Journal of Behavioral and Experimental Economics 114 (2025) 102328

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

journal homepage: www.elsevier.com/locate/jbee

Behavioral interventions and market efficiency: The case of a volatile retail electricity market

Rimvydas Baltaduonis [a,*], Jūratė Jaraitė [b], Andrius Kažukauskas [b]

[a] Economics Department, Gettysburg College, 300N Washington St, Gettysburg, PA 17325, USA
[b] Faculty of Economics and Business Administration, Vilnius University, Saulėtekio al. 9, Vilnius 10105, Lithuania

ARTICLE INFO

JEL classification:
C92
D61
D91
L94
Q41

Keywords:
Electricity
Laboratory experiment
Market efficiency
Peak prices
Peer comparison
Price notifications

ABSTRACT

Numerous field experiments have demonstrated that various monetary and informational incentives encourage demand response by increasing awareness about peak electricity prices and potentially inefficient energy use. However, very little is known about the effects of such interventions on overall market efficiency. We conducted a laboratory experiment with 200 participants to test the effects of different interventions on consumer decisions and overall market efficiency in a market reminiscent of a retail electricity market. We investigate two types of incentives—monetary information in the form of notifications about surge prices and non-monetary informational incentives in the form of peer comparisons—separately and together. We find that notifications about surge prices are effective interventions for reducing resource use and increasing market efficiency during surge-price periods. During these periods, the combination of peak-price notifications and peer-comparison information exhibits the highest efficiency.

1. Introduction

Keeping the balance between demand and affordable supply in the power system is very challenging during peak demand hours. In these hours, only expensive and usually more polluting supply options are left to keep the power system in balance. With the increasing share of intermittent renewables in electricity generation, peak demand hours will likely be more volatile than they have been. One way to help balance the power system and moderate power prices is through changes on the demand side to make it more responsive and flexible.

Power demand, especially in the residential sector, is very unresponsive to peak electricity prices (see, e.g., Lanot and Vesterberg 2021), primarily for the following reasons. First, the literature on preferences for electricity tariffs shows that a considerable share of residential electricity users prefer flat electricity prices (see, e.g., Torriti et al. 2011, Vesterberg 2018, Directorate-General for Energy, 2022). In times of increasing and volatile electricity prices, we might expect the preference for fixed electricity price contracts to increase. Second, a considerable share of households still does not have access to smart electricity meters (see, e.g., ACER 2023). Consequently, households may simply be unaware of the size and/or timing of peak electricity consumption and related prices. Third, until recently, electricity has comprised only a modest share of household budgets in the Global North (see, e.g., Jessoe and Rapson 2014); thus, it may have been rational for households not to pay attention to their electricity usage, not to switch to real-time electricity pricing, not to invest in energy-efficient appliances, and not to save electricity during peak demand hours.

Inelastic peak demand has encouraged closer attention to various demand-side management (DSM) strategies that could help reduce peak electricity consumption or move electricity consumption from peak to non-peak hours, lowering consumers' electricity bills at the same time. One type of DSM is to encourage demand response in the residential sector by using various monetary, informational, and behavioral incentives (Buckley, 2020), which may increase awareness about peak electricity prices and potentially inefficient energy use. A number of review studies have shown that these measures are effective in reducing residential electricity consumption (see, e.g., Abrahamse et al. 2005, Darby 2006, Fischer 2008, Ehrhardt-Martinez et al. 2010, Faruqui et al. 2010, Andor and Fels 2018, Buckley 2020). These studies provide evidence that informational and behavioral interventions may be more effective in triggering electricity conservation than purely monetary incentives such as different pricing contracts.

* Corresponding author.
E-mail address: rbaltadu@gettysburg.edu (R. Baltaduonis).

https://doi.org/10.1016/j.socec.2024.102328
Received 18 August 2023; Received in revised form 22 October 2024; Accepted 6 December 2024
Available online 8 December 2024

For instance, in their meta-review study, Ehrhardt-Martinez et al. (2010) found that feedback interventions result in a greater overall reduction in electricity consumption than dynamic pricing, while the latter is more effective at decreasing demand during peak hours. Similarly, Buckley (2020), carrying out a meta-analysis of the most recent studies, concluded that electricity savings of 2–4 % can be achieved through informational and behavioral incentives and that individual and real-time feedback as well as personalized advice on how to save electricity are more effective than pricing strategies and monetary information.

While field experiments provide valuable insights about the effectiveness of various interventions in energy and other resource domains, by nature they cannot assess the implications of effective behavioral interventions on overall market efficiency as easily as the induced-value laboratory experiments can. For example, we might have a situation when two different interventions in the residential electricity market are similar in terms of their effectiveness in achieving residential electricity savings, but it remains unclear which one leads to greater consumer surpluses and overall market efficiency.[1] This observation emphasizes the value of results from laboratory experiments, which could complement or inform the designs of more expensive large-scale controlled experiments in the field or aid in bridging knowledge gaps when field-experimental data are unavailable, especially when the objective is to compare the market efficiency of different interventions.

Although there are other laboratory experiments that test the effectiveness of various electricity pricing schemes (e.g., Adilov et al. 2004, Barreda-Tarrazona et al. 2012, Baltaduonis and Weisz 2014, Atasoy et al. 2018) or investigate the effects of different informational and behavioral incentives on the conservation of electricity and other resources using the common pool resource game (e.g., Buckley and Llerena 2022), to the best of our knowledge, no study has contrasted the effects of various informational interventions on overall market efficiency in the same laboratory setting. Hence, our study contributes to the field of economic laboratory experimental research by reporting on a laboratory experiment that compares consumer decision-making in the presence of critical peak-price notifications, peer comparisons of consumption levels, both interventions together, and no intervention. Furthermore, we compare the effects of these interventions on overall market efficiency. We postulate that during surge-pricing periods and in the absence of dynamic pricing, the simple knowledge of wholesale market prices is not sufficient to induce significant resource-conservation behavior and that additional measures, such as surge-price notifications, are needed to increase awareness about expected critical peak periods. In addition, we argue that peer comparisons combined with surge-price notifications should strengthen the response, especially among consumers who need stronger nudges or have difficulty understanding monetary information. Regarding market efficiency, we predict that during price surge events, alerting information should enable consumers to make consumer surplus–enhancing choices, which should increase overall market efficiency.

Our first major result shows that during peak periods of cost shocks, notifications about surge prices were the most effective in reducing resource use. On average, these interventions reduced resource use by approximately 14 % relative to the treatment providing price information only. This result suggests that when coupled with real-time critical peak notifications, price acts as an effective measure to achieve resource savings at a desired time.

Next, we find that during surge-pricing periods, the combination of peak-price notifications and peer comparison information resulted in the highest efficiency. Surge-price notifications alone are ranked as the second-best intervention in terms of economic efficiency. For all interventions, efficiency gains did not extend beyond peak pricing events into the other periods, implying that efficiency gains from advance price alerts are short-lived. Finally, we observe that most of the efficiency gains resulting from these informational interventions are produced in low-efficiency markets.

Can we draw any conclusions about the external validity of our results? Our laboratory experiment was specifically designed as a pilot study for a randomized country-wide field experiment, which aimed to test the same behavioral interventions in the case of residential electricity prosumers.[2] The results of the field experiment are summarized in Kažukauskas et al. (2024). Although the field experiment could not evaluate the effects of the same behavioral interventions on market efficiency, it allowed us to compare the results on the effectiveness of our selected interventions and to provide some insights about the usefulness of social comparison information. Interestingly, in the field as in the laboratory we found that: first, the combination of critical peak price notifications and social comparison information was the most effective intervention in reducing overall net electricity use; second, critical peak price notifications alone were sufficient to induce conservation behavior during critical peak hours; third, social comparison on its own was not effective; and finally, we did find evidence that individuals who received social comparison information perceived it as less useful than those who did not receive such information. Altogether, the results from the field confirm the main findings from the lab. Furthermore, it suggests that our laboratory experiment can be an effective exploratory tool to study impacts of behavioral interventions when lengthy, costly and properly executed randomized control trials in the field are not an option.

Our results are timely as they provide new evidence from the economic experimental laboratory that consumers could help to reduce energy demand during peak hours and, more importantly, could positively contribute to energy-market efficiency. Together with more sophisticated digital technologies and pricing strategies, various informational interventions could increase the role of demand response in balancing volatile power systems. Overall, these results indicate that peak energy issues may be alleviated by using low-cost and easily implemented informational feedback.

The rest of the paper is structured as follows. In Section 2, we present the design and procedures of our laboratory experiment as well as the hypotheses that we test. In Section 3, we describe the experimental data and provide some descriptive results. In Section 4, we outline and discuss the results of the regression analysis and examine some policy implications. We conclude in Section 5.

2. Experimental design, procedures, and hypotheses

2.1. Environment and market institutions

Like Baltaduonis and Weisz (2014), we are among the first to analyze electricity consumer behavior by extending the work of Rassenti et al. (2001, 2003), who conducted a laboratory economics experiment with an electricity market structure exhibiting cyclical consumer demand. Our experiment focuses on buyers' decisions at the retail level. To isolate the behavioral effects of notifications about critical peak days as well as average levels of peer consumption, we examine a rather simple environment compared to actual retail electricity markets: intertemporal substitution of consumption is absent, and the non-market procurement of electricity is not possible. The relative market performance in terms of total surplus (the sum of Marshallian consumer surplus and producer surplus) under a flat-rate pricing (FRP) contract is measured in a cyclical demand and competitive supply environment found in electricity markets while controlling for the unilateral market power of the buyers. All

[1] We refer the reader to Harrison and List (2004), Levitt and List (2007) for a comprehensive overview and comparison of field and laboratory experiments, as well as a discussion of how laboratory experiments can complement field experiments.

[2] This study is registered in the AEA RCT Registry and the digital object identifier (DOI) is https://doi.org/10.1257/rct.6379-1.0.

aspects in the experimental design were carefully chosen to capture the key stylized features of retail electricity markets. For example, even though some power utility jurisdictions are gradually introducing more dynamic pricing options for electricity bills, the FRP contract continues to be a popular choice for billing purposes by consumers, where that choice is available to them.[3]

2.1.1. Environment

2.1.1.1. Demand. In each period, termed a "day," four buyers, who belong to an independent market group, are presented with units that they can purchase. The generic nomenclature of experimental goods is adopted purposefully in order to avoid suggestive behavior that participants engage in their daily lives. The quantities of units available for purchase vary cyclically across different days. There are four days in a "week" and a total of two weeks in a "month." At the end of each month, the buyers receive a monthly bill for the purchases made. Each day represents a separate market pricing period: Day 1 is an off-peak period (low demand, night), Day 2 is a shoulder period (medium demand, morning), Day 3 is a peak period (high demand, afternoon), and Day 4 is another shoulder period (medium demand, evening). These cycles of four pricing periods are designed to mimic the typical fluctuations in demand for electricity during a 24 h period (daily load curve), which are reflected in day-ahead electricity markets, where the market-clearing price is determined hourly by the most expensive supply offer fulfilling the demand.

Fig. 1 and Table A1 in Appendix A present the aggregate demand and supply curves during the 15 months of the experiment, depicting a typical supply for all days. On critical days, which in our experiment corresponds only to Day 3, this typical supply gets multiplied by a factor representing a shock to a system during some peak periods ("critical peak periods"). The four buyers in a market are each denoted by "B" followed by an identification number (see Table A1 in Appendix A). The marginal benefits derived from consuming different units are parameterized and distributed among buyers in a way that induces inelastic market demand in the neighborhood of competitive prices while making sure that buyers do not have the unilateral market power to profitably deviate from their competitive equilibrium consumption levels, that is, to singlehandedly cause oligopsonistic market prices. Notably, the market demand on Days 2 and 4 is identical on the aggregate level, but the units carrying the same marginal benefit are not assigned to the same buyer more than once during these two days. This assignment serves as a control for whether unit distribution among buyers has an effect on market efficiency on Days 2 and 4.

2.1.1.2. Supply. The market supply reflects the increasing marginal cost of production and is stationary throughout the experiment (see Supply 1 curve in Fig. 1), with the exception of seven Day 3 instances ("events") that experience the following positive supply shocks built in as multipliers for the typical supply costs: (1) Month 6, Week 1: multiplier = 1.9; (2) Month 7, Week 2: multiplier = 4; (3) Month 9, Week 2: multiplier = 2.6; (4) Month 10, Week 1: multiplier = 2.6; (5) Month 10, Week 2: multiplier = 2.6; (6) Month 13, Week 1: multiplier = 1.9; and (7) Month 14, Week 2: multiplier = 10.9.

