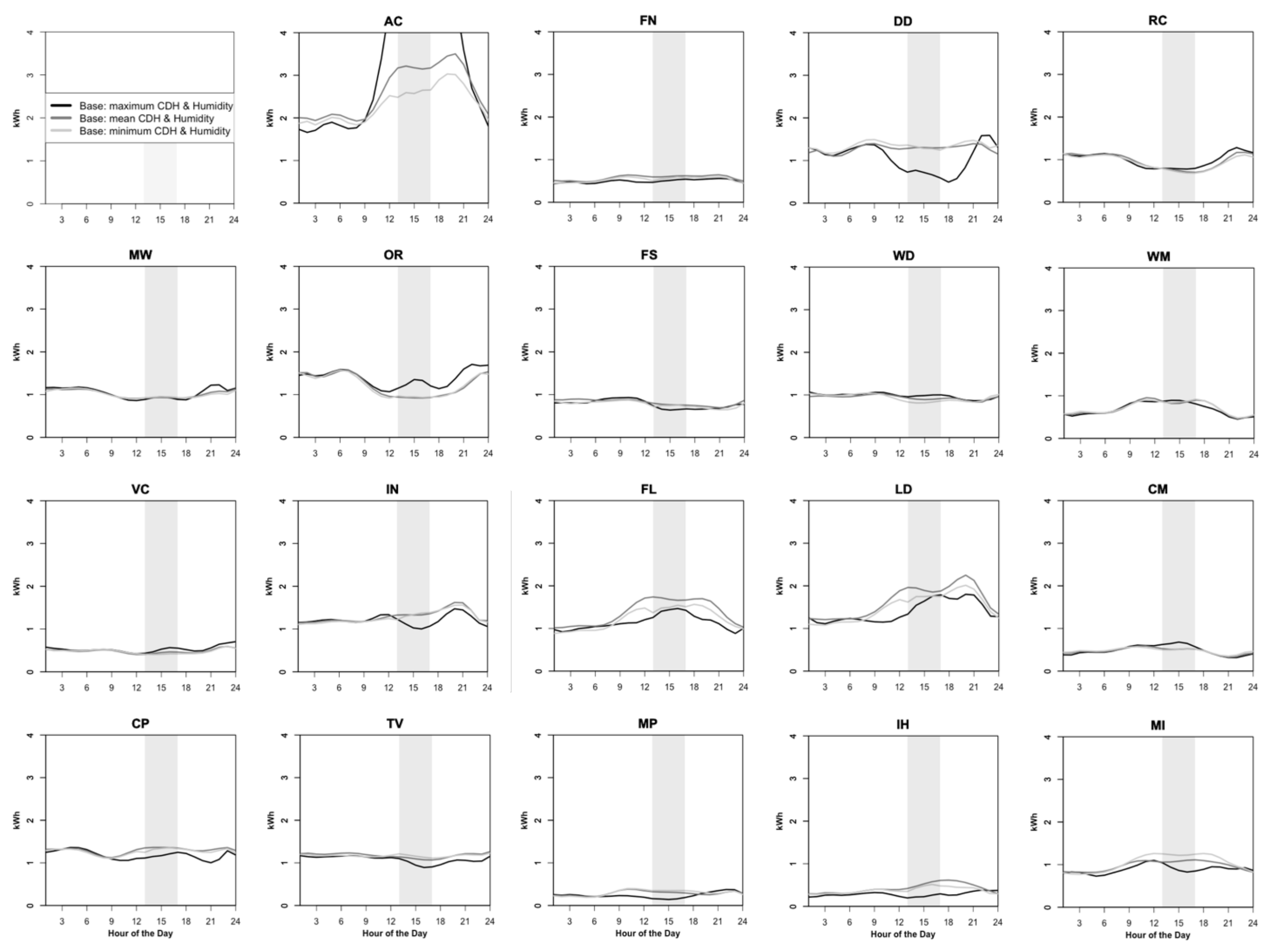
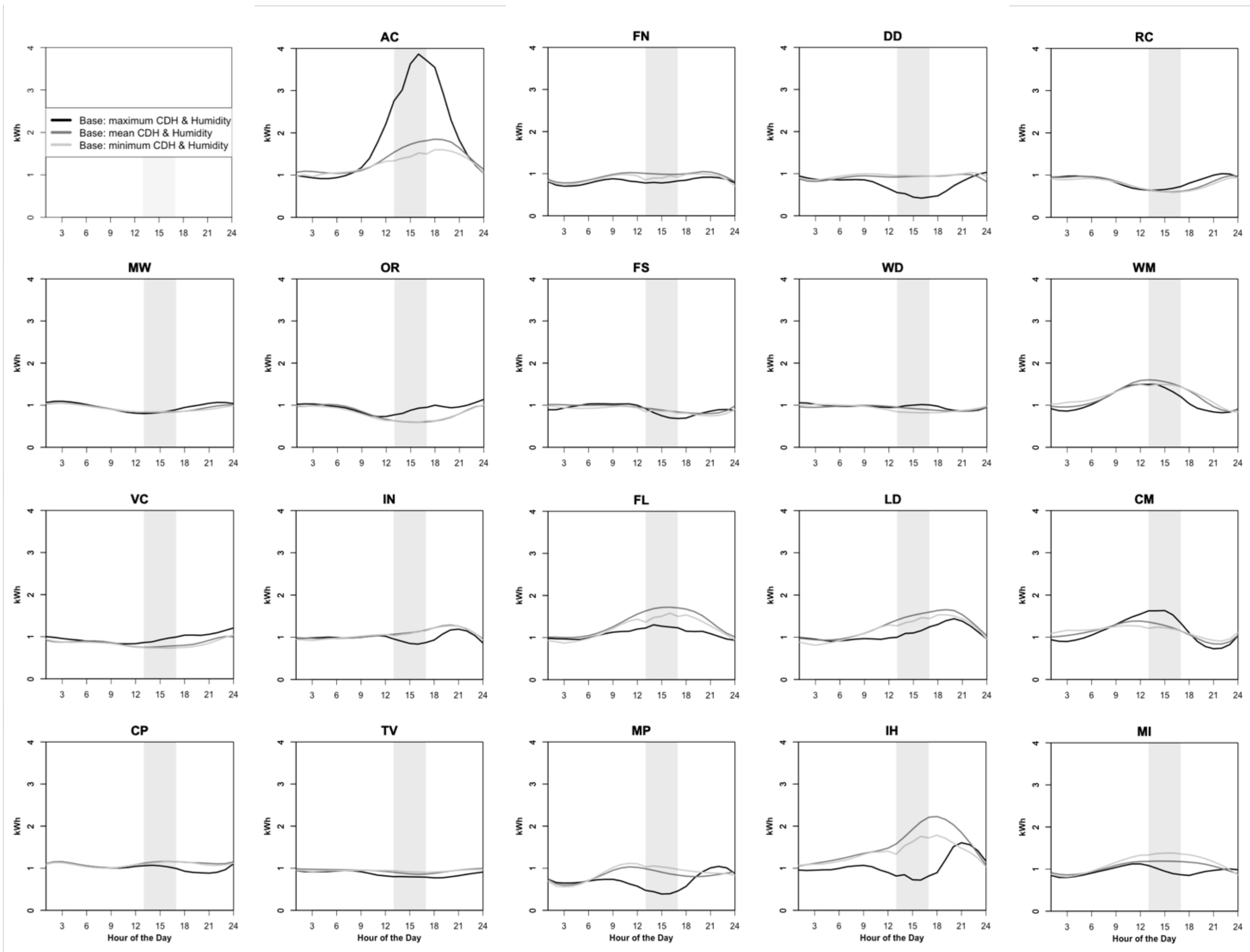


