



# Real-time feedback on electricity consumption: evidence from a field experiment in Italy

Giacomo Marangoni · Massimo Tavoni

Received: 19 May 2020 / Accepted: 1 December 2020  
© The Author(s), under exclusive licence to Springer Nature B.V. part of Springer Nature 2021

**Abstract** Smart meters can help citizens in optimizing energy consumption patterns. However, mixed evidence exists on their effectiveness in reducing energy demand and especially in levelling off the daily peaks of electricity load curves. Here, we evaluate the impact of providing real-time feedback on electricity consumption from a field trial in Italy. We combine standard regressions with machine learning techniques on high-frequency data to quantify impacts on both levels and patterns of electricity use. Results indicate that real-time feedback can moderately decrease electricity consumption (between 0.5 and 1.9% depending on model specification), but that it does not promote load shifting throughout the day by itself. Machine learning reveals evidence of significant household heterogeneity in the behavioral response.

**Keywords** Energy conservation · Real-time feedback · Residential load curves · High-frequency data · Smart meters

---

G. Marangoni (✉) · M. Tavoni  
Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Milan, Italy  
e-mail: giacomo.marangoni@polimi.it

G. Marangoni · M. Tavoni  
RFF-CMCC European Institute on Economics and the Environment, Centro Euro-Mediterraneo sui Cambiamenti Climatici, Milan, Italy

## Introduction

Europe has set ambitious policy targets for energy and climate, intending to develop a sustainable, low-carbon society. Existing policy packages, such as the 2030 energy and climate strategy and the EU Green Deal, have emphasized a variety of proposed tools. These include carbon pricing, incentives to renewable and other low-carbon energy, and energy efficiency measures. These policies, pushed together with the liberalization of the EU electricity markets, are having profound repercussions on the energy industry. While our energy supply systems undergo a necessary transition toward lower carbon-emitting alternatives, power utilities are also considering the potential coming from the demand side. Utilities no longer provide electricity or gas alone but are now providers of energy services, aimed at improving customer experience.

Avoiding inefficient behaviors and appliances in households (HHs) have multiple benefits. On one hand, households can reduce their energy bills by reducing or shifting consumption. On the other, power providers can avoid potentially costly and carbon-intensive on-peak production, as well as investments in networks and plants to sustain otherwise higher future peak demands. Since members of households are often unaware of their precise consumption levels over time, one way of engaging them is via frequent information feedback on their consumption. Thanks

to the increasing deployment of smart meters, i.e., electronic devices that record consumption of electric energy sub-hourly for monitoring and billing purposes, real-time feedback is now a viable and affordable option. Smart meters are indeed being deployed at a very fast rate, with almost 200 million devices expected to be operational in the EU in 2020.<sup>1</sup> Such technology has the capability of recording and communicating energy consumption levels at a high temporal resolution (e.g., 15 min or less), thus enabling consumers to take control of their energy bills.

Policy makers have promoted information programs, claiming their high potential and cost-effectiveness for energy conservation. The rationale for this policy intervention is that users are often unaware of their energy footprint, either because of the high costs of acquiring this information or because of behavioral factors such as limited attention, present bias, and limited salience (Allcott and Mullainathan 2010). However, critics remain due to mixed empirical findings.

This paper applies for the first time a combination of econometric and machine learning models to a large-scale dataset of high-frequency residential power consumption readings recently collected by a major energy utility in Italy. The goal is to provide new empirical evidence on the still debated extent to which real-time feedback of residential power consumption can promote energy conservation behaviors. We improve on previous studies in two ways. First, we offer novel estimates of the energy conservation effect and its persistence over time, testing for alternative model formulations based on subsamples varying in size, information set, and time resolution. Second, we exploit the high-frequency data to study how real-time feedback impacted daily power consumption not only in absolute levels but also in their distribution throughout the day. Our study eventually contributes to a developing literature set to inform cost-effective behavioral interventions for improving residential energy savings, with lessons possibly extensible to contexts beyond the Italian one they originated from.

## Literature review

People's perceptions of energy consumption can have significant discrepancies with reality (Attari et al.

2010). In particular, low-energy activities tend to be overestimated and high-energy activities tend to be largely underestimated. This justifies the exploration of behavioral interventions to promote lower energy use and more environmentally friendly decisions. So far, the majority of studies has provided strong empirical support for social comparison interventions (Allcott 2011). Evidence for other types of interventions is still mixed (Andor and Fels 2018), including for information feedback type of programs which is the focus of this paper.

Several field experiments have been conducted to understand the causal impact of such information programs. On one hand, information feedback seems to strengthen the price elasticity of demand, making interventions like dynamic pricing more effective (Jesso and Rapson 2014). On the other, it supposedly induces electricity savings (Schleich et al. 2013; Attari et al. 2014; Lynham et al. 2016). This result is driven both by the process of learning about the energy consumption of different activities as well as of reminding of one's energy use, with possibly the former being a stronger driver (Lynham et al. 2016). Regarding relative average consumption reductions, two papers performed a meta-analysis of the many available studies (Karlin et al. 2015; McKerracher and Torriti 2013). We refer the readers to the references therein for exhaustive coverage of the literature on the topic.

Karlin et al. (2015) consider 42 studies with an effect size on power consumption ranging from  $-8$  to above 20%. Their meta-model estimates an overall reduction between 4 and 12%, depending on different aggregation schemes. Several treatment variables moderate this relationship, including frequency of feedback, medium, comparison message, duration, and combination with other interventions (e.g., goal, incentive). Feedback turns out to be more effective especially when combined with goal-setting or external incentive interventions, and is somewhat brief (e.g., less than 3 months) or quite long (e.g., longer than 1 year). McKerracher and Torriti (2013) come up with more conservative estimates in their meta-analysis. Especially focusing on the lower estimates of more recent studies with larger sample sizes and more representative sample selection and recruitment methods, the authors argue that a realistic, large-scale conservation effect from feedback is in the range of 3–5%. These estimates are significantly lower than what

<sup>1</sup> See <https://ses.jrc.ec.europa.eu/publications/reports/benchmarking-smart-metering-deployment-eu-27-focus-electricity>

are suggested in earlier studies but are supposed to be more robust, given the very large sample sizes of recent trials. It should be remarked that even if relative reductions do not appear to be large, they are in fact considered very significant from the industry viewpoint. If these results can be generalized, given the large share of energy consumed in the residential and commercial sectors, they can go a long way in improving energy efficiency, especially in the light of the very low price elasticities observed in residential energy. Among available types of feedback, real-time household-level feedback seems to be the most effective in overcoming imperfect information and inattention biases, with potentially large conservation effects (Tiefenbeck et al. 2016).

As technology becomes cheaper and consumption data more readily available, interest in high-frequency feedback has been rising. Houde et al. (2013) take advantage of hourly data from a long-lasting real-time feedback field experiment and also look into time-of-day reduction effects and persistence of the effect over time. In this case, access to feedback leads to an average reduction in household electricity consumption of 5.7%, persisting for up to 4 weeks. Another experiment combined in-home displays for real-time feedback with interventions involving moral suasion and economic incentives (Ito et al. 2018). Moral suasion led to 8% short-run power usage reductions, while economic incentives doubled reductions and made them more persistent, supporting the synergy between information and dynamic pricing.

Despite the magnitude of the consumption savings, a recurring finding is high heterogeneity in the impact estimates. Apart from differences in experimental setups, both socio-demographic and psychological variables can explain the variety in consumption levels and response to information feedback (Abrahamse and Steg 2009). Nonetheless, often the relationship between these variables and treatment effects is not evident (Houde et al. 2013), and larger studies are still required.

Availability of hourly consumption data can provide further insights on how people behave. Machine learning, and in particular clustering, has been applied in the past to understand patterns of consumption better. Although several approaches are documented (Wang et al. 2015), some are of more immediate application and interpretation, like the one by Kwac et al. (2014), focusing on the direct clustering of normalized

load shapes. These methods have mostly been used in the context of static load profiling rather than for detecting behavioral changes.