Since the focus of this study is on demand-side behavior in different information environments, the aggregate market supply is modeled as a competitive process and is therefore implemented via true cost bidding by robot sellers (electricity producers) in the wholesale market. Any intermediary, such as a regulated utility company or electricity retailers that purchase energy on the wholesale market to supply their retail customers, is also modeled to represent perfectly competitive outcomes. In other words, they are captured by robots that merely pass through the dynamic wholesale costs by transforming them into a respective FRP rate that consumers have to pay for their consumed energy.

2.1.1.3. Knowledge. All participants in the experiment are aware that each participant knows only the marginal benefit of the units available to them. Individual participants do not know the marginal benefit of the units available to other buyers, but they learn the marginal costs on the supply side since the wholesale unit price is presented to the buyers in real time. Thus, every participant is aware that the aggregate demand for units determines the market price during each day. The participants also know that they are all billed under the same type of pricing contract, that is, FRP.

2.1.2. Market institutions

All consumers are enrolled in an FRP contract where all costs associated with production are equally distributed over the total quantity of units produced regardless of the timing of consumption. To capture the essence of this pricing contract, a uniform price per unit is calculated as the weighted average of the wholesale market prices during the month and charged for all retail purchases of that month. This type of pricing contract exhibits the highest degree of aggregation of wholesale price signals and the lowest exposure of retail consumers to market fluctuations. The participants pay a uniform price for all units purchased each month, but market prices vary quite considerably throughout the month. However, all buyers are able to observe daily wholesale market prices in real time.

To study the information effects on retail market performance, we conduct five treatments employing a between-subject design. Our baseline treatment (T0) has no supply shocks while the other four treatments include seven instances of supply shocks on Day 3 and different informational environments starting in Month 6:

- Treatment 1 (T1): only the positive supply shocks, no messages about surge pricing and no peer comparisons;
- Treatment 2 (T2): peer comparisons available;
- Treatment 3 (T3): messages about surge pricing but no peer comparisons;
- Treatment 4 (T4): messages about surge pricing and peer comparisons available.

The treatment-specific informational environments are implemented only after the fifth experimental month to allow for learning and observe any changes in within-subject behavior when the interventions are introduced. Thus, all five treatments are designed to have the first five months be identical in all aspects, which also serves as a control in our between-subject experimental design to ensure that the behavior is statistically similar across all treatments before the interventions are introduced.

2.1.2.4. Messages about surge pricing. Notifications about upcoming critical peak days are shown to the participants as a way to make them more attentive to their consumption during periods of extremely high market prices (see Fig. B1 in Appendix B). However, from an individual perspective, notifications about surge prices in the market are not particularly relevant for a consumer surplus–maximizing person because the price she pays is a uniform price determined *ex-post* as a result of the decisions made by all participants.

2.1.2.5. Messages about peer comparison. Starting in Month 6, the buyers in T2 and T4 can access accumulated historical information

[3] As indicated by Directorate-General for Energy (2022), flat rate tariffs are more popular than dynamic pricing among residential consumers in most EU countries. On the PAPowerSwitch website - a simple tool that presents all available pricing contracts for electricity residential customers in Pennsylvania (USA) - the vast majority of contract options are fixed price contracts. As the adoption of different or more sophisticated billing practices proliferates, this research could be extended to include other pricing schemes (e.g., block tariffs).

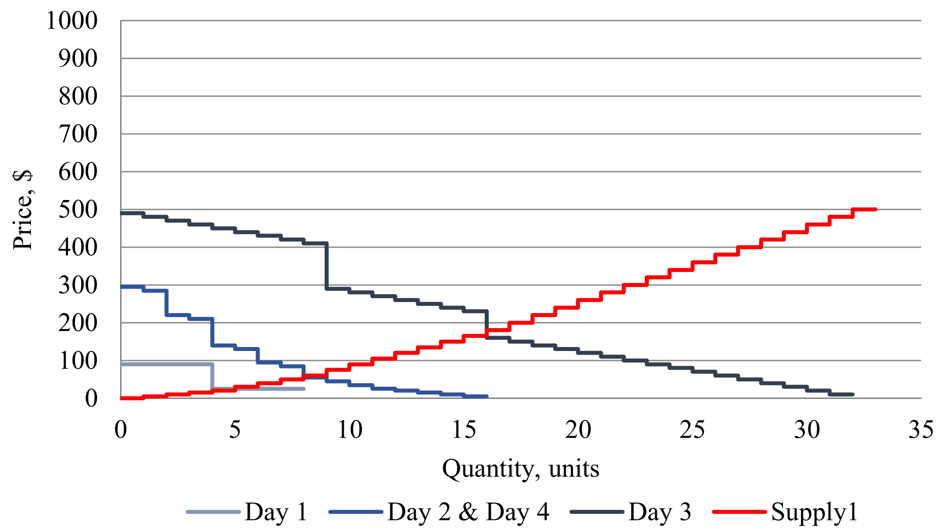

Fig. 1. Market demand and supply during a typical week.
Notes: This figure represents the demand (marginal benefit) schedules for Days 1–4 and the supply (marginal cost) schedule, which is the same for all experimental days.

about their own consumption on a specific day of a specific week as well as the historical market averages by clicking on a button on their screens (see Fig. B2 in Appendix B). Specifically, we compute and make accessible to the buyers information about the average number of units they purchased on a specific day of the week as well as the average number of units purchased by their peers in their market (see Fig. B3 in Appendix B). The historical averages of personal and peer consumption are also available at the monthly level (see Fig. B4 in Appendix B). The buyers could voluntarily choose whether or not to click on these peer comparison buttons. The buyers in T2 and T4 received the following instructions message: "From now on, information about historical monthly and daily consumption will be available to you by clicking on the respective buttons in the upper right corner. This will contain both individual and market averages."

2.2. Procedures

All experimental sessions were conducted in Gettysburg Lab for Experimental Economics (GLEE) during the months of April, May and September in 2021. Altogether, 200 participants were recruited, with 40 participants assigned to each treatment, resulting in 10 independent market groups per treatment. The participants were undergraduate students attending Gettysburg College who were randomly recruited from the Gettysburg College undergraduate student list, which includes all current student email addresses.

The treatments were randomized at the session level. All market groups in the same session participated in the same treatment since the experimental instructions were played out loud in order to create common knowledge about the market environment and rules among the participants.

Each participant drew a random card to be seated at one of the computer terminal stations, from which they could not view the other participants' monitors. The recorded video of the instructions was played to the participants at the start of each session, and paper instructions were available at the computer terminals (see the text of the instructions in Appendix B).

On all decision screens, the participants had to click on "Purchase Unit" buttons consecutively to buy units. If participants wished to cancel a purchase, they were able to click the "Undo Purchase" buttons in the opposite order they had selected units to buy. The participants were given 15 s to decide how many units they wished to purchase on a specific day. A table revealing the units available for purchase and their individual resale value was presented to the left of the buttons. If 15 seconds expired while the buyer was on the peer comparison screen, their current purchases (if any) were recorded as final for that day and the whole market advanced to the next day. On the decision screens, the participants were able to view their current balance, the number of units they had purchased, the latter's resale revenue, and the market price per unit. They were not able to see their costs, profit, or flat price per unit since those were calculated at the end of each month (see Fig. B5 in Appendix B).

While we recognize that resale value approach cannot perfectly capture all complexities of real-world consumer preferences, however, historically, the resale value approach has been widely used in lab experiments as a straightforward way to induce buyer preferences and consumer surplus. One of its key strengths lies in its simplicity, making it easier for participants to understand how their decisions can lead to profits. This clarity helps ensure that participants are more engaged and make informed decisions that reflect real-world behavior. Moreover, by linking choices to resale values, we simulate market conditions where participants can readily perceive the gains resulting from the difference between resale values and cost, thereby facilitating an effective measure of consumer surplus in the lab.

At the end of each month, participants viewed a monthly bill, which presented the flat price per unit calculated as the weighted average of the market prices during the month. They also saw the total units purchased, total resale revenue, total costs, total profit, and information on their updated current balance (see Fig. B6 in Appendix B).

During each session, the participants also answered two questionnaires, which asked them to rate the following statements about the usefulness of daily or monthly historical peer comparison information from "strongly disagree" to "strongly agree" (see Fig. B7 in Appendix B): "It would be useful for me to have historical daily/monthly consumption containing both individual and market averages." The first questionnaire was presented to the participants immediately before the treatment phase and the second one after it.

Each session lasted approximately 75 min. The participants were paid a show-up fee of 10 USD in addition to any earnings they made during the experiment. On average, the participants earned 11.23 USD during the experiment, not including the show-up fee of 10.00 USD. The median was 11.75 USD. Earnings ranged between 0.00 USD and 18.75 USD without the show-up fee.

2.3. Hypotheses

The purpose of our laboratory experiment is to compare the market efficiency outcomes of the above-described treatments and understand whether advance notifications about critical peak prices with and without peer comparisons (T3 and T4) lead to higher economic efficiency than simple real-time information about prices (T1). In addition, we aim to compare participants' decisions on resource conservation in the abovementioned interventions.

We predict that in the absence of dynamic pricing, advance notifications about expected surge pricing are needed to raise awareness about critical peak prices to increase overall market efficiency. Hence, the first hypothesis states that in the presence of cost shocks, real-time price information alone (T1) is not as economically efficient as the same price information combined with notifications about surging prices (T3). We also expect T3 to be more effective than T1 in terms of resource conservation. Our expectations are based on the findings of similar experimental field studies, which show that notifications/nudges to save energy during peak hours can be effective (see, e.g., Brandon et al. 2019, Ito et al. 2018).

Furthermore, we expect that peer comparison information should increase overall market efficiency and strengthen conservation behavior. Peer (or social) comparisons are some of the most popular behavioral interventions that have been used to induce behavioral change in consumers, particularly for energy and water resources. A large body of literature provides evidence that peer comparison information is effective in reducing residential energy and water consumption (for a review, see Každukauskas et al. 2021). Most of these experimental field studies generally argue or assume that more information (peer comparison with or without other monetary information) can increase overall efficiency.

Hence, our second hypothesis is that critical peak-price notifications along with peer comparison information (T4) lead to higher market efficiency than price-related information only (T1 or T3) and peer comparison information only (T2). Based on the findings of Brandon et al. (2019), whose experimental field study is arguably the most similar to ours, we expect that resource conservation and market efficiency will be higher under T4 than under T1–T3.

Next, we postulate that additional information of any type—monetary or non-monetary—is more important for resource users who normally fail to make choices that increase total surplus. Therefore, our third hypothesis is that peak-price notifications and peer comparison information together or by themselves (T4, T3, and T2) result in higher efficiency gains in markets with low-efficiency performance.

3. Descriptive evidence

Data was collected from five sessions of each treatment for a total of 25 sessions. Each session comprised 120 periods, representing the experimental days in the 15 experimental months (each experimental month contains eight experimental days).

Table 1
Number of participants by gender, nationality, and age across treatments.

		T0	T1	T2	T3	T4	Total
Gender	Female	25	27	28	26	22	128
	Male	14	13	11	13	17	68
	Other	1	0	1	1	1	4
Nationality	U.S.	33	38	37	36	39	183
	Other	7	2	3	4	1	17
Age	18–19	12	6	11	5	6	40
	20	11	13	13	8	12	57
	21	14	11	10	17	12	64
	22–23	3	10	6	10	10	39
Total		40	40	40	40	40	200

Notes: This table presents the major characteristics of the participants across the treatment groups. T0 refers to the treatment with no supply shocks, T1 to the treatment with cost shocks but no messages about surge pricing and no peer comparisons, T2 to the treatment with only peer comparisons, T3 to the treatment with messages about surge pricing but no peer comparisons, and T4 to the treatment with messages about surge pricing and peer comparison information.

A total of 24,000 observations of daily purchase decisions (16,000 observations excluding the pre-treatment/learning phase) were collected for all five treatments (T0–T4) using the above-described laboratory experiment. Table 1 provides the gender, age, and nationality distribution for all 200 participants in the laboratory experiment. The characteristics of the participants are very similar across the treatment groups (insignificant statistical differences between the groups), and our samples are dominated by young, female, and U.S.-citizen students.

We report two types of data for each treatment: market efficiency scores and purchased units. The latter outcomes are important since they could be compared to the results of the field experiments that test monetary and behavioral interventions similar to our laboratory experiment.