Fig. A6. Appliance-level estimates of hourly baseline consumption obtained from the 5^{th} degree polynomial model in the high, medium, and low temperature and humidity scenarios



A4. Prediction of Period- and Individual-level Demand Response

Section 5.1 discusses period-wise demand response. The peak-time demand response of appliance j during the campaign, the additional demand response on event days, and the full event-day demand response are denoted by $\Delta_{j,\mathbf{P}}^{cmpgn}$, $\Delta_{j,\mathbf{P}}^{event}$, and $\Delta_{j,\mathbf{P}}^{full}$, respectively, with their off-peak period counterparts by, $\Delta_{j,\mathbf{OP}}^{cmpgn}$, $\Delta_{j,\mathbf{OP}}^{event}$, and $\Delta_{j,\mathbf{OP}}^{full}$, which are calculated as follows:

$$\begin{aligned}
\Delta_{j,\mathbf{P}}^{cmpgn} &= \sum_{h\in\mathbf{P}} \left[1-\exp(-\widehat{\beta_{jh}^{cmpgn}})\right]\bar{Y}_h \\
\Delta_{j,\mathbf{P}}^{event} &= \sum_{h\in\mathbf{P}} \left[1-\exp(-\widehat{\beta_{jh}^{event}})\right]\bar{Y}_h \\
\Delta_{j,\mathbf{P}}^{full} &= \sum_{h\in\mathbf{P}} \left[1-\exp(-\widehat{\beta_{jh}^{cmpgn}}-\widehat{\beta_{jh}^{event}})\right]\bar{Y}_h \\
\Delta_{j,\mathbf{OP}}^{cmpgn} &= \sum_{h\in\mathbf{OP}} \left[1-\exp(-\widehat{\beta_{jh}^{cmpgn}})\right]\bar{Y}_h \\
\Delta_{j,\mathbf{OP}}^{event} &= \sum_{h\in\mathbf{OP}} \left[1-\exp(-\widehat{\beta_{jh}^{event}})\right]\bar{Y}_h \\
\Delta_{j,\mathbf{OP}}^{full} &= \sum_{h\in\mathbf{OP}} \left[1-\exp(-\widehat{\beta_{jh}^{cmpgn}}-\widehat{\beta_{jh}^{event}})\right]\bar{Y}_h
\end{aligned}$$

where $\bar{Y}_h$ is the average electricity consumption of the treatment subjects in hour h during the campaign. The confidence intervals of these period-wise demand responses are obtained from parametric bootstrapping with all standard errors of the coefficients clustered at the level of day-of-the-sample-of-individual.

Section 5.2 predicts individual-level aggregate demand responses. The individual-level demand responses in peak period in terms of the campaign effect, the marginal event-day effect, and the full event-day effect are calculated, respectively, as follows:

$$\begin{aligned}
\Delta_{\mathbf{P}}^{cmpgn}(i) &= \sum_{j\in J} \Delta_{j,\mathbf{P}}^{cmpgn}\ D_i^j \\
\Delta_{\mathbf{P}}^{event}(i) &= \sum_{j\in J} \Delta_{j,\mathbf{P}}^{event}\ D_i^j \\
\Delta_{\mathbf{P}}^{full}(i) &= \sum_{j\in J} \Delta_{j,\mathbf{P}}^{full}\ D_i^j.
\end{aligned}$$