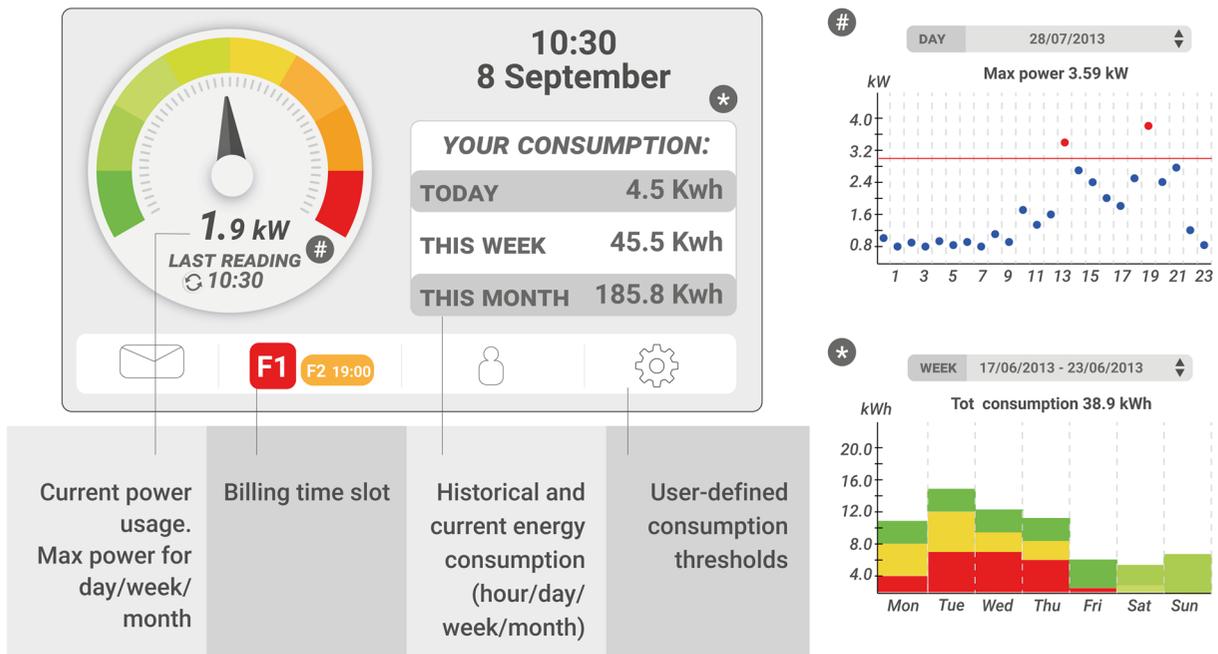
This work combines the classical regression approach, as in Houde et al. (2013), with a machine learning approach, as in Kwac et al. (2014), to understand better if and how power consumption behavior changed concurrently with the installation of in-home displays providing real-time feedback in a large sample of Italian households.

Only one other peer-reviewed article uses this same dataset (Fontana et al. 2019), but with important differences. Their identification strategy relies only on a before–after comparison, while here we exploit the gradual phase-in design to disentangle the effect from an already decreasing energy consumption trend. This could justify the fact that Fontana et al. (2019) find a higher average conservation effect. Furthermore, the authors focus on correlating the effect on hourly power consumption to appliances ownership, rather than considering aggregate behavior like we do here. They find no statistically significant evidence for such correlation, which might be consistent with our inability as well of detecting clear sub-daily load shifts.

### The in-home display

In 2011, a large European power utility company started a 3-year-long pilot project in the area of Isernia (mid-south Italy) to test new smart-grid-related technologies and inform future network restructuring plans. One part of the project dealt with customer engagement for demand response. In this context, a kit was distributed to thousands of end-users to enable active participation by making people aware of how much electricity they were consuming.

The main interaction with the kit occurs via a display installed in the house, informing users about instantaneous consumption, as well as daily, weekly, and monthly summaries (Fig. 1). The display also provides information about the current billing slot (days are split into 3 billing periods: peak (F1), intermediate (F2), and off-peak (F3)) and the time at which the next slot will enter into force. If users enter information about their billing tariffs, they also get feedback on monetary expenditures. Users can set goals, and are also warned by an acoustic signal when their power consumption exceeds the contractual obligation (set at



**Fig. 1** Features of the in-home display for real-time power consumption information feedback, distributed to households in the province of Isernia

3 kW for most customers).

The display was distributed to residents of Isernia City and surrounding municipalities for free. The company focused on this area for technical reasons, related to the feasibility of high-frequency consumption data measurement and transmission. The display was not randomly allocated: it was first distributed outside the city, and subsequently in the city. The company advertised this opportunity through media campaigns. They faced initial difficulties in recruiting enough volunteers, possibly for concerns about privacy or simply lack of advertisement. They subsequently intensified the promotional campaign, by hosting meetings within local communities, in schools and other public spaces. This had the effect of increasing participation quite rapidly.

Overall, the trial was not carried out according to the golden rules of randomized controlled trials. Hence, the design is subject to possible self-selection, a trait common to all voluntary-based programs. Furthermore, no data were collected on consumers outside the program, which could have been used to miti-

gate the selection bias, e.g., by balancing untreated and treated groups via propensity score matching. For treatment estimation, we compensate for the lack of an untreated group by exploiting the gradual phase-in design of the pilot, as done in similar nonrandomized observational studies. In particular, we test for potential imbalances between those who receive the display earlier and those who receive it later, and control for biasing factors in those model specifications where these factors appear explicitly. We also repeat the analysis focusing on the central time periods of the experiment, where the balance between untreated and treated is potentially greater (see “[Regression of daily electricity consumption](#)”). In terms of external validity, we can only provide a comparison between the studied sample and available historical statistics to have a sense of how dissimilar our sample is from the population. Despite its limits, the trial still provides useful information on energy consumption behavior, it is the first one carried out in Italy on such a large scale, and it can still inform expectations on similarly designed future campaigns by energy utilities.

## Methods

The kit, containing the in-home display used for consumption feedback testing, was distributed to thousands of households in the province of Isernia in Italy. Recruitment occurred voluntarily and was supported through several channels. Between June and September 2012, the kit was promoted with informational days at schools, mass marketing, and collaborations with public authorities and institutions. The official recruitment started in November 2012, while informal tests were running since August. Among those who adhered to the initiative, only a sub-sample, which will be called the “Client” sample, had the in-home display still active at the end of the test period, namely December 2014. This sample excludes non-domestic or non-resident customers, as well as those with a power contract other than 3 kW or 4.5 kW, as patterns of consumption may be very different in these cases. For each household in the “Client” sample, the utility provided data on monthly energy consumption between January 2012 and December 2014, split by billing time slot.<sup>2</sup> Extra available information includes contractual power and municipality at the moment of joining the program, as well as date of delivery and version of the display. With “Survey” sample, we denote a subset of the “Client” households that agreed to provide also information on the demographics of family members, the number of appliances available in the house, and some characteristics of the dwelling.

For another subset of the “Client” households, which will be referred to as the “Load” sample, it was possible to obtain higher frequency readings of energy consumed, i.e., every 15 min, at least for a fraction of the full 3-year span. The collected load curves are re-sampled to 1-h time steps for this analysis. To filter outliers on the higher end, we assume that electricity can be withdrawn at most with power exceeding 10% the contractual value. Hourly readings exceeding 3.3 kWh or 4.95 kWh are thus removed for 3-kW or 4.5-kW contracts households respectively. These thresholds