To calculate market efficiency scores for each day, we divide the achieved total surplus by the maximum possible total surplus. To obtain the achieved total surplus for each day, we calculate the consumer and producer surpluses for each day. Consumer surplus for a day is equal to the following:

$$\begin{aligned} \textit{Consumer Surplus} = {} & (\textit{Resale Revenue Buyer 1} - (\textit{Total Units Purchased} * \textit{Price per Unit})) \\ & + (\textit{Resale Revenue Buyer 2} - (\textit{Total Units Purchased} * \textit{Price per Unit})) \\ & + (\textit{Resale Revenue Buyer 3} - (\textit{Total Units Purchased} * \textit{Price per Unit})) \\ & + (\textit{Resale Revenue Buyer 4} - (\textit{Total Units Purchased} * \textit{Price per Unit})) \end{aligned} \tag{1}$$

where Resale Revenue is the induced buyer's value of the purchased units. Because we are using an FRP regime, the flat rate ("Price per Unit") is calculated as the weighted average cost for the units purchased over the eight days of the month:

$$\textit{Price per Unit} = \frac{(\textit{Total Units Purchased Day 1} * \textit{Market Price per Unit Day 1}) + \ldots + (\textit{Total Units Purchased Day 8} * \textit{Market Price per Unit Day 8})}{(\textit{Total Units Purchased Day 1} + \ldots + \textit{Total Units Purchased Day 8})} \tag{2}$$

Producer surplus for a day is calculated as follows:

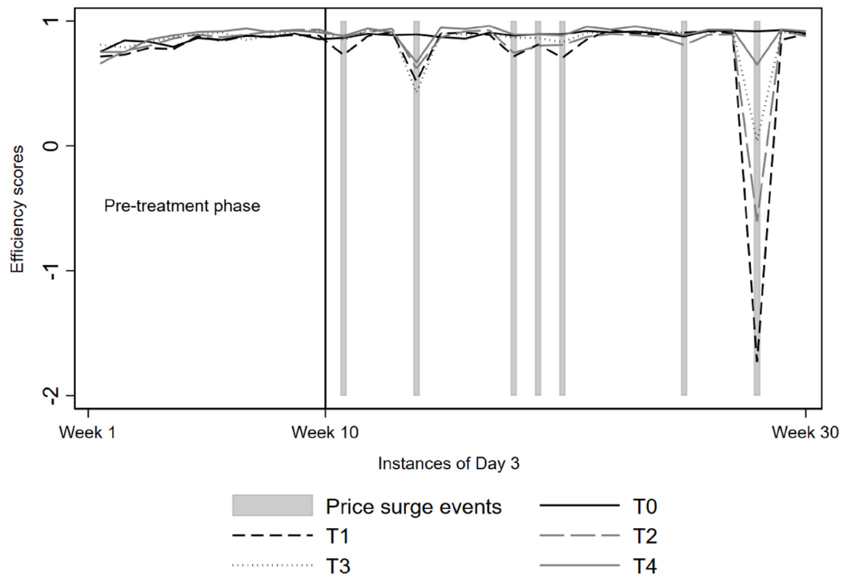

Fig. 2. Efficiency score averages by treatment group for Day 3.
Notes: This figure presents the average efficiency scores across the treatment groups T0-T4 for instances of Day 3, which include the seven price surge events indicated by the gray columns. T0 refers to the treatment with no supply shocks, T1 to the treatment with cost shocks but no messages about surge pricing and no peer comparisons, T2 to the treatment with only peer comparisons, T3 to the treatment with messages about surge pricing but no peer comparisons, and T4 to the treatment with messages about surge pricing and peer comparison information. In total, there are 30 instances of Day 3, including the seven price surge events.

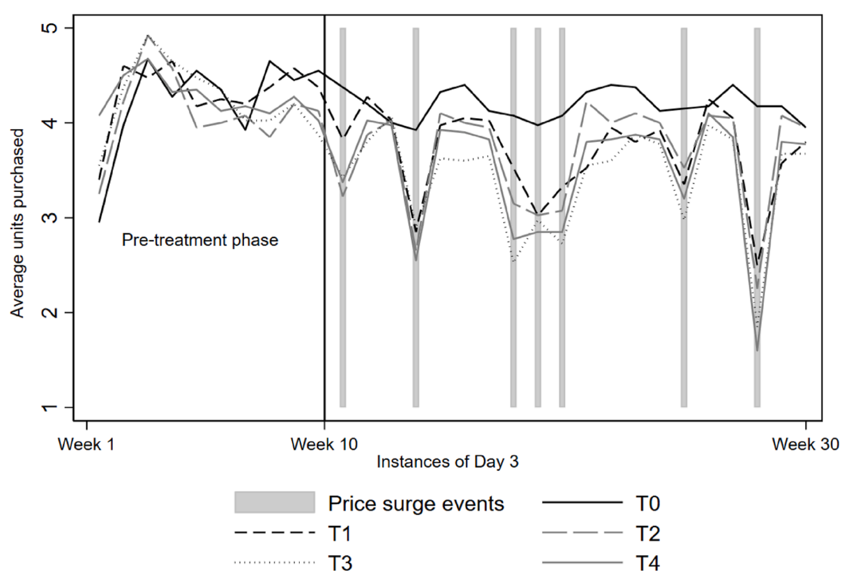

Fig. 3. Averages of purchased units by treatment groups for the instances of Day 3.
Notes: This figure presents the average purchased units across the treatment groups T0-T4 for instances of Day 3, which include the seven price surge events indicated by the gray columns. T0 refers to the treatment with no supply shocks, T1 to the treatment with cost shocks but no messages about surge pricing and no peer comparisons, T2 to the treatment with only peer comparisons, T3 to the treatment with messages about surge pricing but no peer comparisons, and T4 to the treatment with messages about surge pricing and peer comparison information. In total, there are 30 instances of Day 3, including the seven price surge events.

$$\begin{aligned} \textit{Producer Surplus} = & (\textit{Price per Unit} * \textit{Total Units Purshased}) \\ & - \textit{Producer Costs} \end{aligned} \tag{3}$$

Producer Costs for each day are determined by the marginal cost (MC) curve and the number of units purchased that day (see Table A1 in Appendix A).

Once the consumer and producer surpluses have been determined for each day, they can be added to compute the total surplus for a day. For example, the daily efficiency score for Day 1 is calculated as follows:

$$\textit{Daily Efficiency} = \frac{\textit{Total Surplus Day } 1}{\textit{Maximum Possible Surplus Day } 1} \tag{4}$$

For monthly efficiency calculations, the following is used:

$$Monthly\ Efficiency = \frac{Total\ Surplus\ Day\ 1 + \ldots + Total\ Surplus\ Day\ 8}{Maximum\ Possible\ Surplus\ Day\ 1 + \ldots + Maximum\ Possible\ Surplus\ Day\ 8} \quad (5)$$

We begin our descriptive analysis by plotting the simple mean values of the efficiency scores and purchased units, respectively, for all instances of Day 3 (T0–T4), which is the day in the experimental weeks when seven price surge events happened (Figs. 2 and 3). These events are highlighted by the shaded areas in both figures. Fig. 2 demonstrates that as the participants became more familiar with the experimental laboratory setting and gained more experience with the induced market values, the average market efficiency scores for all treatment groups tend to increase in our pre-treatment phase (Months 1–5). When the seven events of supply shocks take place, the participants in all treatment groups (T1–T4) responded by reducing the number of purchased units, which contrasts with the treatment group without cost shocks T0 (Fig. 3). However, the size of the responses differs across these groups: the least responsive treatment group is T1, which received neither notifications about surge pricing nor peer comparison information, whereas the most responsive is T4, which received both price surge notifications and peer comparison information.

However, being less or more responsive to cost shocks does not automatically mean that the participants made better or worse decisions in terms of overall market efficiency. We find that some treatment groups that were exposed to identical supply shocks had difficulty keeping their efficiency scores at the efficiency level they achieved before price surge events (e.g., in Fig. 2, the third, fourth, and fifth instances of cost shocks and the subsequent efficiency scores for T1 and T2). Furthermore, we observe that some treatment groups managed to optimize their choices better than others. The most notable treatment group is T4, which received both price surge notifications and peer comparison information. By being among the most responsive in terms of purchased units, the participants in this group also managed to maintain the highest average efficiency scores among the treatment groups that experienced cost shocks (T1–T3) during the critical peak days. The worst-performing treatment group in terms of average efficiency scores was T1, which also was the least responsive in lowering the average purchased units on critical days. T1 received neither price surge notifications nor peer comparison information (see Figs. 2 and 3).

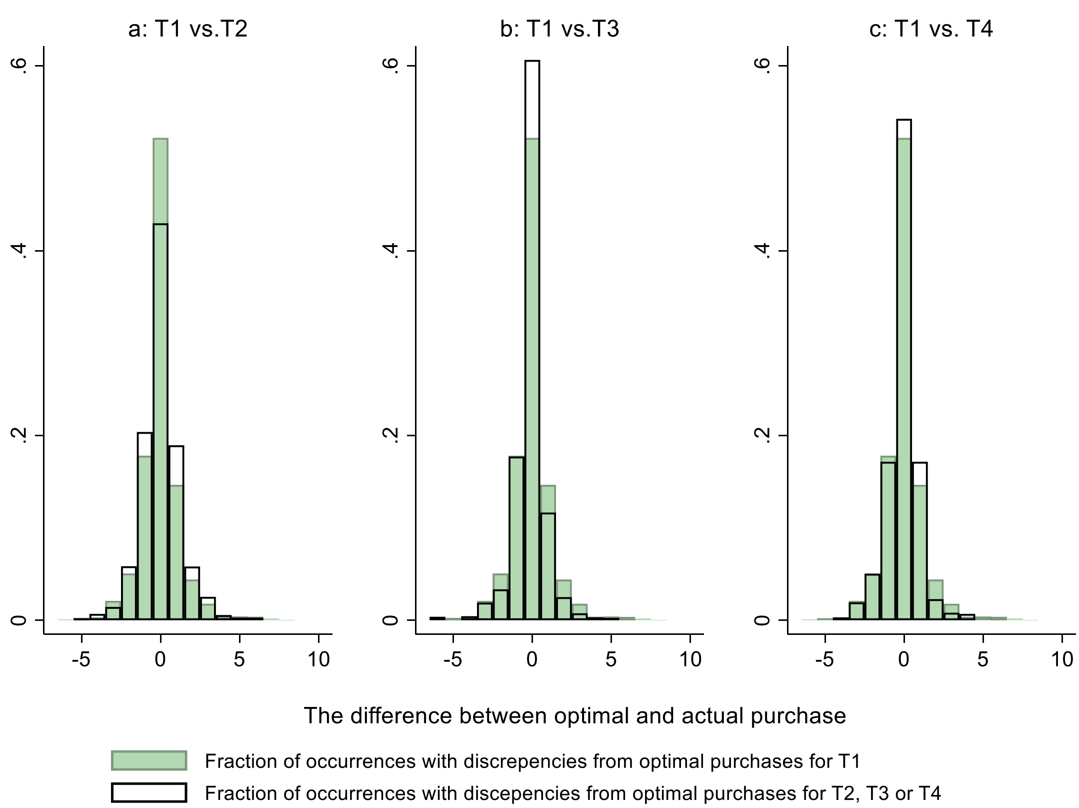

Fig. 4. Histograms of period differences between optimal and actual purchases by individual buyers during the treatment phase (T1–T4).
Notes: This figure contains three panels (a, b, and c), each showing a two-layered histogram comparing the distributions of the difference between the optimal and actual levels of purchased units per day by individual buyers during the treatment phase (Months 6–15) for two treatment groups (T1 vs. T2, T1 vs. T3, and T1 vs. T4). If the difference between the optimal and actual quantities of purchased units equals 0, the participants managed to optimize their consumption levels; if this difference is greater or <0, the participants did not behave as optimally as they could have. The green bars represent this distribution for T1 and the white-transparent bars for either T2, T3, or T4. Specifically, panel a compares the distribution of T1 to the distribution of T2, panel b that of T1 to T3, and panel c that of T1 to T4. T1 refers to the treatment with cost shocks but no messages about surge pricing and no peer comparisons, T2 to the treatment with only peer comparisons, T3 to the treatment with messages about surge pricing but no peer comparisons, and T4 to the treatment with messages about surge pricing and peer comparison information.

Table 2
Efficiency scores and average purchased units by treatment group (T0–T4).