The individual-level demand responses in off-peak period are similarly given as:

$$\begin{aligned}
\Delta_{\mathbf{OP}}^{cmpgn}(i) &= \sum_{j\in J} \Delta_{j,\mathbf{OP}}^{cmpgn}\ D_i^j \\
\Delta_{\mathbf{OP}}^{event}(i) &= \sum_{j\in J} \Delta_{j,\mathbf{OP}}^{event}\ D_i^j \\
\Delta_{\mathbf{OP}}^{full}(i) &= \sum_{j\in J} \Delta_{j,\mathbf{OP}}^{full}\ D_i^j.
\end{aligned}$$

A5. Demand Response by Business Type

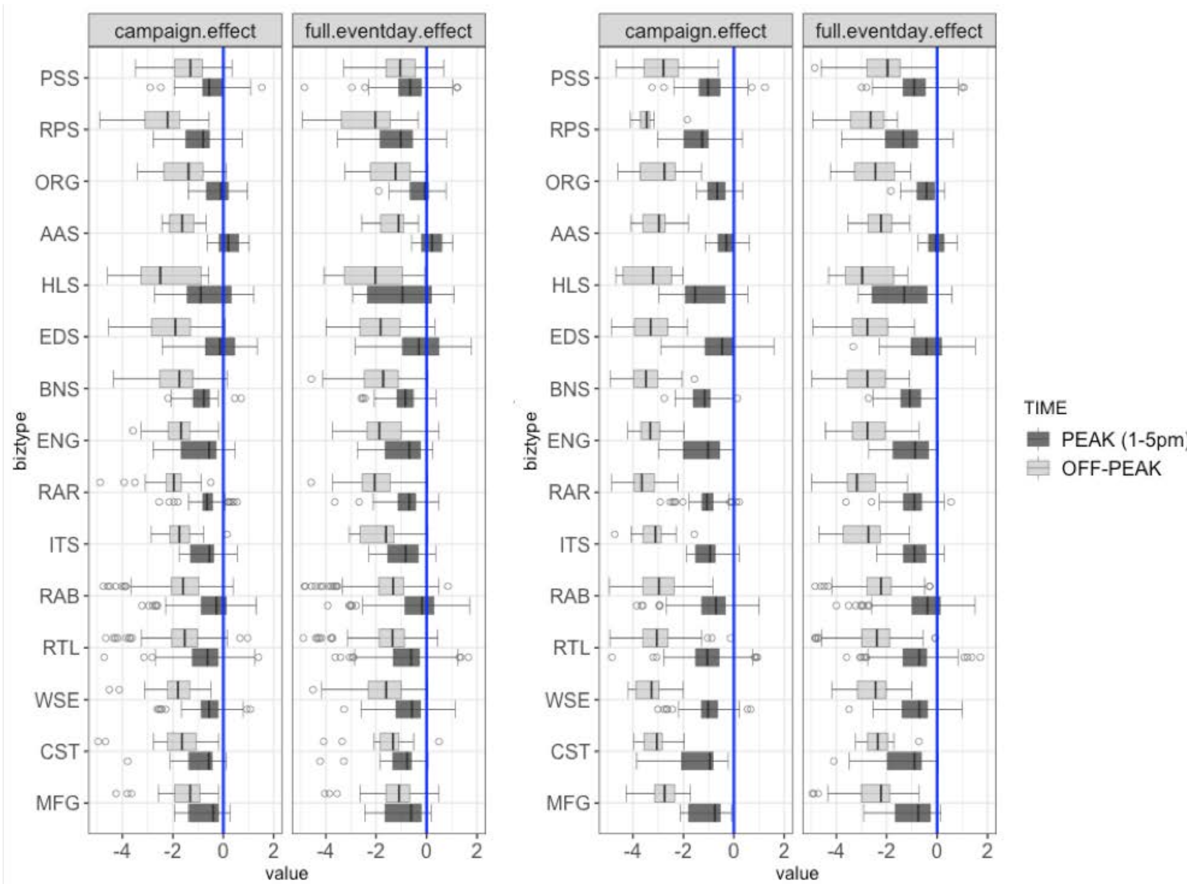
A more workable, perhaps politically palatable strategy than the individual-level recruitment approach discussed in Section 5.2 would be to segment the entire market based on business types and tune-up recruitment activities based on them. To get a glimpse of how the subjects' electricity demand response and its distribution might vary across business types, we group the individual-level predictions according to their types, which are displayed in Figure A7.

The sectoral distribution of demand response makes three points. First, seen from the median of the distribution predicted from the unrestricted model, while demand response remains smaller in peak period than in off-peak period for all business types during the campaign (*campaign.effect*), such separation is somewhat less evident on event days (*full.eventday.effect*). That is, despite not statistically significant, dynamic price events seem to provide some additional stimulus for more peak-time curtailment and associated more load shifting into the off-peak period, at least in terms of sectoral distribution.

Second, while all of the fifteen business types present negative median peak-time response on event days, their between-sector variance is not considerable. More significant is within-sector heterogeneity, which is indicated by the box-and-whiskers spanning in both directions. Note that sectors delivering business-to-business services, such as business support (BNS), engineering service (ENG), construction (CST), and small manufacturing (MFG), exhibit peak-time distribution that is bunched on the left to generate relatively robust peak-time curtailment. Except the few sectors, no other business type ensures most of its member businesses generating such robust peak-time curtailment. The suggestion is that utility firms may consider approaching these high-impact sectors first, rather than the others without robust negative demand response.

The last point, which may relieve some concerns about program performance, is that the predicted median peak-time demand response remains negative, so does the mean, during the campaign in most business types, except sectors directly dealing with customers, such as restaurant & bar (RAB), art & sports (AAS), education service (EDS), and organization (ORG). Therefore, "no regrets" second-best strategy for the program administrator would be to draw on business types exhibiting strictly negative mean peak-time demand response and to recruit their member businesses using a stratified sampling based on accessible observables, such as monthly electricity bills and fixed capacity payments.