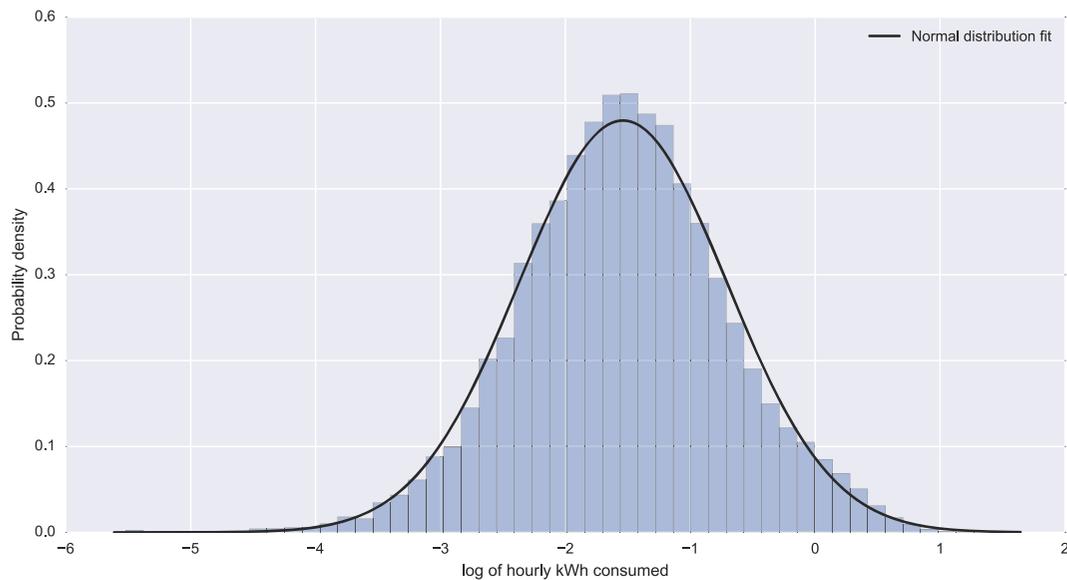
represent the same power levels above which service would generally be discontinued after a while. Still, only a tiny fraction of observations exceed these extremes, corresponding to around 99.999% quantiles of their respective datasets. A slightly more restrictive threshold is assumed on the lower end, removing data below the 0.1% quantile (i.e., 4 Wh). This should filter very low readings which might correspond to either faulty sensors, blackouts, or empty houses, i.e., not interesting cases for the analysis. Gaps up to 2 h of missing data are interpolated linearly from available values. Days remaining with missing points after this interpolation are removed. For monthly and daily data, we consider as outliers and exclude observations with power consumption below the 5th percentile or above the 95th percentile, interpolate linearly over time only when a single central reading is missing among 3 consecutive ones, and discard HHs left with less than 2 years of monthly data or 60 days of daily data.

There is no clear best sample among the three for our impact evaluation purposes, since those with higher number of observations have also a lower number or different set of variables available. Hence we decided to include all of the three in the analysis that follows. Table 1 compares some key variables from the “Client”, “Survey” and “Load” samples with 2012 statistics at the province (“Isernia”), region (“Molise”) and national (“Italia”) level (see Table S1 in the Appendix for further summary statistics). The “Client” and “Load” samples could be compared only on the basis of 2012 average yearly consumption due to availability of data. Households who joined and stick to the experiment seem to have consumed much more than the average family in Isernia in 2012, both when considering the estimate from ISTAT of 1973 kWh and the one from TERNA of 2248 kWh (not reported in the table). Both average and standard deviation of consumption across households do not change much between the sub-samples considered. When looking at survey data, we observe that the program involved families with an above-average number of members and rooms in the house. Ownership of washers, dryers, dishwashers, electric boilers is also above the regional average, while the presence of air conditioning is more marginal. This also holds when considering the intersection of the “Survey” and “Load” samples. Thus, it appears that the sample of our analysis is not fully representative, pointing to selection bias. The bias was not apparently

<sup>2</sup>time of use is classified into 3 categories: F1 (on-peak), from Monday to Friday 8am–7pm, national holidays excluded; F2 (intermediate), from Monday to Friday 7am–8am and 7pm–11pm, plus Saturday 7am–11pm; F3 (off-peak), from Monday to Saturday 11pm–7am, plus Sunday and national holidays. Depending on the contract, different time slots may have different prices, with the most popular billing having a higher price for F1 and a lower price for both F2 and F3.

**Table 1** Comparison of variables between different sub-samples from this study and official historical statistics for year 2012

	Client	Curves	Survey	Survey + Curves	Isernia	Molise	Italia
Number of HHs	1834	966	804	439	36563	131059	25872613
Avg. Domestic Power Consumption in 2012 [kWh/HH]	2814 (1103)	2715 (1046)	2831 (1076)	2772 (1069)	1973		2299
Avg. Domestic Power Consumption in 2012 [kWh/Capital]			960 (515)	992 (574)	954		1186
Avg. Number of Family Members			3.3 (1.2)	3.2 (1.2)	2.4	2.4	2.3
% Females in Families			49.1 (22.0)	49.4 (23.3)	51.0	51.2	51.5
Avg. Age			42.1 (15.6)	44.4 (16.6)		44.9	43.8
Avg. Number of Rooms per HH			5.3 (0.9)	5.3 (0.9)	4.6	4.5	4.2
% HHs w/ Fridge			99.4	98.9		99.8	99.6
% HHs w/ Washer			99.0	98.9		98.1	96.2
% HHs w/ Dishwasher			62.3	61.7		35.9	39.3
% HHs w/ Freezer			65.4	64.5		25.8	25.3
% HHs w/ Dryer			12.4	11.4		3.0	3.3
% HHs w/ Electric Boiler			13.8	14.4		8.6	13.6
% HHs w/ Air Conditioning			18.9	18.7		23.5	28.0



**Fig. 2** Distribution of hourly power consumption is well approximated by a log-normal distribution

driven by higher per capita energy consumption: this matches quite well the official statistics of the Isernia municipality. However, families participating in the trial are more numerous, with 3.2 persons per family as opposed to the population average of 2.4. This might be attributed to the initial school campaigns, which were the most successful in getting the program going according to the utility.

Hourly energy consumption computed from the “Curves” dataset appears to have a log-normal distribution (see Fig. 2), in agreement with the literature (Kwac et al. 2014). It includes 966 households and 26,304 time periods, for a total of 14,805,000 non-null observations (i.e.,  $\sim 60\%$  of all client-hour combinations).

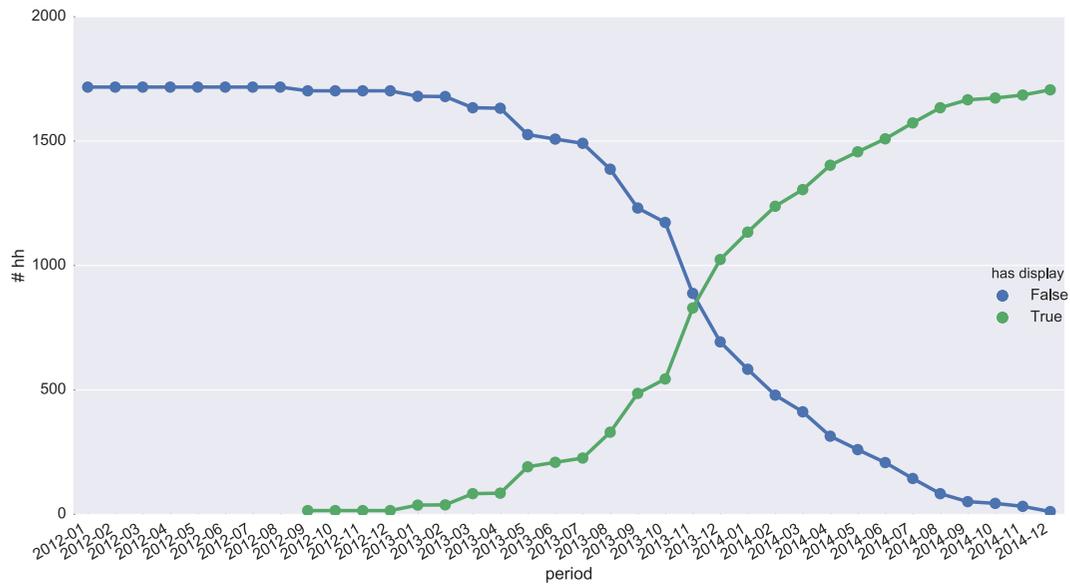
#### Regression of daily electricity consumption

First, we quantify the average power conservation effect of having an in-home display. If we were to look only at conditional averages of consumption between those with and without an in-home display over time, we would risk to attribute to the display the merit of an already decreasing trend in demand ( $\sim 3\%$  reduction per year in 2013 and 2014, based on ISTAT and TERNA data). Hence, it is important to rely on some other identification strategies.