Outcome units	Samples	T0	T1	T2	T3	T4
Average daily efficiency scores	Pre-treatment days (Months 1–5)	0.85	0.85	0.84	0.86	0.89
	All treatment days (Months 6–15)	0.87	0.83	0.84	0.89	0.89
	Peak days (instances of Day 3)	0.90	0.71	0.79	0.83	0.90
	Critical days	0.89	0.38	0.58	0.68	0.82
	Non-peak days (instances of Days 1, 2, and 4)	0.86	0.87	0.86	0.91	0.88
Average monthly efficiency scores	All treatment months (Months 6–15)	0.89	0.87	0.89	0.89	0.93
Average purchased units per day	Pre-treatment days (Months 1–5)	2.46	2.42	2.36	2.32	2.38
	All treatment days (Months 6–15)	2.34	2.12	2.16	2.00	2.06
	Peak days (instances of Day 3)	4.19	3.68	3.67	3.40	3.48
	Critical days	4.11	3.20	2.99	2.77	2.74
	Non-peak days (instances of Days 1, 2, and 4)	1.72	1.60	1.66	1.54	1.58
Average monthly purchased units	All treatment months (Months 6–15)	18.71	16.98	17.31	16.02	16.46

Notes: This table presents the average efficiency scores and average purchased units by individual buyers across different periods and treatment groups (T0–T4). "Pre-treatment days" refer to the first 40 experimental periods (Months 1–5), that is, the so-called learning phase. "All treatment days" refer to experimental periods 41 to 120 (Months 6–15), which followed the pre-treatment phase. "Peak days" are instances of Day 3. "Critical days" are instances of cost shocks, that is, price surge events. "Non-peak days" are instances of Days 1, 2, and 4. T0 refers to the treatment with no supply shocks, T1 to the treatment with cost shocks but no messages about surge pricing and no peer comparisons, T2 to the treatment with only peer comparisons, T3 to the treatment with messages about surge pricing but no peer comparisons, and T4 to the treatment with messages about surge pricing and peer comparison information.

The efficiency score levels are clearly correlated with the severity of the cost shocks (see Section 2.1.1 for the sizes of the supply shock multipliers). The most challenging event was the last one, which constituted the largest cost shock (the last shaded area in Figs. 2 and 3).

Another way to examine how information treatments contributed to better choices in terms of economic efficiency is to investigate whether these treatments helped participants to reduce the difference between the optimal and actual levels of purchased units per period. In Fig. 4, we report the distributions of these differences during the treatment phase (Months 6–15) for each treatment (T1: green bars; T2/T3/T4: transparent bars) and contrast them with one another. If the difference between the optimal and actual quantities of purchased units equals 0, the participants managed to optimize their consumption levels; if this difference is greater or less than 0, the participants did not behave as optimally as they could have. The participants in the peer comparison group T2 did not manage to increase the fraction of occurrences with optimal choices (0 values); instead, this fraction decreased by nine percentage points (see panel a in Fig. 4). Nevertheless, this group managed to improve their average efficiency by reducing the number of large mistakes. For instance, T2's fraction of deviations from optimal purchases that are greater than or equal to 6 units was half the size of that observed in T1 (0.34 % in T2 and 0.69 in T1). Panel b in Fig. 4 shows that surge-price notifications (T3) helped participants to increase the fraction of occurrences resulting in optimal purchases relative to the baseline treatment (T1) by eight percentage points, although they were slightly less successful in reducing large mistakes among participants in the treatment, with a fraction of deviations from optimal purchases greater than or equal to six units of 0.41. Finally, treatment group T4, which managed to keep the highest efficiency levels, as shown in Fig. 2, not only reduced extreme mistakes (with a fraction of deviations from optimal purchases greater than or equal to six units of only 0.09) but also increased the fraction of cases of optimal purchases by two percentage points (see panel c in Fig. 4). Thus, it appears that providing peer comparison information (T2 and T4) helps to lower the likelihood of extreme deviations from optimal purchase levels compared to the treatments without this information.

The descriptive statistics for each treatment group in Table 2 confirm what was visible in Figs. 2 and 3. The table provides the averages of daily efficiency scores and purchased units for the following experimental periods: the pre-treatment days (periods 1–40, Months 1–5) and all treatment days (periods 41–120, Months 6–15), which are then split into peak days (instances of Day 3), critical days (instances of Day 3 with cost shocks) and non-peak days (instances of Days 1, 2, and 4). Additionally, Table 2 provides the average monthly efficiency scores and average monthly purchased units for the treatment months.

On average, participants in the treatment groups T3 and T4 managed to maintain the highest daily efficiency scores (0.89) among the treatment groups that experienced cost shocks when all treatment days are considered. However, the average efficiency score is much higher for T4 than that for T3 when only peak days are considered (0.90 vs. 0.83). Meanwhile, participants in T1 and T2 performed the worst and second-worst (0.71 and 0.79), respectively, in terms of market efficiency scores during the peak days. Additional detailed descriptive statistics for all variables by treatment group can be found in Table C1 of Appendix C. For the convenience of the reader, we also report the percentage changes in the average daily efficiency scores relative to the pre-treatment period of each group (see Table C2 in Appendix C).

4. Regression analysis

We then estimate the random-effects regression models to capture the statistical differences between the treatment groups in terms of efficiency scores and purchased units in response to the exogenous supply shocks. In this analysis, we only focus on the treatment days (Months 6–15, or periods 41–120) and the treatment groups that were subjected to price surge events (T1–T4), with T1 serving as the baseline treatment. In doing so, we compare "apples" with "apples," that is, the choices made by participants in T1–T4 that were exposed to critical peak events. This means that we leave out of the analysis the pre-treatment or "learning" phase (Months 1–5, or periods 1–40) and the treatment group T0.

We estimate the following panel regression model with random effects[4] for units purchased:

$$y_{it} = \alpha_1 + \alpha_2 T_2 + \alpha_3 T_3 + \alpha_4 T_4 + u_i + \varepsilon_{it}, \tag{6}$$

where y_{it} is the number of units purchased by subject i at time t, α_1 is a constant, T_k are the dummy variables indicating whether the subject is in one of the treatment groups (T_1 serves as the baseline), u_i are the random effects, and ε_{it} is an idiosyncratic error term. The estimated coefficients α_2-α_4 measure the average causal treatment effects of our monetary and non-monetary interventions on the purchased units.

[4] We use Breusch and Pagan Lagrangian multiplier test for random effects: The null hypothesis that variances across entities is zero is rejected suggesting the random-effects model over the pooled OLS model.

Table 3
Effects of informational interventions on purchased units by individual buyers.

Variables	(1) All treatment days	(2) Peak days	(3) Critical days	(4) Non-peak days	(5) Monthly
Peer comparison only (T2)	0.042	−0.015	−0.214	0.060	0.332
	(0.138)	(0.312)	(0.347)	(0.113)	(1.100)
Price notification only (T3)	−0.120	−0.277★	−0.432★★	−0.068	−0.960
	(0.138)	(0.312)	(0.347)	(0.113)	(1.100)
Both (T4)	−0.065	−0.196	−0.457★★	−0.022	−0.523
	(0.138)	(0.312)	(0.347)	(0.113)	(1.100)
Constant	2.123✩✩✩	3.681✩✩✩	3.200✩✩✩	1.603✩✩✩	16.983✩✩✩
	(0.097)	(0.221)	(0.245)	(0.080)	(0.778)
Observations	12,800	3200	1120	9600	1600

Notes: This table presents the results of the estimated panel regression model with random effects as described in Eq. (6). The dependent variable is daily (or monthly) purchased units by individual buyers. "All treatment days" refer to experimental periods 41 to 120 (Months 6–15), which followed the pre-treatment phase. "Peak days" are instances of Day 3. "Critical days" are instances of cost shocks, that is, price surge events. "Non-peak days" are instances of Days 1, 2, and 4. T1 refers to the treatment with cost shocks but no messages about surge pricing and no peer comparisons, T2 to the treatment with only peer comparisons, T3 to the treatment with messages about surge pricing but no peer comparisons, and T4 to the treatment with messages about surge pricing and peer comparison information. ★★★/✩✩✩ $p < 0.01$.
★★/✩✩ $p < 0.05$, and ★/✩ $p < 0.1$ indicate significance levels, where filled stars ★ indicate significance levels preserved under randomization inference with clustering at market level (not available for constant term), while empty stars ✩ indicate significance levels that are sustained by the standard errors.

The following panel-data tobit regression model with random effects,[5] with right-censoring at 1, is employed for efficiency scores:

$$\theta_{mt} = \beta_1 + \beta_2 T_2 + \beta_3 T_3 + \beta_4 T_4 + \nu_m + \mu_{mt}, \tag{7}$$

where θ_{mt} is the efficiency score of market m (consisting of four buyers) at time t, β_1 is a constant, T_k are the dummy variables indicating whether a market efficiency outcome belongs to a particular treatment (T_1 serves as the baseline), ν_m are the random effects, and μ_{mt} is an idiosyncratic error term. The estimated coefficients β_2-β_4 measure the average causal treatment effects of our monetary and non-monetary interventions on the efficiency scores.

In our main model (see Eqs. (6) and (7)), we rely on standard errors without accounting for the correlation within markets. Hence, we employ randomization inference (RI) that was originally developed by Fisher (1953) and later advanced by Rosenbaum (2002) to account for such correlation. In addition, RI places no distributional assumptions on the errors and is valid even in small samples. RI computes the empirical distribution of the treatment estimates for a large number of randomly generated placebo treatments under the null hypothesis of no effect using a simulation method. The critical value of the treatment effect to be used for the inference test can be determined from a large number of simulations. We conduct the RI test using 1000 replications in the *ritest* Stata command developed by Heß (2017). The results from the RI test are presented along standard significance levels for our main results. In Appendix D, we provide a detailed mathematical representation of the RI procedure applied to our main panel regression models.

4.1. *Effects of information provision on purchased units*

First, we consider the effects of surge-price notifications and peer comparison on purchased units. Table 3 shows the effects of the treatments on purchased units. In columns 1, we report the average treatment effects for all experimental days (Days 1–4 during Months 6–15) of the treatment phase. To evaluate how participants from different treatment groups responded during and after critical peak days, we consider the following three subsamples: we report the results for peak days only (instances of Day 3) in column 2, for critical days only (instances of price surge events) in column 3 and for non-peak days (instances of Days 1, 2, and 4) in column 4. Column 5 presents the treatment effects aggregated at the monthly level. Next, we present and discuss the results of the regression models, and we consider the conventional significance level of at least 5 %.

On critical days, participants in the T3 and T4 groups, which received price surge notifications, reduced their purchases by 0.432 and 0.457 units, respectively, or approximately 14 % compared to T1 (see column 3 in Table 3 for effect sizes). However, the resource-conservation effect is not significant in T3 and T4 relative to T1 at the monthly level. We do not observe any resource-conservation effects relative to the T1 group in the treatment group that has received only peer-comparison information (T2).

4.2. *Information provision effects on efficiency scores*

In Table 4, we present the estimated average treatment effects of our informational interventions on participants' efficiency scores (see Section 3 for details about how these efficiency scores were calculated). We report the regression results for the same experimental periods as in the analysis of purchased units in Table 3. Although peak periods, especially critical days, (columns 2 and 3) have the largest potential to generate gains from trade and, consequently, the strongest propensity to increase market efficiency with optimal decision-making, we are also interested in analyzing any spillover effects of our informational interventions into non-peak periods (column 4) as well as aggregate effects on market efficiency scores at the monthly level (column 5).

As in our descriptive analysis, we find that when we consider only peak or critical days, the largest efficiency gain is achieved by providing both types of information (notifications about price surges and peer comparisons, T4). On peak days, participants in the T4 group increased their daily efficiency score by 0.202 points on average relative to participants in the T1 group (column 2 in Table 4). Participants in the T3 group managed to increase their average efficiency score by 0.138 points as well (see column 2 in Table 4). The positive "spillover" effect is not present (column 4 in Table 4). Meanwhile, pure peer comparison information (T2) had no effect on efficiency scores.

The estimated coefficients for the monthly efficiency scores when all treatment days are considered show that the largest and statistically significant efficiency gains were achieved by providing both types of information (T4) (column 5 in Table 4). Even though peer comparison information was not effective in promoting resource conservation, it may help to avoid large mistakes and extreme deviations from optimal purchase levels (see the descriptive evidence in Section 3), leading to more efficient outcomes in markets that also receive surge-price notifications (T4).

Although, Fig. 2 demonstrates that the participants quickly became familiar with the experimental environment gaining experience with the

[5] Likelihood-ratio test prefers panel tobit over pooled tobit model.

Table 4
Effects of the treatments on market efficiency scores.