Fig. A7. Predicted distribution of the sample's demand responses by business type in peak and off-peak periods during the campaign (*campaign.effect*) and for event days (*full.eventday.effect*) obtained from the main unrestricted model (left) and the 10th-degree polynomial model (right)



Note: The box-and-whisker plots display the lower quartile (the box's left end), the median (the inside band), and the upper quartile (right end). The ends of the whiskers indicate the lowest datum within 1.5 IQR (inter-quartile range) of the lower quartile and the highest datum within 1.5 IQR of the upper quartile.

A6. Group-level Sparsity Modeling

Our modeling exercise has assessed the effect of the PTR program for all appliance categories reported (j) in all hours-of-the-day (h) of all contribution types (ϕ). Such dense modeling, however, may create challenges regarding the interpretation and application of findings. Obviously, one would prefer to identify a smaller set of statistically significant predictors that capture main signal present in the data, facilitating the interpretation of the original model. In doing so, it would be desirable to have all hourly coefficients within the same appliance category being either zero or non-zero simultaneously. The intuition is that small businesses may prioritize a subset of appliances critical for the quality of services, implementing demand response across hours of the day during the campaign, while not paying any extra attention to the other not-worth-to respond appliances. Therefore, any approach to identifying a smaller set of predictors should acknowledge the natural "group" structure that the model itself might not necessarily represent. Yet, hourly coefficients for baseline consumption, as well as weather effects, may not be permitted to drop off because they collectively establish a benchmark to estimate the effects of demand response.

Given that our Wald-type parameter exclusion approach did not rend itself to sparser group-level modeling that is consistent across the appliances,[19] we consider an alternative shrinkage method, presupposing that only a relatively small number of structurally grouped predictors exhibit pronounced demand response effects. For the purpose of generating more sparse coefficient vectors, the group lasso is employed (Meier et al., 2008, Yuan and Lin, 2006). With regard to the main regression model in Eq. (2), the shrinkage method then solves

$$\min_{\boldsymbol{\beta}_j^{\phi} \in \mathbb{R}^{24}} \left[\frac{1}{2} \sum_{i=1}^{N} \left\{ \widehat{e^Y} - \sum_{\phi} \sum_{j} \widehat{\mathbf{e}_j^{\phi}}' \boldsymbol{\beta}_j^{\phi} \right\}^2 + \lambda \sum_{j=1}^{J} \|\boldsymbol{\beta}_j^{\phi}\|_2 \right]$$

where $\widehat{\mathbf{e}}_j^{\phi} \in \mathbb{R}^{24}$ is the vector of regressors for appliance j for contribution ϕ, $\boldsymbol{\beta}_j^{\phi}$ is the vector of group-level parameters to estimate with its euclidean norm denoted by $\|\boldsymbol{\beta}_j^{\phi}\|_2$, and λ is a penalty parameter, an optimal level of which is to be found through the cross validation process (Tibshirani et al., 2015). The trade-off is that as λ increases (decreases), the model becomes sparser (denser), fitting the data less (more) closely. In our sparsity modeling, all parameters for baseline usage $\boldsymbol{\beta_j}^{base}$ and weather effects $\boldsymbol{\beta_j}^{m}$ are left out of the shrinkable set of groups, as they jointly constitute the benchmark electricity usage.

The cross-validation procedure indicates that the mean-squared prediction error is minimized with

[19] Recall that none of the 19 appliance categories fails to reject the simultaneous exclusion for both peak and off-peak hours (Table 4).

$\lambda = 0.000148$, where none of the campaign coefficient groups are dismissed but all of the marginal event-day coefficient groups are. We instead picked the most parsimonious model within one standard error of the minimum, that is, $\lambda = 0.000720$, acknowledging that the trade-off curve itself is estimated with error (Breiman, 2017, Hastie et al., 2009, Tibshirani et al., 2015). Our group lasso procedure led to the dismissal of campaign coefficients for the three appliances (FN, WD, and TV) and of marginal event-day coefficients for all appliances. The same procedure has been applied to the 10^{th} degree polynomial model to dismiss none of campaign coefficients but eleven of marginal event-day coefficients (DD, RC, MW, OR, VC, FL, LD, CM, TV, IH, and MI). The hour-level estimates for demand response obtained by applying the group lasso to the main model and the polynomial model are displayed in Figures A10 and A11, respectively, with their baseline effect counterparts shown in Figures A8 and A9.

First of all, we find notable similarity in the overall trends between hourly coefficients estimated from the sparsity modeling of the main model and those from the polynomial model (Figures A10 and A11). This suggests that the two models are not very much different in capturing main signals from the data and in generating the structure of variables selection and coefficient shrinkage, which provides another empirical support for applying the polynomial parameter dependency to our unrestricted model. Second, with the above-mentioned coefficient groups dismissed, the procedure introduces estimation bias for nearly all selected campaign effect variables toward zero, compared to the denser cases (Figures 4 and 6). Such shrinkage seems most evident for AC and FN, indicating that, with the sparsity assumption, the cooling devices fall short of generating strong effects throughout the campaign, suggesting that cooling service for small businesses may not necessarily be used as intensively during the campaign as the denser model predicts. Third, with the sparsity assumption, the individual-level distribution of demand response on event days becomes almost indistinguishable from the demand response during the campaign (not shown). This is due to the variables selection procedure removing the marginal event-day effects for nearly all appliances. In addition, the case with the sparsity assumption, compared to the case without, introduces minor estimation bias to the campaign effect toward zero, shifting the mode of its distribution slightly to the right while in our case fitting the data within one standard error of the minimum.