Impact is usually evaluated against a counterfactual consumption; in this context, the hypothetical one of those who already received the display if they had not received it. Ideally, a control group is sampled to provide such counterfactual. Since we do not have access to the latter, we exploited as an alternative the gradual phase-in of the experiment: we build the counterfactual on the basis of the consumption of those who have not received the display yet at any point in time. Figure 3 shows how households progressively received the display. Although our identification strategy is not ideal, recent research seems to indicate that high-frequency data can be used to estimate causal effects in non-experimental research designs.<sup>3</sup> This identification strategy works as long as the in-home display delivery date is plausibly random with respect to HH characteristics. Evidence supporting this hypothesis is documented at the end of this section.

We model the impact evaluation problem with an ordinary least square (OLS) regression, as done previously in the literature in similar experimental settings

<sup>3</sup>See for example, David Rapson work entitled “Can high-frequency data and non-experimental research designs recover causal effects? Validation using an electricity usage experiment,” with Katrina Jessoe and Douglas Miller, presented at 2015 AERE (<http://aere.org/summer/documents/AERESummerConference2015Program.pdf>)



**Fig. 3** Number of observed clients with and without display at each month in the 3 years of the experiment

(Houde et al. 2013). We use panel data given that we have repeated measurements of power consumption for each household over time. We propose six model specifications to explain the response of daily power consumption to having received an in-home display (Table 2). All models share the same structure:

$$y_{i,t} = \alpha D_{i,t} + \sum_j \beta_j x_{j,i,t} + \epsilon_{i,t} \quad (1)$$

where  $D$  indicates the presence of the display at time  $t$  in household  $i$ ,  $x_j$  is one of the other independent variables, possibly dependent only on  $i$  or  $t$ , or both,  $\alpha$  and  $\beta_j$  are the corresponding regression coefficients, and  $\epsilon$  is the error term. Unless stated otherwise,  $y$  is the log of daily power consumption, obtained dividing monthly levels by the number of days in a month, or resampling high-frequency data when available. To better explore the different information sets available for the three different samples, and to test the robustness of results to different timescales, subsamples, and model structures, we consider multiple models.

The first two models are fit to the “Client” sample, which is the largest of the three and is balanced over time, although limited to monthly observations. Given the very few observables available other than consumption, we capture the time-independent but HH-dependent unobservable factors driving power demand with HH fixed effects. The difference in the

two models is how they capture time-dependent but HH-independent factors: either with time fixed effects (at monthly resolution, from Jan 2012 to Dec 2014) or with a 2-degree polynomial of average temperature plus monthly and yearly dummies. Historical temperature statistics are computed from local weather stations records, which cover only part of the full-time span of the experiment and hence slightly limit the number of observations.

The second two models focus on the “Load” sample, which covers fewer households than the “Client” sample but introduces high-frequency measurements. The main difference with the first two models is that time-dependent variables have a daily resolution. The daily fixed effect model spans a period from 1st Jan 2012 to 31st Dec 2014. When not including time fixed effects, we add to the time-dependent variables above also day-of-the-week dummies.

The fifth model applies to the “Survey” sample, which is also smaller than the “Client” sample but allows for correlating power consumption with HH surveyed characteristics. We replace the HH-fixed effects with these variables, which include family size, socio-demographics, appliance ownership, and number of rooms in dwelling.

In the last model, we change the dependent variable: the regression of daily consumption is set up as in model (2), but it excludes nights and weekends

(i.e., in the F1 billed time slot). This is when most of the energy-intensive household activity tends to take place.

The choice of these models reflects the varied nature of the dataset. Models (1), (2), and (6) cover the highest number of HHs, but they have the least amount of covariates and coarsest time resolution. Models (3), (4), and (5) cover less HHs, but provide further details on sub-monthly consumption and on the peculiarity of each HH. Models (2) and (4) offer plausible alternatives on the role of time in affecting consumption, modelled either via weather effects combined with seasonality dummies, or via time fixed effects. The former is more physically grounded, while the latter is not subject to the limits in quality and availability of weather data. Lastly, model (6) taps into possible on-peak effects, versus the overall effect on which the other models focus. Given all these trade-offs, a single-model assessment risks missing the nuances of our dataset and to provide a less robust assessment.

For each model, we run a separate OLS regression. These regressions imply a comparison at each time period between the sets of households with and without an in-home display. If the characteristics of these two sets of households are different, the regression may be biased. We control for this bias in several ways. First, we find that HH power consumption in 2012, i.e., one distinctive baseline characteristic available for all HHs, is not predictive of when a HH receives the display (see Table S2 in the Appendix for the results of the related OLS). Second, we focus on the survey sample to test whether other HH characteristics affect the probability of receiving the display earlier. We formulate the problem as an OLS, where the independent variables include all of the surveyed information, and the dependent variable is equal to one if a HH received the display before the average delivery date, zero otherwise. Only having a rechargeable console has a statistically significant but marginal effect (see Table S3 in the Appendix). Should this

**Table 2** Models used for estimating the impact of having a display on daily power consumption levels

Model name	(1) CLIENT	(2) CLIENT Month FE	(3) LOAD	(4) LOAD Day FE	(5) SURVEY	(6) CLIENT F1
Sample						
Client	x	x				x
Survey					x	
Curves			x	x		
Dependent variable						
log(daily consumption from monthly data)	x	x			x	
log(daily on-peak consumption from monthly data)						x
log(daily consumption from high-freq data)			x	x		
% change in daily consumption from last year (from monthly data)						
Independent variables						
Display received	x	x	x	x	x	x
Municipality					x	
Year dummies	x		x		x	
Month dummies	x		x		x	
Weekday dummies			x			
Avg. Temp. + (Avg. Temp.) <sup>2</sup>	x		x			
HH fixed effects	x	x	x	x		x
Month/Year fixed effects		x				x
Day/Month/Year fixed effects				x		
Survey variables					x	

affect the impact evaluation, we take into explicit account this and all the other surveyed factors in model (5), which focuses on the survey sample. Third, we rerun the regressions focusing only on central observations (in the period from 2013-05 to 2014-04, see Fig. S3 in the Appendix), where both baseline consumption and survey characteristics are potentially more balanced (see Figs. S1 and S2 in the Appendix), but also where the number of observations is considerably reduced. Measurements within the same HH give rise to concerns for serial correlation: we compute intervals of confidence for the coefficients with a cluster-robust covariance estimator, treating each individual as a cluster, and present both confidence intervals for the main models. A question which recurs in this literature is how long the effect of feedback lasts over time. To further explore this point, we implement four additional OLS models with fixed effects for both HHs and time periods. Here, the binary variable controlling for having received the display is replaced with several dummies, controlling whether the display has been received within the last month, two months ago, and so on. We fit this model to the three samples introduced in this study, and consider also the case excluding nights and weekends while keeping the wider “client” sample (F1 model).

### Clustering of load shapes

High-frequency data can be used to assess not only changes in overall electricity consumption, but also shifts in consumption throughout the day. Reallocating electricity from high-peak periods to off-peak ones can yield significant economic and environmental benefits, especially in countries—such as Italy—characterized by high penetration of intermittent renewable energy sources. We normalize load curves with hourly resolution by the total daily consumption. Such load shapes represent how households allocate their daily budget of consumption over hours.

A possible quantitative model to describe the consumption behavior of one HH is the distribution of recurring patterns of its daily load. The problem is that no daily load of an occupied building will present itself twice with the exact same smart meter data. A common practice in this field is to classify load curves into a small set of clusters (Zhou et al. 2013; Rhodes et al. 2014; Haben et al. 2015; Kwac et al. 2014). By

grouping together daily curves that are not too dissimilar, we end up with a limited number of profiles that repeat themselves within the lifestyle of one HH, and within the pool of HHs considered. Once we identify these prototypical profiles, we can build a discrete distribution of their frequency, and detect behavioral changes by checking for meaningful deviations in these distributions before and after an intervention. We focus on normalized demand because it is more representative of the repeated behaviors, and hence of the lifestyle, of a HH (Kwac et al. 2014), rather than absolute levels which were instead the focus of our regression analysis. Here, we assume that HHs change how they allocate consumption over the day only in response to the in-home display intervention.