Variables	(1)	(2) Period efficiency scores	(3)	(4)	(5) Monthly efficiency scores (all days)
	All treatment days	Peak days	Critical days	Non-peak days	
Peer comparison only (T2)	−0.014	0.060	0.192	−0.021	0.018
	(0.045)	(0.074)	(0.150)	(0.035)	(0.026)
Price notification only (T3)	$0.075^{☆}$	$0.138^{★★/☆}$	$0.305^{★★/☆☆}$	0.043	0.025
	(0.045)	(0.074)	(0.150)	(0.035)	(0.026)
Both (T4)	0.042	$0.202^{★★/☆☆☆}$	$0.454^{★★/☆☆☆}$	0.005	$0.056^{★★/☆☆}$
	(0.045)	(0.074)	(0.150)	(0.035)	(0.026)
Constant	$0.860^{☆☆☆}$	$0.732^{☆☆☆}$	$0.387^{☆☆☆}$	$0.884^{☆☆☆}$	$0.869^{☆☆☆}$
	(0.032)	(0.052)	(0.106)	(0.025)	(0.019)
Observations	3200	800	280	2400	400

Notes: This table presents the results of the estimated panel-data tobit regression model with random effects as described in Eq. (7). The dependent variable is the daily (or monthly) market efficiency score for individual markets. "All treatment days" refers to experimental periods 41–120 (Months 6–15), which followed the pre-treatment phase. "Peak days" are instances of Day 3. "Critical days" are instances of cost shocks, that is, price surge events. "Non-peak days" are instances of Days 1, 2, and 4. T1 denotes the treatment with cost shocks but no messages about surge pricing and no peer comparisons, T2 the treatment with only peer comparisons, T3 the treatment with messages about surge pricing but no peer comparisons, and T4 the treatment with messages about surge pricing and peer comparison information. Standard errors are presented in parentheses. Standard errors are presented in parentheses. $^{★★★/☆☆☆}$ $p < 0.01$. $^{★★/☆☆}$ $p < 0.05$, and $^{★/☆}$ $p < 0.1$ indicate significance levels, where filled stars $^{★}$ indicate significance levels preserved under randomization inference with clustering at market level (not available for constant term), while empty stars $^{☆}$ indicate significance levels that are sustained by the standard errors.

Table 5
Effects of the treatments on efficiency scores in high- and low-efficiency markets.

Variables	High efficiency				Low efficiency			
	All treatment days (1)	Peak days (2)	Critical Days (3)	Monthly (4)	All treatment days (5)	Peak days (6)	Critical days (7)	Monthly (8)
Peer comparison only (T2)	−0.052	−0.034	−0.033	−0.026	0.014	0.130	0.344	$0.048^{★/☆☆}$
	(0.066)	(0.117)	(0.217)	(0.051)	(0.045)	(0.086)	(0.227)	(0.019)
Price notification only (T3)	0.046	0.038	0.082	−0.010	−0.025	0.143	0.384	0.017
	(0.057)	(0.101)	(0.188)	(0.044)	(0.064)	(0.122)	(0.321)	(0.026)
Both (T4)	0.041	0.094	0.240	0.013	0.042	$0.279^{★★★/☆☆☆}$	$0.592^{★★★/☆☆☆}$	$0.085^{★★★/☆☆☆}$
	(0.066)	(0.117)	(0.218)	(0.051)	(0.045)	(0.087)	(0.227)	(0.019)
Constant	$0.907^{☆☆☆}$	$0.836^{☆☆☆}$	$0.627^{☆☆☆}$	$0.913^{☆☆☆}$	$0.827^{☆☆☆}$	$0.657^{☆☆☆}$	0.224	$0.840^{☆☆☆}$
	(0.047)	(0.083)	(0.154)	(0.036)	(0.032)	(0.061)	(0.160)	(0.013)
Observations	1600	400	140	200	1600	400	140	200

Notes: This table presents the results of the estimated panel-data tobit regression model with random effects as described in Eq. (7) but for two samples of markets: low-efficiency and high-efficiency markets. The dependent variable is the daily (or monthly) market efficiency score for individual markets. "All treatment days" refers to experimental periods 41–120 (Months 6–15), which followed the pre-treatment phase. "Peak days" are instances of Day 3. "Critical days" are instances of cost shocks, that is, price surge events. T1 denotes the reference treatment with cost shocks but no messages about surge pricing and no peer comparisons, T2 the treatment with only peer comparisons, T3 the treatment with messages about surge pricing but no peer comparisons, and T4 the treatment with messages about surge pricing and peer comparison information. $^{★★★/☆☆☆}$ $p < 0.01$.
$^{★★/☆☆}$ $p < 0.05$, and $^{★/☆}$ $p < 0.1$ indicate significance levels, where filled stars $^{★}$ indicate significance levels preserved under randomization inference with clustering at market level (not available for constant term), while empty stars $^{☆}$ indicate significance levels that are sustained by the standard errors.

induced market values before our treatment phase (Months 6–15), still, some considerable learning could have happened in the treatment phase that could significantly affect and confound our treatment effects. As a robustness test for considerable learning effects, we included monthly dummies in our main models to capture these learning effects. The inclusion of these dummies does not change our main results in any meaningful way (see Tables C3 and C4 in Appendix C).

4.3. *High-efficiency versus low-efficiency markets*

Our first and second hypotheses, according to which during surge-pricing periods, more information is better for resource conservation and overall market efficiency, were confirmed only in part. From a policy perspective, we expect that more information is most beneficial for consumers who normally are in less efficient and underperforming markets. To check whether this was the case in our laboratory experiment, we compare the effects of our monetary and non-monetary treatments in low- and high-efficiency markets.

We divide participants into low-efficiency and high-efficiency groups according to how their pre-treatment efficiency scores compare to the average pre-treatment efficiency scores of all markets during the last experimental week of the pre-treatment phase (Week 10). Participant groups with a below-average efficiency level are defined as low-efficiency markets, and vice versa. We choose the last week before the start of the treatment phase for such identification to avoid data noise caused by the learning process at the beginning of our laboratory experiment.

Table 5 shows that during peak/critical days (columns 2/3 and 6/7 in Table 5), significant efficiency gains were induced only in low-efficiency markets and only by the treatment combining critical peak notifications and peer comparisons (T4). Interestingly, the monthly regressions (columns 4 and 8 in Table 5) reveal that at the aggregate level, the inclusion of peer comparison (T2 and T4) enhances the welfare of low-efficiency markets but has no effect on high-efficiency markets. To summarize the results of Table 5, most of the gains in efficiency resulting from these informational interventions are produced by low-efficiency markets.

The implications of these findings are vital for policymakers as they

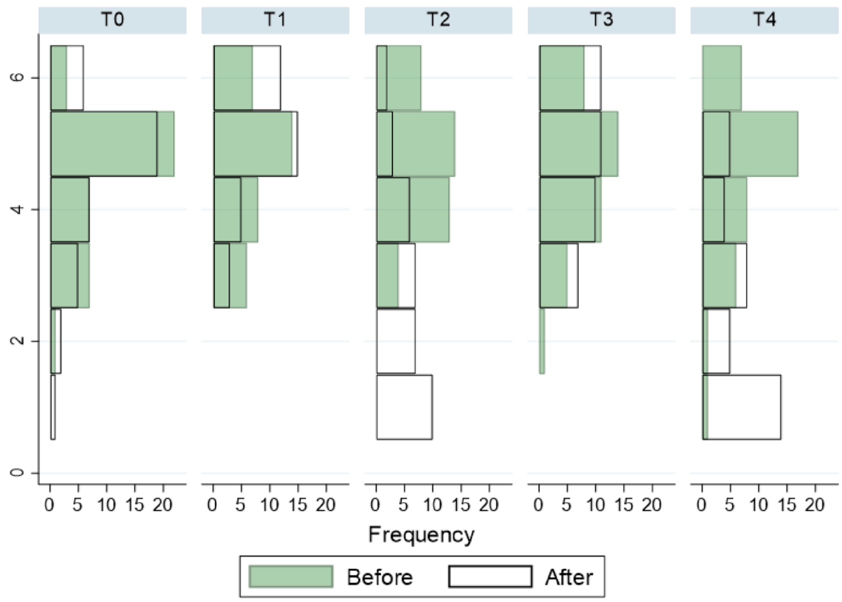

Fig. 5. Histograms of perceived usefulness (0–6) of peer comparison information by treatment group (T0–T4) before and after the treatment phase.
Notes: T0 denotes the treatment without cost shocks, T1 the reference treatment with cost shocks but no messages about surge pricing and no peer comparisons, T2 the treatment with only peer comparisons, T3 the treatment with messages about surge pricing but no peer comparisons, and T4 the treatment with messages about surge pricing and peer comparison information. We use the Kolmogorov–Smirnov test for the equality of distribution to see if preferences about peer comparison information vary between participants who did not receive such information (T0, T1, and T3) and those who did (T2 and T4). We find that the distributions of "before" rankings (indicated by green bars) are statistically not different between the two participant groups, but the distributions of "after" rankings (indicated by transparent bars) are statistically different.

suggest which informational treatment is effective in increasing welfare in low-efficiency markets during surge-pricing events. Our results show that a combination of monetary and non-monetary information is more effective in increasing welfare among underperforming consumers. In other words, consumers who fail to make optimal purchases should not be targeted only with monetary information because this information alone may not lead them to make better choices.

4.4. Do consumers perceive peer comparison information as useful?

We find that peer comparison information alone (T2) does not increase overall market efficiency, and the use of such information declined progressively over the course of our experiment (see Fig. E1 in Appendix E). To determine how the participants perceived the benefits of such information before and after they were exposed to our treatments, we asked the participants in all treatments to rate how useful historical daily/monthly consumption information containing both individual and market averages would be to them (see Section 2.2. for details about the questionnaires).

The distribution of the answers across the treatment groups is presented in Fig. 5. In all treatment groups, more than half of the participants perceived historical peer comparison information as very useful before the start of the treatment period, answering that they "strongly agreed" or "agreed" with the statement (values 6 and 5 in Fig. 5, respectively). The rating of this information did not change at the end of the experiment for groups that did not receive it during the treatment period (T0, T1, and T3). However, the participants that received this information (T2 and T4) rated it as less useful after having had access to it. In fact, most participants "strongly disagreed" or "disagreed" with the statement (48 % in T2 and 53 % in T4). We test the statistical difference between the distributions of "before" and "after" rankings by employing the Kolmogorov-Smirnov test and find that while these distributions are not statistically different across the treatments (T0–T5) at the beginning of the experiment, the distributions of preferences between participants who did not receive peer comparison information (T0, T1, and T3) and those who did (T2 and T4) are statistically different at the end of the experiment.

The finding that participants in the T2 and T4 groups perceived peer comparison information as less useful at the end of the experiment than at the beginning supports the results of the regression analysis, according to which peer comparison information alone, on average, fails to increase market efficiency (Table 4), with low-efficiency markets being an exception (Table 5). In any case, the negative shift in the perception of peer comparison's usefulness observed in the laboratory could indicate serious limitations in employing this behavioral intervention in the field.

5. Conclusions

By decreasing our electricity use, we can reduce our reliance on fossil fuels, significantly decarbonize our economies, and improve overall energy security. Many field experiments have suggested that various informational incentives encourage demand response by lowering electricity use. However, very little is known about the effects of such interventions on overall market efficiency. This paper contributes to our understanding of how information affects consumer purchase decisions as well as market efficiency. Our study is the first experimental (laboratory or field) study to contrast the effects of various informational interventions on overall market efficiency in the same experimental setting. The participants' choices provide us with useful insights not only into how consumers might respond to the implementation of informational programs but also into how overall market efficiency can change in increasingly volatile electricity markets.

In our laboratory experiment, we investigated two types of incentives—monetary information in the form of notifications about surge prices and non-monetary informational feedback in the form of peer comparisons—separately and together. We found that under volatile market conditions reminiscent of retail electricity markets, the combination of surge-price notifications and peer comparison information led to the largest market-efficiency gains. Surge-price notifications alone were the second-best intervention in terms of market efficiency. Furthermore, peer comparison alone did not produce the desired effects on resource conservation or market efficiency. In fact, it even led some consumers further away from their optimal purchase levels, although it simultaneously helped to avoid extreme deviations from optimal behavior. Interestingly, the participants in our two treatments that include peer comparison information significantly shifted their opinions about the usefulness of such information, rating it much less favorably after being exposed to the feedback. Further, the treatment using both informational interventions improved the average welfare the most in initially underperforming low-efficiency markets. Thus, we conclude that the monetary (price surge notifications) and non-monetary (peer comparisons) information was useful in achieving higher market efficiency when both types of feedback were provided.

In terms of external validity, our laboratory experiment was designed as a pilot study for a randomized, country-wide field experiment that tested the same behavioral interventions among residential electricity prosumers. The field experiment results offer valuable insights into the effectiveness of these interventions, though it did not assess their impact on market efficiency. Notably, both the field and laboratory experiments produced consistent findings: (1) the combination of critical peak price notifications and social comparison information was the most effective in reducing overall net electricity consumption; (2) critical peak price notifications alone successfully promoted conservation behavior during peak hours; (3) social comparison information, when used independently, was not effective; and (4) participants who received social comparison information perceived it as less useful than those who did not. Overall, the field experiment confirms the key findings from the laboratory, reinforcing the potential of these interventions in energy conservation efforts across the globe.