Overall, our group-level sparsity modelling exercise points to the possible role that a small set of appliances, not all of them, can play in capturing the main signal present in the data *and*, as well as the strict dominance of the campaign effect over the marginal event-day effect in nearly all appliances covered. To compare the sparse model with the aggregate counterpart, Figure **??** plots mean hourly electricity usage fitted by the estimated coefficients before and after the group lasso. As shown, the two figures present no noticeable difference in the average treatment effects both on event days and on non-event days during the campaign. Caution has to be taken, however, in its implementation. Admittedly, the sparsity modeling can simplify the interpretation of the denser model and allows program developers and their customers to focus on a fewer number of demand response measures available. The downside is that, as all marginal event-day effects are not identified and many of campaign effects are lessened, the information cannot be taken into account in prescribing more customized appliance-specific demand response measures or curtailment contracts.

Fig. A8. Appliance-level estimates of hourly baseline consumption in the high, medium, and low temperature and humidity scenarios after the group lasso applied to the main model

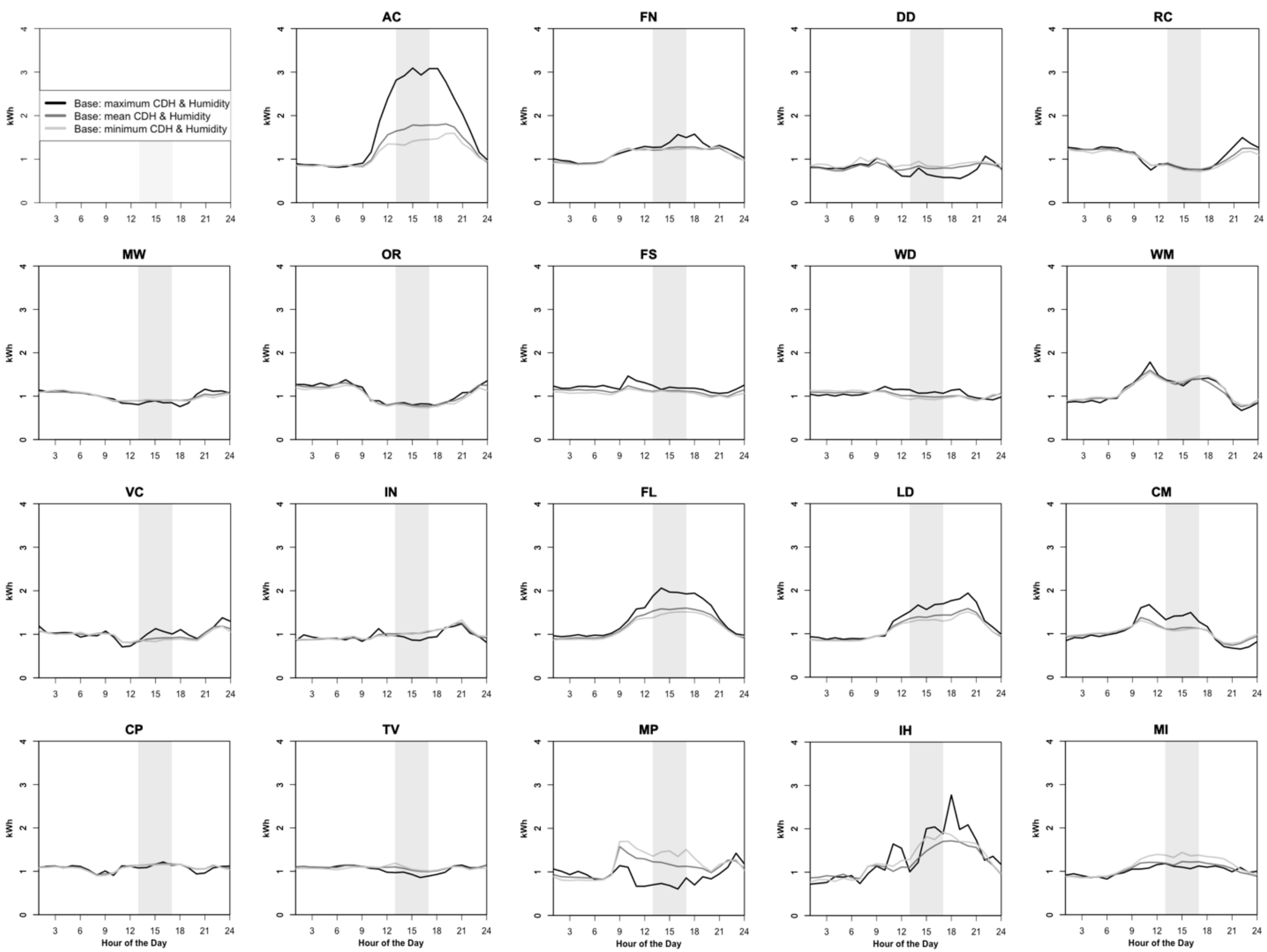


Fig. A9. Appliance-level estimates of hourly baseline consumption in the high, medium, and low temperature and humidity scenarios after the group lasso applied to the 10^{th} degree polynomial model