To cluster the daily load shapes collected in the “Load” sample, we apply a non-supervised machine learning technique called K-means clustering (MacQueen 1967; Hastie et al. 2009). K-Means, one of the most popular non-hierarchical statistical clustering algorithm, is used here to find representative load shapes (or centroids). These represent prototypical daily consumption patterns. The algorithm identifies those sets of centroids that minimize a certain measure called inertia. Inertia is the sum over clusters of the sum of squared distances from a centroid to all members of its cluster. Iteratively, the algorithm improves the position of the centroids so that distances to nearby points decrease. While K-means always converge, it may converge to a local minimum, potentially depending on the initialization of the centroids. To alleviate this problem, the computation is done several times (namely 36), with a re-initialization scheme called “k-means++,” with provably better results than random initialization (Arthur and Vassilvitskii 2007).

K-Means needs an exogenously provided number of clusters to operate. We focus on a range between 8 and 20, spanning reasonable and manageable numbers of prototypical load shapes, already used in the literature (Räsänen and Kolehmainen 2009). We look at how inertia improves by increasing the number of clusters, and search for a discontinuity in the derivative of such relationship.

After having identified the centroids, frequencies of occurrence of centroids are calculated for each household, distinguishing the days before and after the arrival of the in-home display. Seventy days of data are required to build such frequency vectors. If any

behavioral change happened in the patterns of daily consumption due to the in-home display, this should be reflected in the before–after difference of such vectors.

## Results

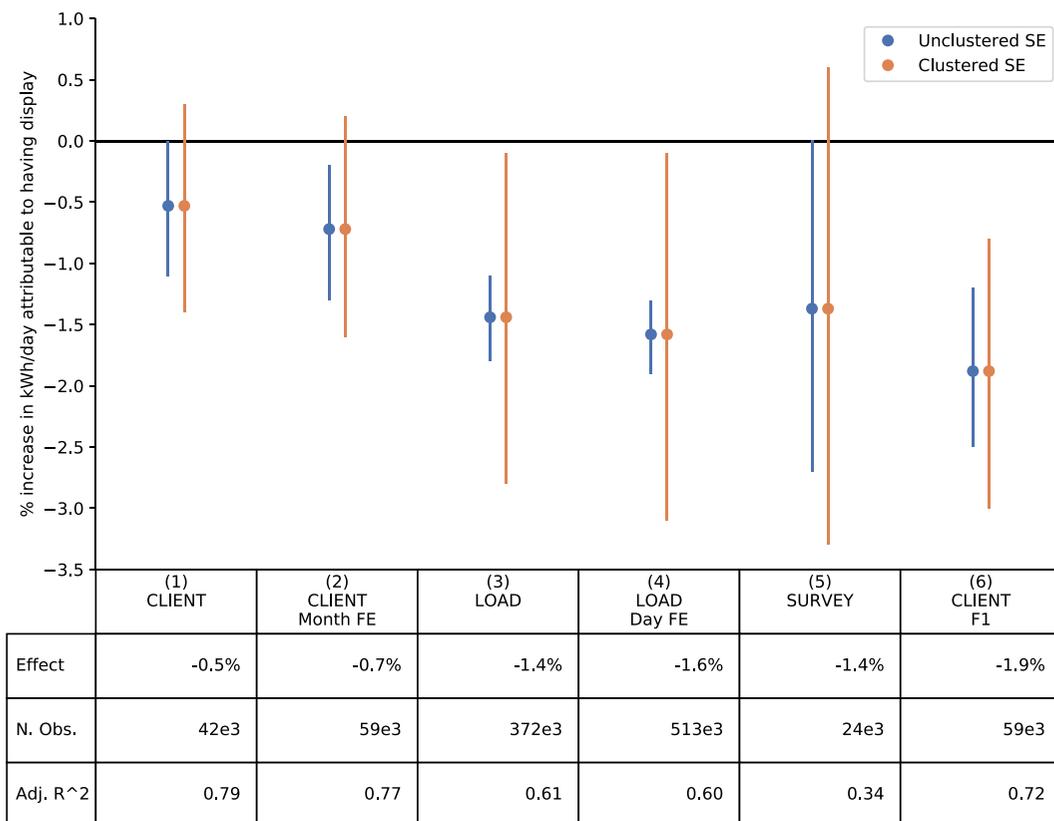
### Regression of daily consumption

The regression coefficient of primary interest is the one related to having received a display (i.e.,  $\alpha$  in Eq. 1). This coefficient represents the average percentage increase in consumption due to the presence of an in-home display (Fig. 4).

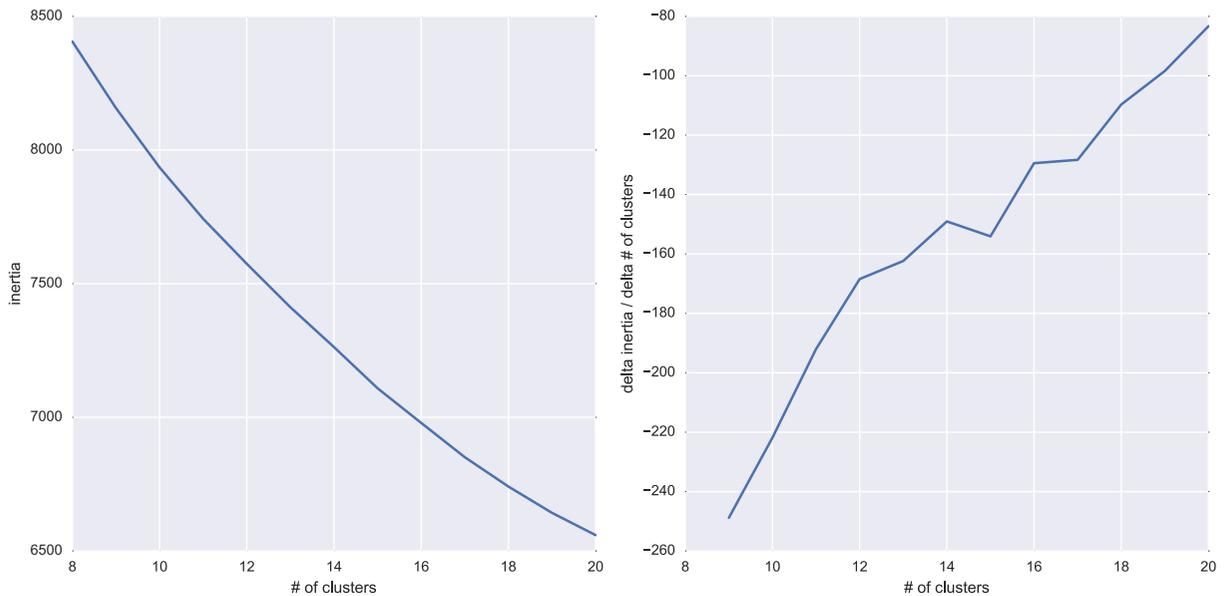
The estimated effect has the expected direction and the reduction in consumption ranges between 0.5 and 1.9%, depending on the model. The effect is smaller in models fitting the wider “Client” sample, and larger in those fitting the “Load” sample, except when focusing

on the on-peak hours. In the models based on high-frequency data, uncertainty ranges shrunk considerably thanks to the abundance of observations with a non-robust covariance estimator. Confidence intervals widen when clustering standard errors, suggesting large serial correlations between observations of the same HH. Further regression results are reported in the Appendix (Tables S4–S9). If we restrict our analysis to observations in the central period, we obtain comparable levels of power reduction attributable to the display, with a decrease in statistical significance for most models (see Fig. S3 in the Appendix).

Figure 5 shows the estimated effect of having received the display for different lengths of time. We do not find a clear decreasing trend in effectiveness of the display as months pass by after the arrival of the in-home display. An average declining trend in power demand reduction is observable in the first months, but the trend is not statistically significant, and seems to be reverted when considering longer time frames.