Based on these findings, the following policy implications can be drawn for real-world electricity markets, particularly in the context of addressing peak electricity demand and promoting energy conservation. First, adopt price-surge notifications as they have been shown to effectively reduce consumption during peak hours. Such notification might help to achieve energy conservation goals (e.g., in October 2022, the EU Council 2022 issued a regulation "on emergency intervention to address high energy prices," which required each EU member state to reduce electricity consumption by 5 % during peak hours) in a cheap and timely manner.

Second, use combined interventions for greater impact. The combination of surge-price notifications with peer comparison information was the most effective intervention in both conserving electricity and improving market efficiency, especially in underperforming or low-efficiency markets. Policymakers should consider integrating non-monetary interventions (such as social comparisons of electricity consumption) with price-based incentives to maximize conservation efforts. This dual approach is particularly valuable in markets with inefficiencies, where consumers may not initially respond optimally to either type of information alone.

Third, refine the use of peer comparison information. While peer comparison alone did not yield significant conservation effects and even led some consumers away from optimal behavior, it still helped mitigate extreme deviations from efficient consumption. Policymakers should be cautious in deploying peer comparison as a standalone tool. Instead, it should complement monetary incentives, ensuring that it supports behavior change rather than causing confusion or inefficiency among certain consumer segments.

Finally, focus on non-efficient users as our results show that combined interventions had the strongest impact in initially low-efficiency markets. Policy efforts should therefore target non-efficient users, where the potential for improving consumption behavior is greatest. Tailored feedback and notifications could be crucial in helping these consumers adjust their behavior, improving both individual welfare and overall market efficiency.

As for limitation of this study, our results suggest that price notifications led to greater reductions in purchases during periods of rising prices (high-scarcity periods), improving efficiency. However, our study does not investigate how price notifications during grid surplus (low-scarcity periods) affect purchases and market efficiency. The potential asymmetry in responses to supply shocks in opposite directions remains an important avenue for future research.

CRediT authorship contribution statement

Rimvydas Baltaduonis: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Resources, Software, Supervision, Writing – original draft, Writing – review & editing. **Jūratė Jaraitė:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Writing – original draft, Writing – review & editing. **Andrius Kažukauskas:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Supervision, Writing – original draft, Writing – review & editing.

Acknowledgments

The authors would like to thank participants at the Gettysburg College Economics Brownbag Seminar, the 2022 Winter Meeting of the Annual Lithuanian Conference on Economic Research, the 60th Annual Meeting of the Public Choice Society, the CRRI Advanced Workshop in Regulation & Competition at Rutgers University, the Helsinki Graduate School of Economics Seminar, the CERE Seminar at Umea University, and the Annual Conference of the European Association of Environmental & Resource Economists for helpful comments. Tsvetomir Petkov's valuable assistance with z-Tree programming is also greatly appreciated. This research is funded by the European Regional Development Fund (project No 01.2.2-LMT-K-718-02-0007) under grant agreement with the Research Council of Lithuania (LMTLT).

Appendix A

Table A1

Table A1
Demand and supply schedules.

	Supply 1	Demand, Day 1		Demand, Day 2		Demand, Day 3		Demand, Day 4	
Unit	MC, $	MB, $	Buyer	MB, $	Buyer	MB, $	Buyer	MB, $	Buyer
1	5	90	B1	295	B4	490	B4	295	B2
2	10	90	B2	285	B3	480	B4	285	B1
3	15	90	B3	220	B2	470	B4	220	B4
4	20	90	B4	210	B1	460	B3	210	B3
5	30	25	B1	140	B2	450	B3	140	B4

(*continued on next page*)

Table A1 (*continued*)

	Supply 1	Demand, Day 1		Demand, Day 2		Demand, Day 3		Demand, Day 4	
Unit	MC, $	MB, $	Buyer	MB, $	Buyer	MB, $	Buyer	MB, $	Buyer
6	40	25	B2	130	B1	440	B3	130	B3
7	50	25	B3	95	B4	430	B2	95	B1
8	60	25	B4	85	B3	420	B2	85	B2
9	75			55	B2	410	B1	55	B3
10	90			45	B1	290	B4	45	B4
11	105			35	B4	280	B4	35	B2
12	120			25	B3	270	B4	25	B1
13	135			20	B4	260	B3	20	B1
14	150			15	B2	250	B3	15	B3
15	165			10	B1	240	B2	10	B2
16	180			5	B3	230	B1	5	B4
17	200					160	B1		
18	220					150	B1		
19	240					140	B2		
20	260					130	B2		
21	280					120	B4		
22	300					110	B4		
23	320					100	B3		
24	340					90	B3		
25	360					80	B3		
26	380					70	B1		
27	400					60	B1		
28	420					50	B2		
29	440					40	B2		
30	460					30	B1		
31	480					20	B1		
32	500					10	B2		

Notes: This table provides the demand (marginal benefit or “resale value,” MB) schedules for Days 1–4, and the supply (marginal cost, MC) schedule, which is the same for all days. MB and MC are expressed in computer dollars ($).

Appendix B

Instructions for the experiment

Figs. B1–B7

This is an experiment in the economics of decision-making. If you follow the instructions carefully and make good decisions, you may earn a considerable amount of money, which will be paid to you in cash at the end of the experiment.

In this experiment, you will be purchasing units as a **buyer**. Every **15 s**, which we will call a “day,” the computer will present *Units* for you to purchase. You can decide how many units you want to purchase by clicking on the “Purchase Unit” buttons. The computer will record your purchases as final at the end of each day. There will be **4** days in a “week” and a total of **2** weeks in a “month.” At the end of each month, you will receive a monthly bill to pay for your monthly purchases. At that time, you will be able to see your *Profits (Losses)* from the choices you have made.

The amount of units you purchase and their corresponding *Resale Values* will determine the amount of money you make. Your *Resale Values* will be your private information and may vary among buyers. The *Cost* of purchased units will be a uniform *Price per Unit* that will be determined at the end of the month.

Depending on the number of units purchased by all participants, the computer will generate the market demand for the day. The market demand will be matched with the market supply, producing the *Market Price per Unit* of the day. At the end of the month, all buyers will be charged the uniform *Price per Unit* for all their purchases of that month. The *Price per Unit* will be calculated as the weighted average of the *Market Prices* during the month.

Your daily Profit = Resale Revenue - Costs =

= (Resale Value of Unit 1 Purchased + … + Resale Value of the Last Unit Purchased) - (Price per Unit x Units Purchased)

At the end of each month, your daily profits (losses) will update your *Current Balance*. Your initial *Current Balance* is **0** computer $. At the end of today’s experiment, your remaining *Current Balance* will be converted into cash at a rate of **X**[6] computer $ to 1 USD.

If you have any questions at any time, please raise your hand and a monitor will come to assist you.

[6] At the end of the session, each participant’s current balance was converted into cash using individualized exchange rates that depended on the identity of the buyer, that is, 1,800 computer $/USD for Buyer 1, 2500 computer $/USD for Buyer 2, 3400 computer $/USD for Buyer 3, and 4300 computer $/USD for Buyer 4. These exchange rates were selected to allow for equitable earnings given the induced heterogeneity in the participant roles.

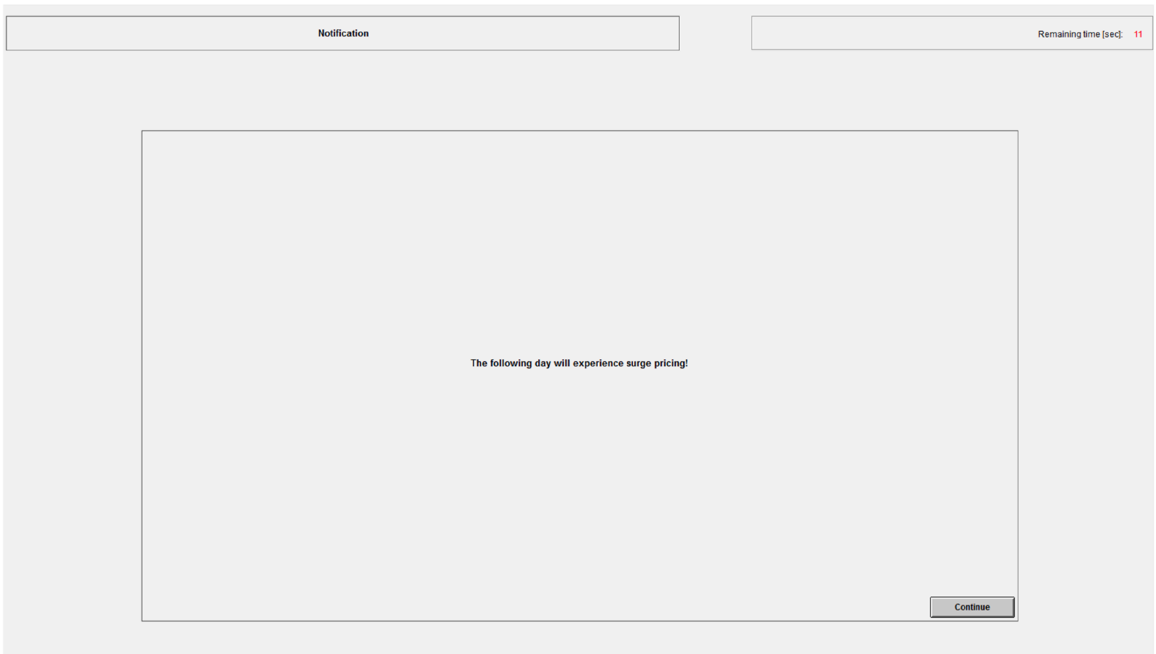

Fig. B1. Notification of expected surge pricing.

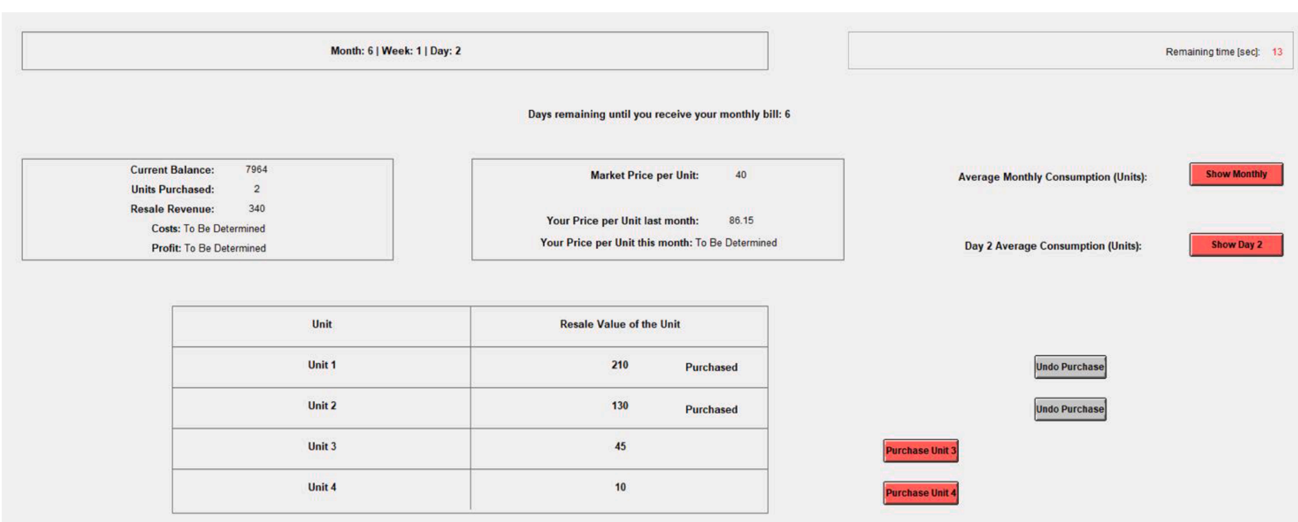

Fig. B2. Purchasing screen with buttons to the information on historical averages.