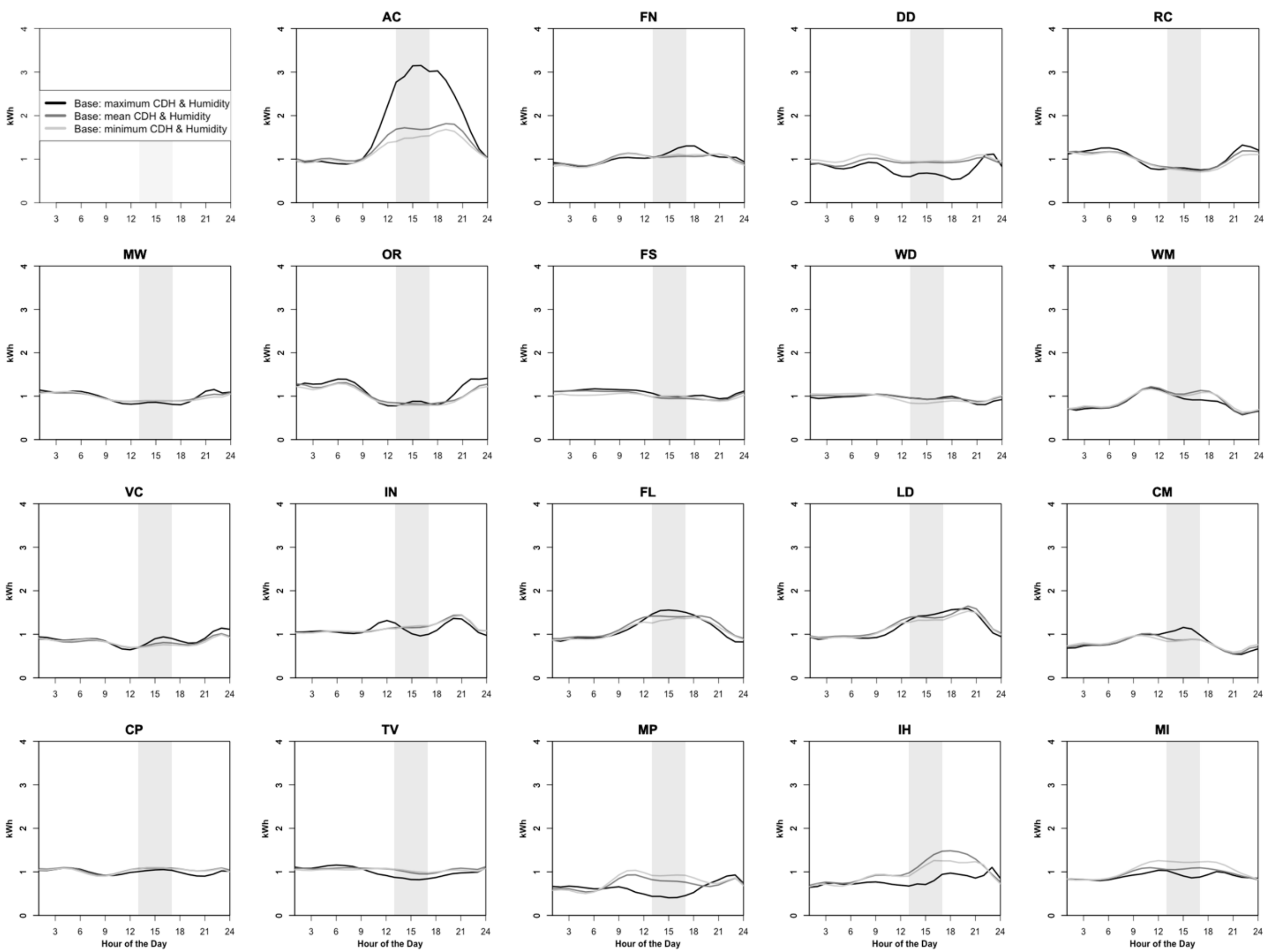


Fig. A10. Appliance-level estimates of the campaign treatment effect and the marginal event-day treatment effect after the group lasso applied to the main model