**Fig. 4** Percentage increase in daily power consumption attributed to having received the in-home display



**Fig. 5** Effect of the in-home display over time. All models have fixed effects for households and time periods, as well as variables controlling for having received the display within the last month (0–1 months), more than 1 month ago but less than 2

(1–2 months), and so on. Values plotted refer to the estimated coefficients of these variables. The first three models are fitted to the three different samples mentioned. The last one excludes consumption in nights and weekends

### Clustering of load shapes

Figure 6 shows the improvement of clustering accuracy, defined in terms of inertia (see “Clustering of load shapes”), as the number of clusters increase. We pick 14 as the preferred number of centroids for the following analyses, in correspondence with the first local maximum of the derivative.

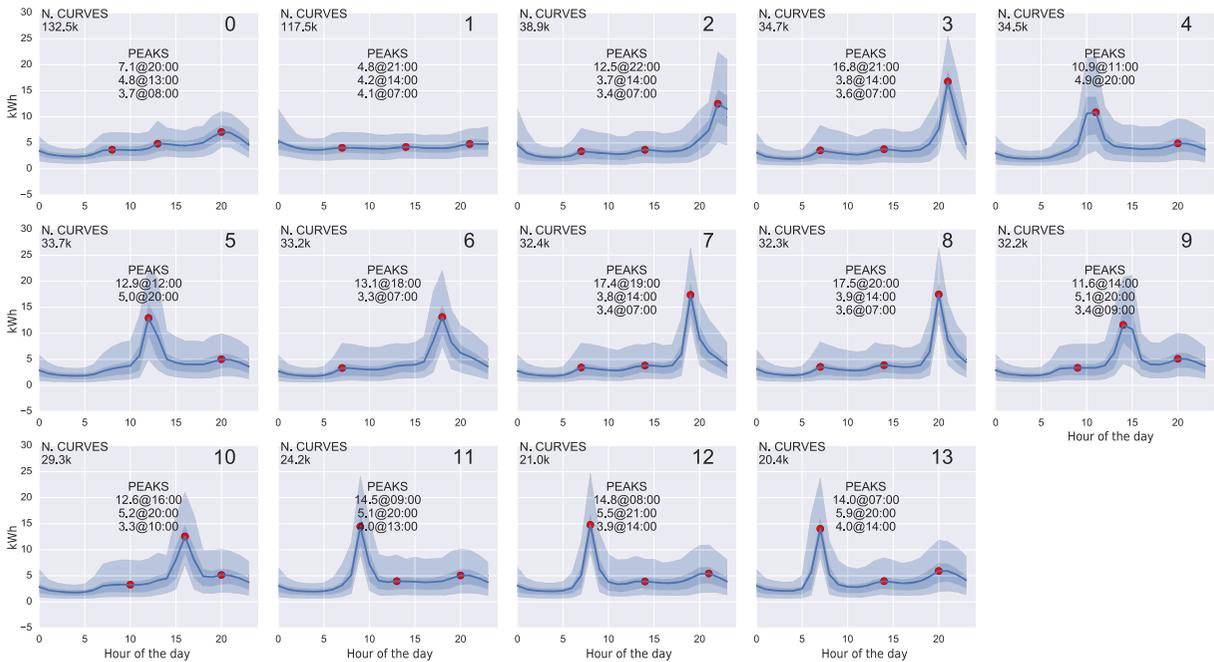
The 14 centroids, or prototypical days, are shown in Fig. 7. Clusters are sorted by the number of actual load shapes they represent (see N. CURVES within subplots). Shapes are clearly separated by the characteristics of their peaks, i.e., their relative height and hour of occurrence.

For each HH, frequencies of occurrence of centroids are calculated for the days before and after the arrival of the in-home display. If there was a common trend in behavioral change due to the in-home display, we should see a significant change in such frequencies, moving away from some representative shape toward other ones. As shown in Fig. 8, this is not emerging from the data. The average behavior of the sampled clients, as coded in these vectors of frequencies, does not change significantly conditionally on having or not a display.

Nonetheless, if we plot the change in frequencies before and after the delivery of the in-home display for each client, a wide heterogeneity emerges (Fig. 9). Most households seem to have maintained stable consumption patterns over time, as indicated by the mass of the distributions of frequency gravitating around 0. Still, several households exhibit much more flexibility in consumption, as shown by the long tails of some of the 14 distributions.

### Discussion

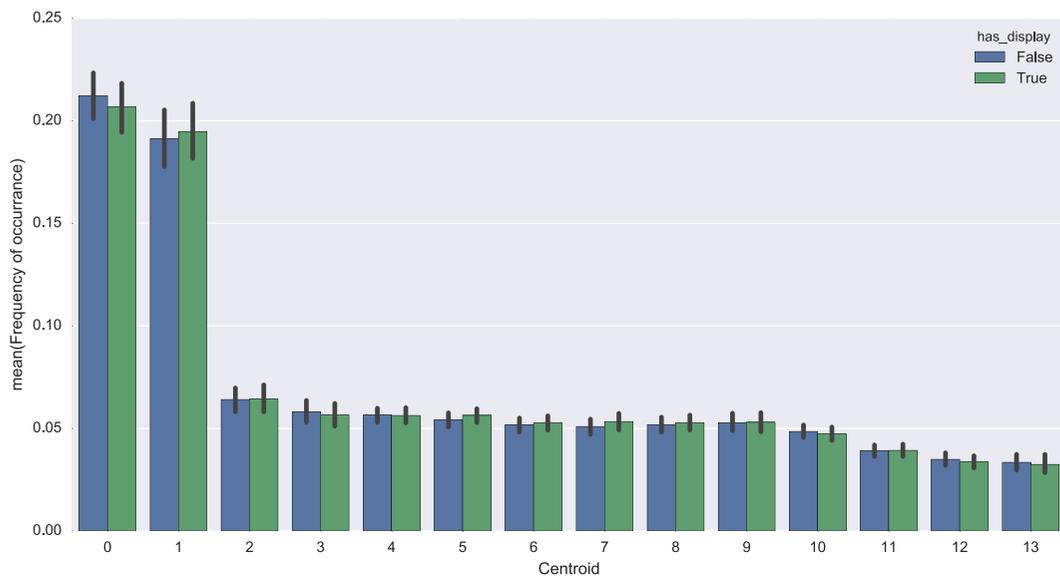
The reduction effect detected with the regression analysis is small (0.5–1.9% on average), but it is statistically significant for half of the models, with confidence intervals within  $-0.6$  to  $3.3\%$ , and the magnitude is in line with the most recent studies. Such a small average conservation effect can have several explanations in the context of this study. The in-home display alone did not provide any monetary incentive or direct message promoting energy conservation, which are documented to be more effective means than information feedback alone. The geographical area of interest is relatively mountainous,



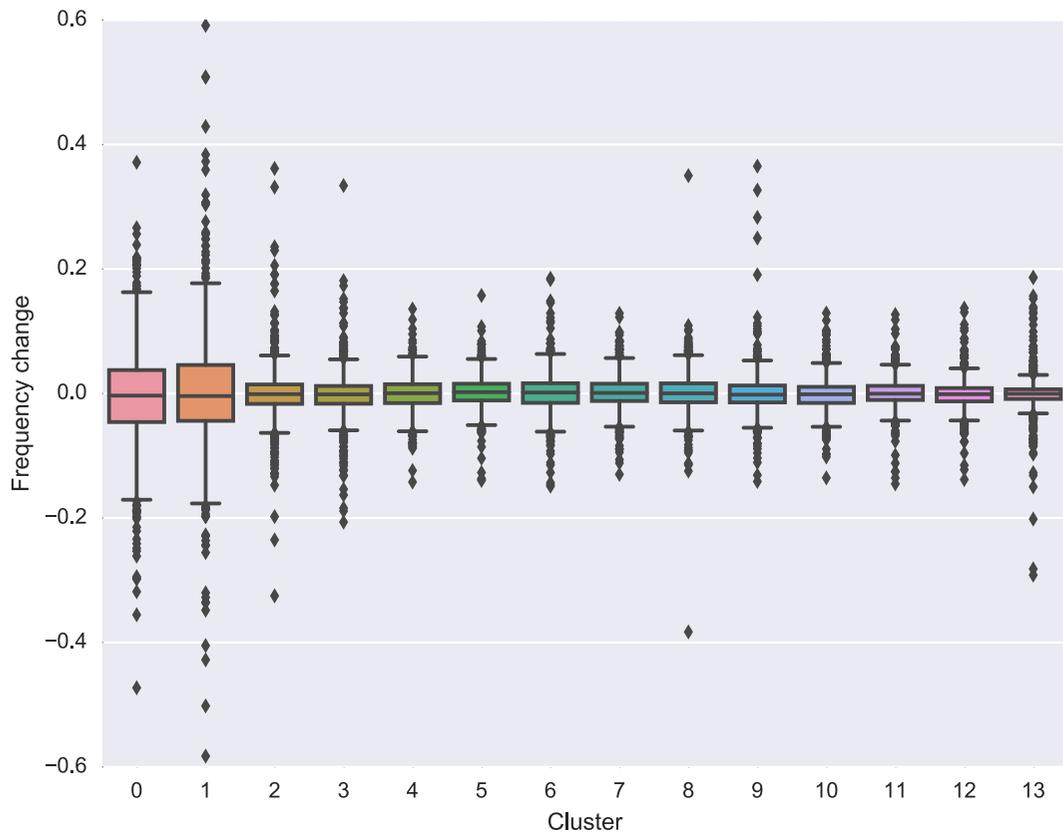
**Fig. 6** Inertia as a function of number of clusters used in K-means (on the left), and derivative of the same graph (on the right)