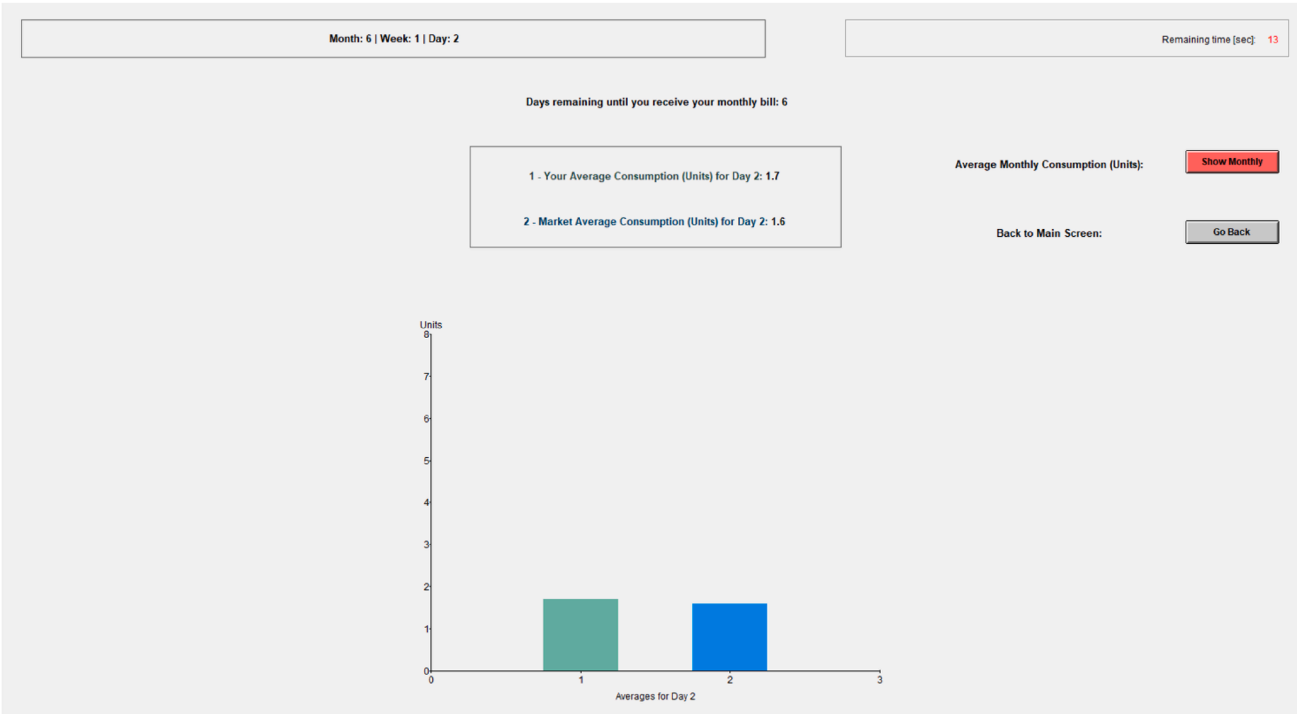

Fig. B3. Information about historical daily averages.

Fig. B4. Information about historical monthly averages.

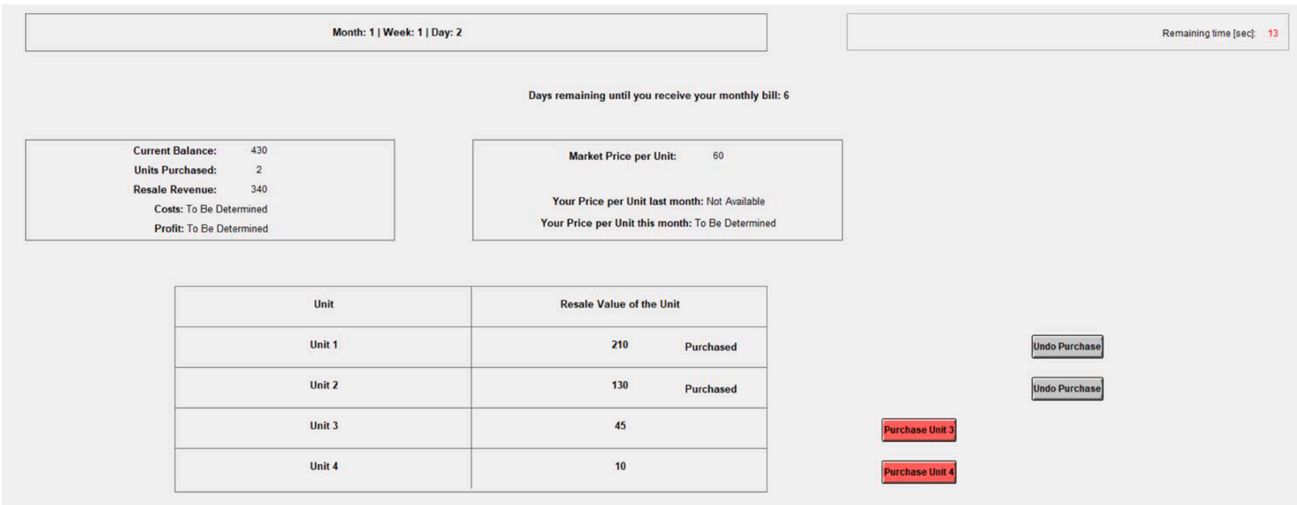

Fig. B5. Purchasing screen.

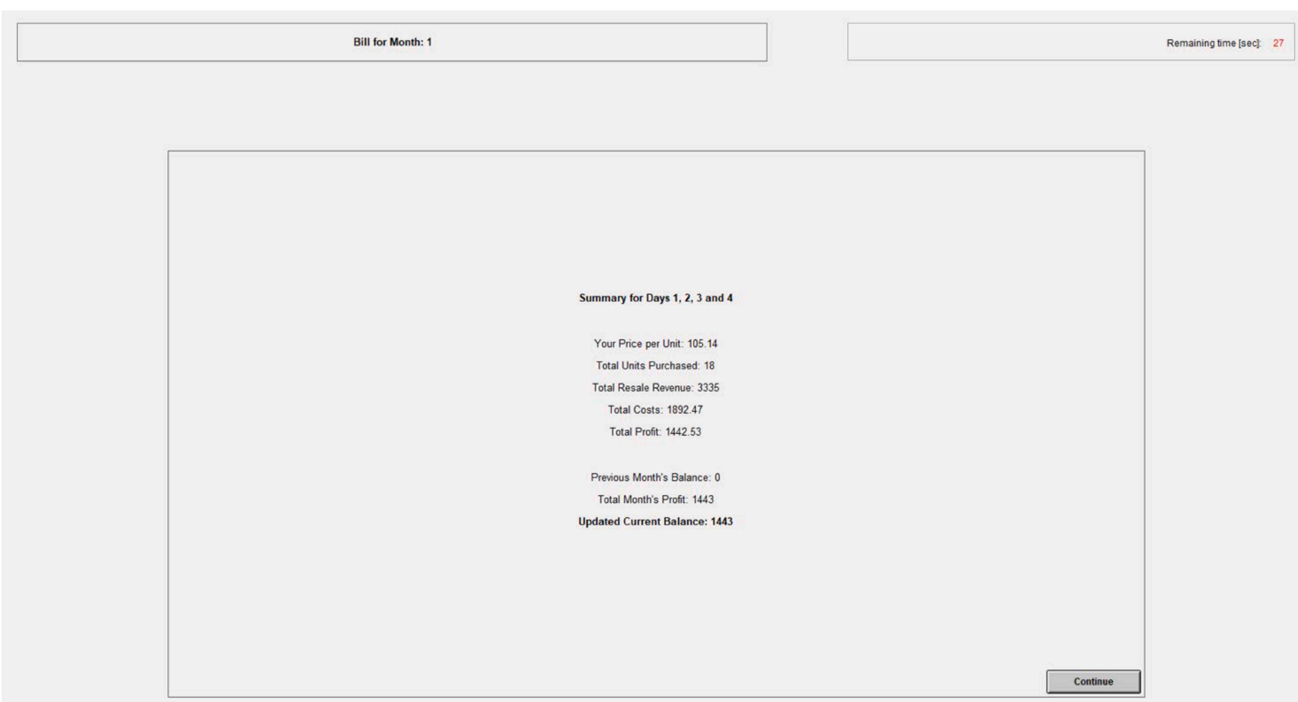

Fig. B6. Monthly bill.

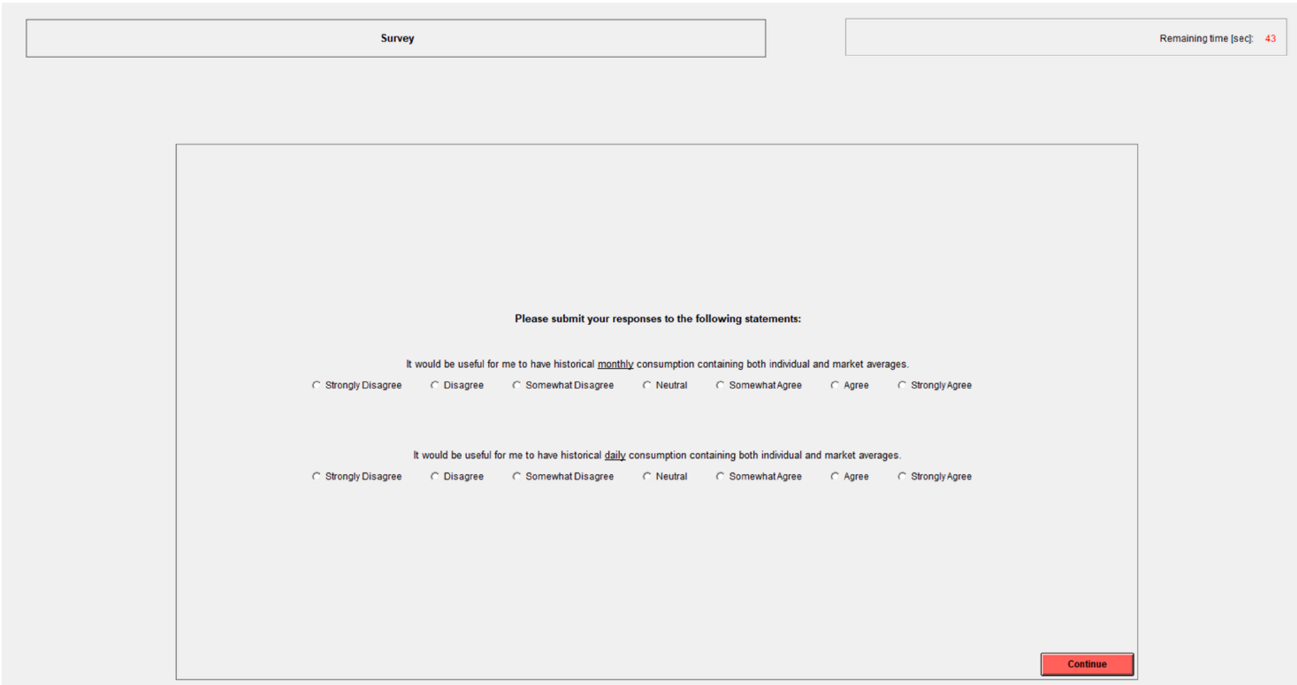

Fig. B7. Questionnaire.

Appendix C

Tables C1–C4

Table C1
Descriptive statistics for the experiment.

Variable	Obs.	Mean	Std. Dev.	Min	Max
Treatment group T0:					
Purchased units	4 800	2.378	1.655	0	8
Period efficiency score	1 200	0.863	0.121	0	1
Monthly efficiency score	150	0.880	0.061	0.705	0.974
Female	4 800	0.625	0.484	0	1
Non-U.S. citizen	4 800	0.175	0.380	0	1
Age	4 800	20.200	0.954	19	22
Treatment group T1:					
Purchased units	4 800	2.224	1.721	0	8
Period efficiency score	1 200	0.839	0.406	−8.477	1
Monthly efficiency score	150	0.858	0.084	0.605	0.988
Female	4 800	0.675	0.480	0	1
Non-U.S. citizen	4 800	0.050	0.164	0	1
Age	4 800	20.625	1.010	19	22
Treatment group T2:					
Purchased units	4 800	2.229	1.592	0	8
Period efficiency score	1 200	0.843	0.207	−2.328	1
Monthly efficiency score	150	0.870	0.070	0.544	0.951
Female	4 800	0.700	0.458	0	1
Non-U.S. citizen	4 800	0.075	0.263	0	1
Age	4 800	20.275	1.024	19	22
Treatment group T3:					
Purchased units	4 800	2.108	1.525	0	8
Period efficiency score	1 200	0.880	0.204	−2.510	1
Monthly efficiency score	150	0.885	0.108	0.232	1.000
Female	4 800	0.650	0.477	0	1
Non-U.S. citizen	4 800	0.100	0.300	0	1
Age	4 800	20.800	0.954	19	22
Treatment group T4:					
Purchased units	4 800	2.165	1.499	0	8
Period efficiency score	1 200	0.888	0.120	−0.254	1
Monthly efficiency score	150	0.912	0.058	0.697	0.974
Female	4 800	0.550	0.498	0	1
Non-U.S. citizen	4 800	0.025	0.156	0	1
Age	4 800	20.650	1.014	19	22

Notes: T0 refers to the treatment with no supply shocks, T1 to the treatment with cost shocks but no messages about surge pricing and no peer comparisons, T2 to the treatment with only peer comparisons, T3 to the treatment with messages about surge pricing but no peer comparisons, and T4 to the treatment with messages about surge pricing and peer comparison information.