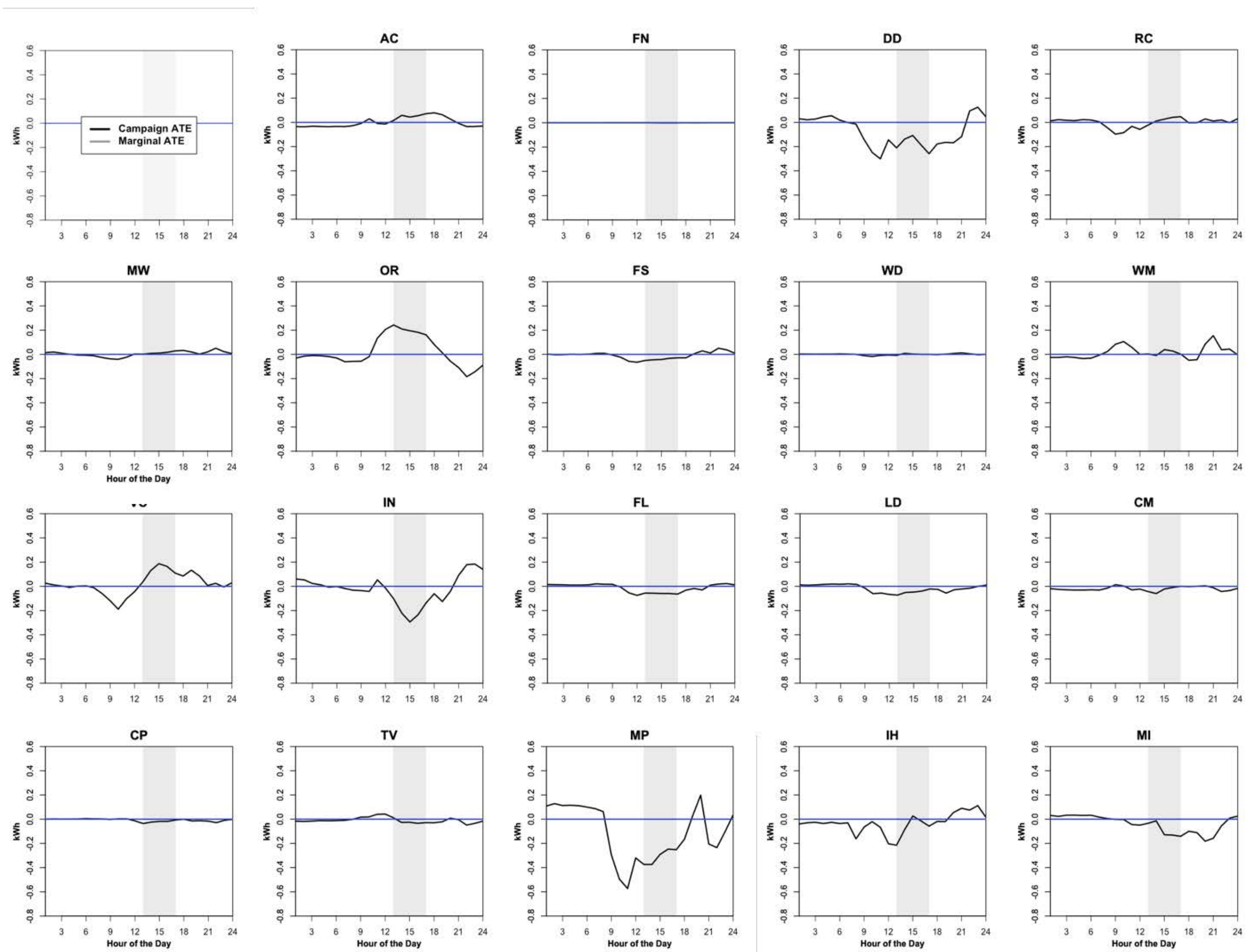
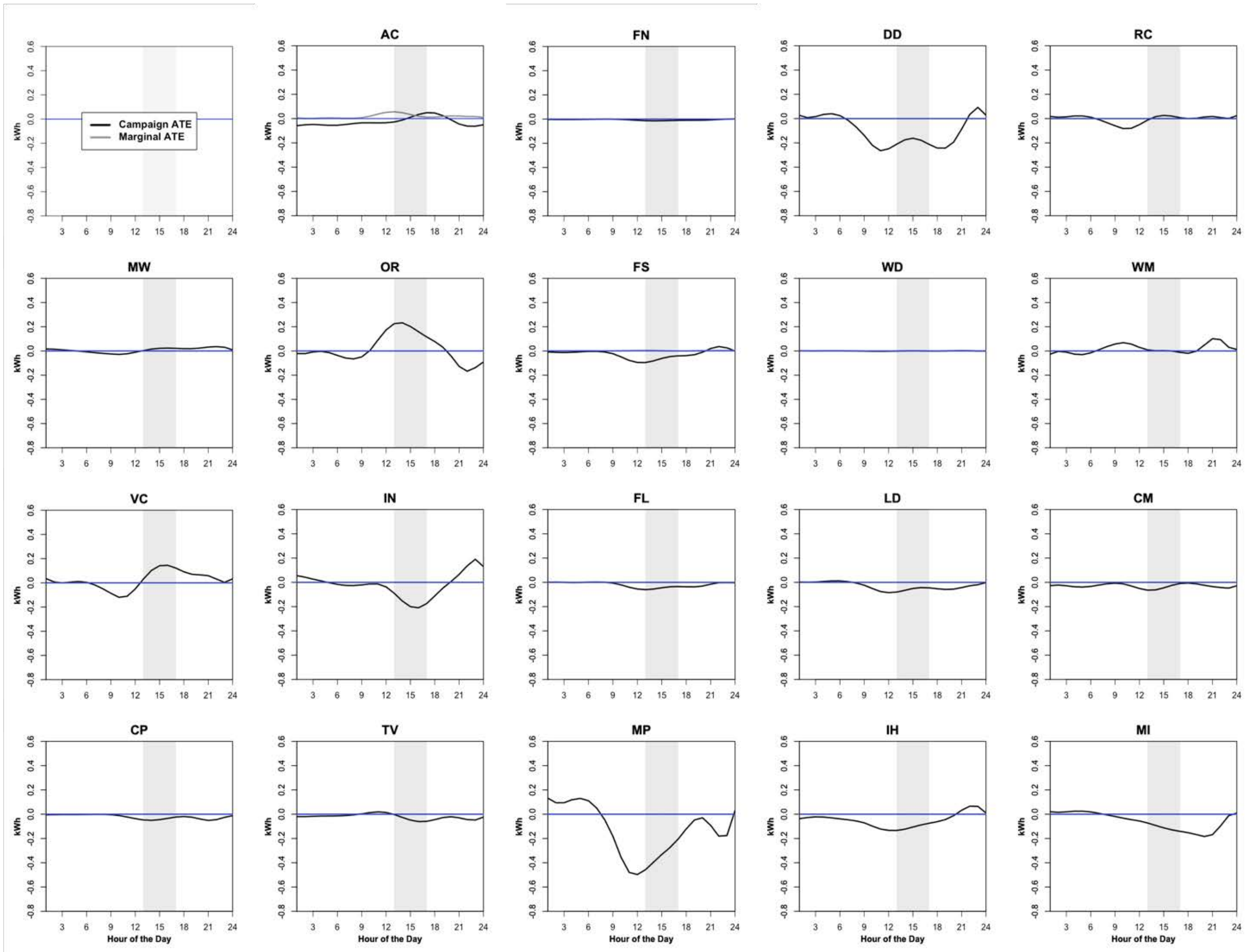