with a climate not requiring airconditioning. Heating on the other side is rarely done via electricity. Power demand is mostly related to lighting and other low-consumption appliances, making it harder to save

more energy. The experiment also took place in a moment and region with trends of already decreasing electricity consumption, leaving even smaller room for additional energy savings.



**Fig. 7** Prototypical load shapes resulting from the K-means clustering. Primary and secondary peaks are described in each subplot, in terms of % level and hour of occurrence

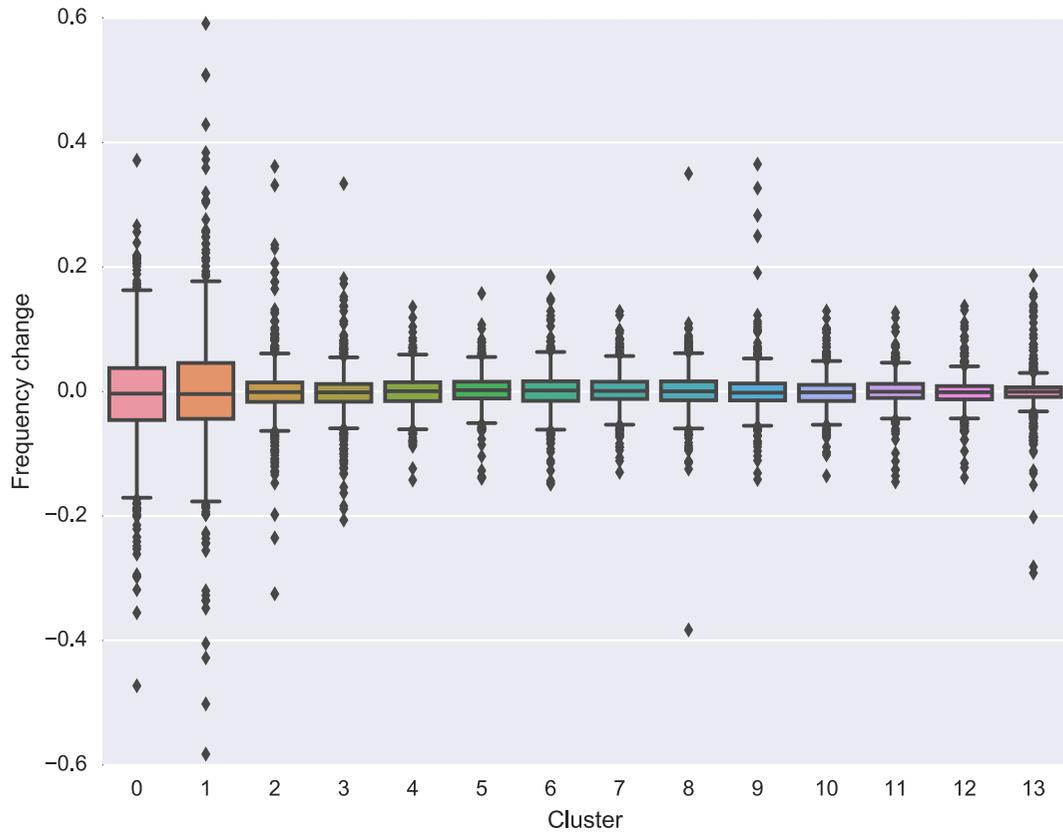


**Fig. 8** Frequencies of occurrence of centroids, averaged across clients, to which load shapes before and after the in-home display arrival are clustered. Black vertical bars represent confidence interval for the means

Different models and subsamples suggest slightly different estimates. The impact is lowest when considering overall monthly consumption across the wider “Client” sample (models 1 and 2). The average effect is possibly lowered by those HHs who were involved in the project but did not engage, and those time periods where engagement might be lower, e.g., during the night, during weekends, and long after the intervention.

Models fitting the “Load” and “Survey” samples support the hypothesis of higher conservation effects than those obtained with comparable “Client” models. HHs in the “Load” sample show higher consumption per capita in the first place (Table 1); hence, they might have more options to reduce consumption and more to gain from energy demand control. Furthermore, related models use daily data, which is mostly available for a fraction of the 3-year time span of the monthly data. Daily data might have picked up a

stronger effect, excluding periods long after the intervention with possibly smaller effects. On the other hand, high-frequency data allow for greater identification power across time periods, with the presence of daily variables, and across HHs, with more observations per HH. The narrowing effect of confidence intervals due to a higher number of observations is compensated by the strong serial correlations between measurements for each HH over time. HHs responding to surveys, fitting into the “Survey” subsample, might have higher interest in the project and higher motivation in reducing consumption in the first place, with respect to the wider audience reached. Nonetheless, confidence intervals increase as well, as the number of observations reduces and an extensive list of surveyed variables drives the estimation of power consumption baselines for each HH. More heterogeneity in the effect of the in-home display emerges as a result.



**Fig. 9** Changes in frequencies of occurrence of centroids, to which load shapes before and after the in-home display arrival are clustered. Each household corresponds to a series of 14 points, one per cluster

We do not find any statistically significant evidence that the conservation effect vanishes over time (Fig. 5). The estimated effect associated to having the display in the first month is similar to that of having the display for more than 5 months. Average reduction in consumption seems to slightly wear off in the first months, but in most models it reverts back to the levels of the first month in the long run. This is consistent with a narrative where households eventually find a consistent way of living with less energy, either by honing energy-conservation practices or by deciding to replace inefficient appliances.

While the in-home display achieved a reduction in power consumption, it was not able to systematically shift consumption peaks over time (Fig. 8). There were individual households that did change their lifestyles (Fig. 9) after receiving the display, but the net effect is zero. Households who switched to flatter daily load profiles counterbalance those who concentrated their consumption around specific peaks. The in-home

display did not set any incentive to shift peaks of consumption. While HHs were reminded about the current time billing slot by the display, price differential across slots is very low, and often HHs prefer flat billing contracts that do not have such differential. Without specific messages or stronger economic incentives, the in-home display did not achieve any systematic load shift. Nonetheless, heterogeneity in the response is very high, a result which is consistent with the literature on the impact of real-time feedback programs.

One additional result stems from the survey, which reveals that the treated group has an above average number of household members, children, rooms, and appliances in house. This is consistent with having more opportunities for energy conservation. Designing an efficient large-scale roll out of in-home displays could take this information into account, and target households fitting these characteristics first. If brought to scale, even a few percent reduction may matter to the power industry.

## Conclusions

Residential consumers determine each day's magnitude and shape of their daily load curve. Can a real-time information feedback device systematically affect these properties? And by how much? We tried to address these policy-relevant questions using data from a recent large-scale trial carried out in Italy.

According to our estimates, which involve six alternative plausible statistical models, an average reduction in daily power consumption of 0.5–1.9% can be attributed to such device, even though with marginal statistical significance. This is in line with more recent experimental studies on the topic.

Average hourly power consumption behavior, measured in terms of distributions of preferences for representative load shapes by each user, seems to remain unchanged before and after the arrival of the display. Nonetheless, a wide heterogeneity has yet to be explained.

Further data on the households would be useful to try to segment the analysis in meaningful groups, identifying those customers for which information feedback led to either significantly lower consumption levels or significantly different consumption habits.

Consumption data at the appliance level would provide invaluable information on behavior and behavioral change, even though at the cost of increasing computational tractability.

Results are specific to the group involved in the study, and are not easily generalizable to the wider population. The trial was carried out in an area of the country with lower than average per capita energy consumption, as well as income, allowing for more limited adjustments and investments. On the other hand, the sample households appear to have more family members than average, possibly young children. Future better experimental designs will also help in avoiding issues of sample representativeness and absence of control group.

All these potential steps for further research could lead to a better understanding of how such interventions perform at a large scale, and whom these interventions could be most beneficial for.

**Acknowledgements** We thank the utility company that collected and provided the data and that made this analysis possible. We also thank Prof. M. Jordan, Prof. G. d'Adda and Prof. J. Bonan for the helpful suggestions.

**Funding** The research leading to these results has received funding from the ERC grant agreement no. 336155 (COBHAM). We also acknowledge support from the EDITS project, sponsored by the Research Institute of Innovative Technology for the Earth (RITE), and from the NewTRENDS project, funded by the European Unions H2020 research and innovation programme under grant agreement no. 893311.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Supplementary Information** The online version contains supplementary material available at (<https://doi.org/10.1007/s12053-020-09922-z>).

## References

- Abrahamse, W., & Steg, L. (2009). How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings? *Journal of Economic Psychology*, 30(5), 711–720. <https://doi.org/10.1016/j.joep.2009.05.006>. <http://www.sciencedirect.com/science/article/pii/S0167487009000579>.
- Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9–10), 1082–1095. <https://doi.org/10.1016/j.jpubeco.2011.03.003>. <http://www.sciencedirect.com/science/article/pii/S0047272711000478>.
- Allcott, H., & Mullainathan, S. (2010). Behavior and energy policy. *Science*, 327(5970), 1204–1205. <https://doi.org/10.1126/science.1180775>.
- Andor, M.A., & Fels, K.M. (2018). Behavioral economics and energy conservation – a systematic review of non-price interventions and their causal effects. *Ecological Economics*, 148, 178–210. <https://doi.org/10.1016/j.ecolecon.2018.01.018>. <http://www.sciencedirect.com/science/article/pii/S092180091731039X>.
- Arthur, D., & Vassilvitskii, S. (2007). k-means++: the advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms* (pp. 1027–1035): Society for Industrial and Applied Mathematics.
- Attari, S.Z., DeKay, M.L., Davidson, C.I., de Bruin, W.B. (2010). Public perceptions of energy consumption and savings. *Proceedings of the National Academy of Sciences*, 107(37), 16054–16059. <https://doi.org/10.1073/pnas.1001509107>. <http://www.pnas.org/content/107/37/16054>.
- Attari, S.Z., Gowrisankaran, G., Simpson, T., Marx, S.M. (2014). Does information feedback from in-home devices reduce electricity use? Evidence from a field experiment. <http://papers.ssrn.com/abstract=2545187>.
- Fontana, M., Tavoni, M., Vantini, S. (2019). Functional data analysis of high-frequency load curves reveals drivers of residential electricity consumption. *PLOS ONE*, 14(6), e0218702. <https://doi.org/10.1371/journal.pone.0218702>.

- <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0218702>.
- Haben, S., Singleton, C., Grindrod, P. (2015). Analysis and clustering of residential customers energy behavioral demand using smart meter data. *IEEE Transactions on Smart Grid, PP(99)*, 1–1. <https://doi.org/10.1109/TSG.2015.2409786>.
- Hastie, T., Tibshirani, R., Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Berlin: Springer Science & Business Media.
- Houde, S., Todd, A., Sudarshan, A., Flora, J.A., Armel, K.C. (2013). Real-time feedback and electricity consumption: a field experiment assessing the potential for savings and persistence. *The Energy Journal*, 34(1). <https://ideas.repec.org/a/aen/journl/ej34-1-04.html>.
- Ito, K., Ida, T., Tanaka, M. (2018). Moral suasion and economic incentives: Field experimental evidence from energy demand. *American Economic Journal: Economic Policy*, 10(1), 240–267. <https://doi.org/10.1257/pol.20160093>. <https://www.aeaweb.org/articles?id=10.1257/pol.20160093>.
- Jessoe, K., & Rapson, D. (2014). Knowledge is (less) power: experimental evidence from residential energy use. *The American Economic Review*, 104(4), 1417–1438. <https://doi.org/10.1257/aer.104.4.1417>.
- Karlin, B., Zinger, J.F., Ford, R. (2015). The effects of feedback on energy conservation: a meta-analysis. *Psychological Bulletin*, 141(6), 1205–1227. <https://doi.org/10.1037/a0039650>.
- Kwac, J., Flora, J., Rajagopal, R. (2014). Household energy consumption segmentation using hourly data. *IEEE Transactions on Smart Grid*, 5(1), 420–430. <https://doi.org/10.1109/TSG.2013.2278477>.
- Lynham, J., Nitta, K., Saijo, T., Tarui, N. (2016). Why does real-time information reduce energy consumption? *Energy Economics*, 54, 173–181. <https://doi.org/10.1016/j.eneco.2015.11.007>. <http://www.sciencedirect.com/science/article/pii/S0140988315003011>.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Oakland, CA, USA*, (Vol. 1 pp. 281–297).
- McKerracher, C., & Torriti, J. (2013). Energy consumption feedback in perspective: integrating Australian data to meta-analyses on in-home displays. *Energy Efficiency*, 6(2), 387–405. <https://doi.org/10.1007/s12053-012-9169-3>.
- Räsänen, T., & Kolehmainen, M. (2009). Feature-based clustering for electricity use time series data. In *International conference on adaptive and natural computing algorithms* (pp. 401–412): Springer.
- Rhodes, J.D., Cole, W.J., Upshaw, C.R., Edgar, T.F., Webber, M.E. (2014). Clustering analysis of residential electricity demand profiles. *Applied Energy*, 135, 461–471. <https://doi.org/10.1016/j.apenergy.2014.08.111>. <http://www.sciencedirect.com/science/article/pii/S0306261914009349>.
- Schleich, J., Klobasa, M., Gözl, S., Brunner, M. (2013). Effects of feedback on residential electricity demand—findings from a field trial in Austria. *Energy Policy*, 61, 1097–1106. <https://doi.org/10.1016/j.enpol.2013.05.012>. <http://www.sciencedirect.com/science/article/pii/S0301421513003443>.
- Tiefenbeck, V., Goette, L., Degen, K., Tasic, V., Fleisch, E., Lalive, R., Staake, T. (2016). Overcoming salience bias: how real-time feedback fosters resource conservation. *Management Science*, 64(3), 1458–1476. <https://doi.org/10.1287/mnsc.2016.2646>. <https://pubsonline.informs.org/doi/abs/10.1287/mnsc.2016.2646>.
- Wang, Y., Chen, Q., Kang, C., Zhang, M., Wang, K., Zhao, Y. (2015). Load profiling and its application to demand response: a review. *Tsinghua Science and Technology*, 20(2), 117–129. [http://www.researchgate.net/profile/Yi\\_Wang137/publication/275365433\\_Load\\_profiling\\_and\\_its\\_application\\_to\\_demand\\_response\\_A\\_review/links/553e58fb0cf20184050eb255.pdf](http://www.researchgate.net/profile/Yi_Wang137/publication/275365433_Load_profiling_and_its_application_to_demand_response_A_review/links/553e58fb0cf20184050eb255.pdf).
- Zhou, K.L., Yang, S.L., Shen, C. (2013). A review of electric load classification in smart grid environment. *Renewable and Sustainable Energy Reviews*, 24, 103–110. <https://doi.org/10.1016/j.rser.2013.03.023>. <http://www.sciencedirect.com/science/article/pii/S1364032113001822>.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.