Fig. A11. Appliance-level estimates of the campaign treatment effect and the marginal event-day treatment effect after the group lasso applied to the 10th-degree polynomial model



A7. Post-Campaign Survey Responses

Fig. A12. Post-campaign survey responses from the treatment subjects on the frequency of undertaking demand response on PTR event days

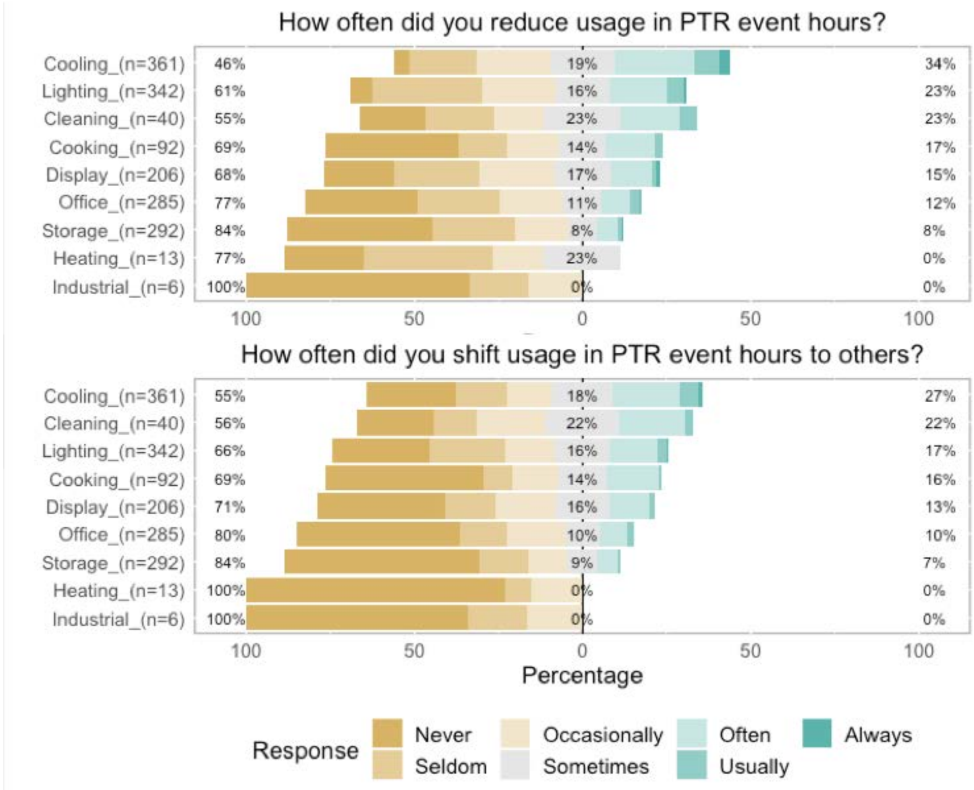


Table A2: Post-campaign survey responses from the treatment subjects on demand response measures used during the campaign and their perceived performances

	Measures implemented (multiple choices)	Measure of greatest perceived impact
Turning off unnecessary lights	392	71
Changing AC's set temperature	297	176
Pulling power cords or shutting off multiple-tap	273	82
Not using AC	190	167
Using fan in lighter wind mode	173	11
Shifting usage to off-peak hours	66	10
Using fridge in lighter cooling mode	61	0
Not using fan	45	3
Using higher efficiency appliances	36	1
Reducing frequency of using appliances	32	9
Going out leaving shops empty	24	3

Observations from the post-campaign focused group interviews, some of which also appear in the main text, are as follows:

"It was not until I joined the program that I got into the habit of pulling power cords...I did not check SMS messages except the first two. I just continued to save energy during the summer, while taking such action for granted." (A golf shop owner)

"The event notices were almost useless to me as I just continued to save electricity use since I joined the program. Whenever I went out for home visiting services, I turned off nearly all appliances not in use and pulled their power cords. I didn't do any of these before joining the program." (A hardware store owner) *"It would have been okay to me to receive the event notice on the same day as I already became accustomed to energy conservation"* (An optical store owner)

"It did not take long before I got used to energy saving lifestyle. I did not do as such before joining the program. To reduce electricity usage, for example, I pulled out one or two light bulbs at my store during the program." (A keysmith store owner)

"If at least one customer asks for air conditioning, I must turn it on no matter what." (A pan-fried rice restaurant owner)

"Customers were very sensitive to the temperature setting of drinks fridges. They complained that beverage was not cool enough and left without purchasing. I could not ignore such response." (A supermarket owner)

"It was hard to curtail electricity usage because customers visited almost anytime. Then I had to start the machine right away without exception" (A flour mill owner)

"I recall that when I received event notices, I tended to turn on the air conditioner a little later in the morning of the event days ... Or sometimes I just left the door open turning it off until customers entered in." (A herbal medicine shop owner)

"As our restaurant is open 24 hours, we leave the air conditioners on all day in the summer. During the campaign, however, I left one or two air conditioners turned off, and when customers asked, I simply turned them on." (A BBQ restaurant owner)

"Although my store wasn't particularly sweltering, when visiting customers felt uncomfortable, that was time to turn on the air conditioner." (A mobile phone shop owner)

"It is almost impossible to turn off the air conditioner as the customers are to sit down nearly two hours." (A beauty parlor owner)

"If the shop is hot, the guests simply leave, developing a bad reputation. Maintaining pleasant environment is important for us." (A barber shop owner)

"The only thing I could do was to endure heat for one or two hours or raise temperature setting. Turning off computers was not a feasible option for business like us." (An interior design office owner)