Table C2
Comparison of the average daily efficiency scores during the treatment period relative to the pre-treatment period of each treatment group, %.

Samples	T0	T1	T2	T3	T4
All treatment days (Months 6–15)	2 %	−2 %	0 %	3 %	0 %
Peak days (instances of Day 3)	6 %	−16 %	−6 %	−3 %	1 %
Critical days	5 %	−55 %	−31 %	−21 %	−8 %
Non-peak days (instances of Days 1, 2, and 4)	1 %	2 %	2 %	6 %	−1 %

Notes: "All treatment days" refer to experimental periods 41 to 120 (Months 6–15), which followed the pre-treatment phase. "Peak days" are instances of Day 3. "Critical days" are instances of cost shocks, that is, price surge events. "Non-peak days" are instances of Days 1, 2, and 4. T0 refers to the treatment with no supply shocks, T1 to the treatment with cost shocks but no messages about surge pricing and no peer comparisons, T2 to the treatment with only peer comparisons, T3 to the treatment with messages about surge pricing but no peer comparisons, and T4 to the treatment with messages about surge pricing and peer comparison information.

Table C3
Effects of informational interventions on purchased units by including learning effects.

VARIABLES	(1) All treatment days	(2) Peak days	(3) Critical days	(4) Non-peak days	(5) Monthly
Peer comparison only (T2)	0.042	−0.015	−0.214	0.060	0.333
	(0.138)	(0.312)	(0.347)	(0.113)	(1.100)
Price notification only (T3)	−0.120	−0.277★	−0.432★★	−0.067	−0.960
	(0.138)	(0.312)	(0.347)	(0.113)	(1.100)
Both (T4)	−0.065	−0.196	−0.457★★	−0.022	−0.522
	(0.138)	(0.312)	(0.347)	(0.113)	(1.100)
Monthly dummies	Y	Y	Y	Y	Y
Observations	12,800	3200	1120	9600	1600

Notes: This table presents the results of the estimated panel regression model with random effects as described in Eq. (6). The dependent variable is daily (or monthly) purchased units by individual buyers. "All treatment days" refer to experimental periods 41 to 120 (Months 6–15), which followed the pre-treatment phase. "Peak days" are instances of Day 3. "Critical days" are instances of cost shocks, that is, price surge events. "Non-peak days" are instances of Days 1, 2, and 4. T1 refers to the treatment with cost shocks but no messages about surge pricing and no peer comparisons, T2 to the treatment with only peer comparisons, T3 to the treatment with messages about surge pricing but no peer comparisons, and T4 to the treatment with messages about surge pricing and peer comparison information. ★★★/✩✩✩ $p < 0.01$.
★★/✩✩ $p < 0.05$, and ★/✩ $p < 0.1$ indicate significance levels, where filled stars ★ indicate significance levels preserved under randomization inference with clustering at market level (not available for constant term), while empty stars ✩ indicate significance levels that are sustained by the standard errors.

Table C4
Effects of the treatments on market efficiency scores by including learning effects.

Variables	(1)	(2) Period efficiency scores	(3)	(4)	(5) Monthly efficiency scores (all days)
	All treatment days	Peak days	Critical days	Non-peak days	
Peer comparison only (T2)	−0.013	0.059	0.192	−0.021	0.018
	(0.045)	(0.073)	(0.150)	(0.035)	(0.026)
Price notification only (T3)	0.075✩	0.135★/✩	0.305★★/✩✩	0.043	0.025
	(0.045)	(0.074)	(0.150)	(0.035)	(0.026)
Both (T4)	0.043	0.203★★★/✩✩✩	0.454★★★/✩✩✩	0.005	0.056★★/✩✩
	(0.045)	(0.074)	(0.150)	(0.035)	(0.026)
Monthly dummies	Y	Y	Y	Y	Y
Observations	3200	800	280	2400	400

Notes: This table presents the results of the estimated panel-data tobit regression model with random effects as described in Eq. (7). The dependent variable is the daily (or monthly) market efficiency score for individual markets. "All treatment days" refers to experimental periods 41–120 (Months 6–15), which followed the pre-treatment phase. "Peak days" are instances of Day 3. "Critical days" are instances of cost shocks, that is, price surge events. "Non-peak days" are instances of Days 1, 2, and 4. T1 denotes the treatment with cost shocks but no messages about surge pricing and no peer comparisons, T2 the treatment with only peer comparisons, T3 the treatment with messages about surge pricing but no peer comparisons, and T4 the treatment with messages about surge pricing and peer comparison information. Standard errors are presented in parentheses. Standard errors are presented in parentheses. ★★★/✩✩✩ $p < 0.01$.
★★/✩✩ $p < 0.05$, and ★/✩ $p < 0.1$ indicate significance levels, where filled stars ★ indicate significance levels preserved under randomization inference with clustering at market level (not available for constant term), while empty stars ✩ indicate significance levels that are sustained by the standard errors.

Appendix D

Randomization Inference (RI) is a non-parametric approach which is particularly useful in experimental settings where the assignment of treatments is random, and it can provide valid inference even in the presence of complex correlation structures. In this appendix, we provide a mathematical explanation of how RI test works, specifically in the context of our regression model specified in Eq. (6) (and in Eq. (7), if the notation is adjusted).

Our treatment indicator vector is $\boldsymbol{T} = (T_2, T_3, T_4)$, and the null hypothesis for RI in case of Eq. (6) is that the treatment effects are zero:

$$H_0: \ \boldsymbol{\alpha_T} = 0$$

To test the hypotheses, we use estimated p-values for permutation tests on the basis of Monte Carlo simulations by implementing the following RI procedure:

First, we randomly permute the treatment labels to create a new treatment vector $\boldsymbol{T}^b$ for each permutation b (where b ranges from 1 to B, and the total number of permutations is 1000).

Second, for each permuted treatment vector $\boldsymbol{T}^b$, we re-estimate the regression model:

$$y_{it}^{\{b\}} = \alpha_1^{\{b\}} + \alpha_2^{\{b\}} T_2^{\{b\}} + \alpha_3^{\{b\}} T_3^{\{b\}} + \alpha_4^{\{b\}} T_4^{\{b\}} + u_i^{\{b\}} + \epsilon_{it}^{\{b\}}$$

Then we compute the test statistic $\widehat{\theta}^{\{b\}}$ for each permutation where $\widehat{\theta}$ represent the observed test statistic of interest, such as the vector of estimated treatment coefficients $\widehat{\boldsymbol{\alpha}}_T$.

Finally, we calculate p-value that is determined by the proportion of permuted test statistics that are more extreme than the observed test statistic $\widehat{\theta}$:

$$Pr\left(\left|\widehat{\theta}^{\{b\}}\right| \rangle \left|\widehat{\theta}\right|\right).$$

Appendix E

Fig. E1

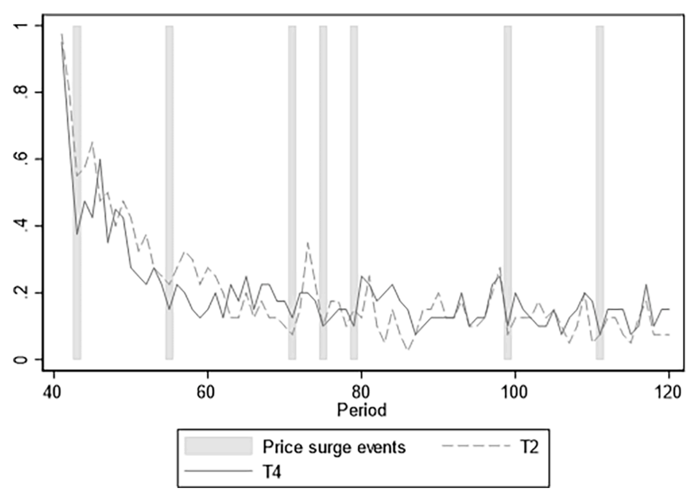

Fig. E1. The proportion of subjects in the T2 and T4 groups who clicked to check information about historic averages.
Note: This figure shows the proportion of subjects in T2 and T4 groups who clicked on buttons to request historic monthly or daily averages of purchased units.

Data availability

Data will be made available on request.

References

Abrahamse, W., Steg, L., Vlek, C., & Rothengatter, T. (2005). A review of intervention studies aimed at household energy conservation. *Journal of Environmental Psychology, 25*(3), 273–291.

ACER. (2023). Demand response and other distributed energy resources: What barriers are holding them back? 2023 Market Monitoring Report. 19 December 2023.

Adilov, N., Schuler, R. E., Schulze, W. D., & Toomey, D. E. (2004). The effect of customer participation in electricity markets: An experimental analysis of alternative market structures. In *Proceedings of the 37th Hawaii international conference on system sciences* (pp. 1–10).

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.

Atasoy, T., Harmsen - van Hout M.J.W., & Madlener, R. (2018). Strategic demand response to dynamic pricing: A lab experiment for the electricity market. FCN Working Paper No. 5/2018, 1–38. Available at SSRN: https://ssrn.com/abstract=3221747 or doi:10.2139/ssrn.3221747.

Baltaduonis, R., & Weisz, J. (2014). Information feedback effects in retail electricity markets. In *Proceedings of the 14th IAEE European conference, sustainable energy policy and strategies for Europe* (pp. 1–55). International Association for Energy Economics. October 28-31, 2014.

Barreda-Tarrazona, I., Garcia-Gallego, A., Pavan, M., & Sebater-Grande, G. (2012). Demand response in experimental electricity markets. *Revista Internacional de Sociologia, 70*, 127–165. Special Issue on Behavioral and Experimental Economics.

Buckley, P. (2020). Prices, information and nudges for residential electricity conservation: A meta-analysis. *Ecological Economics, 172*, Article 106635.

Buckley, P., & Llerena, D. (2022). Nudges and peak pricing: A common pool resource energy conservation experiment. *Journal of Behavioral and Experimental Economics*, Article 101928.

Brandon, A., List, J. A., Metcalfe, R. D., Price, M. K., & Rundhammer, F. (2019). Testing for crowd out in social nudges: Evidence from a natural field experiment in the market for electricity. *Proceedings of the National Academy of Sciences, 116*(12), 5293–5298.

Darby, S. (2006). *The effectiveness of feedback on energy consumption. a review for DEFRA of the literature on metering, billing and direct displays, 486* p. 26). Environmental Change Institute, University of Oxford.

Directorate-General for Energy (2022). Quarterly report on European electricity markets with focus on the impact of wholesale electricity prices on household retail prices. Market observatory for energy. Volume 14.

Ehrhardt-Martinez, K., Donnelly, K.A., & Laitner, S. (2010). Advanced metering initiatives and residential feedback programs: A meta-review for household electricity-saving opportunities. Washington, DC: American Council for an Energy-Efficient Economy.

EU Council (2022). Council Regulation (EU) 2022/1854 of 6 October 2022 on an emergency intervention to address high energy prices.

Faruqui, A., Sergici, S., & Sharif, A. (2010). The impact of informational feedback on energy consumption—A survey of the experimental evidence. *Energy, 35*(4), 1598–1608.

Fischer, C. (2008). Feedback on household electricity consumption: A tool for saving energy? *Energy Efficiency, 1*(1), 79–104.

Fisher, R. A. (1953). *The design of experiments*. Oliver & Boyd, Oxford.

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic literature, 42*(4), 1009–1055.

Heß, S. (2017). Randomization inference with Stata: A guide and software. *The Stata Journal, 17*(3), 630–651.

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.

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

Kažukauskas, A., Broberg, T., & Jaraitė, J. (2021). Social comparisons in real time: A field experiment of residential electricity and water use. *The Scandinavian Journal of Economics, 123*(2), 558–592.

Kažukauskas, A., Jaraitė, J., & Baltaduonis, R. (2024). Energy prosumers and behavioral interventions. Available at https://ssrn.com/abstract=4966190.

Lanot, G., & Vesterberg, M. (2021). The price elasticity of electricity demand when marginal incentives are very large. *Energy Economics, 104*, Article 105604.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic perspectives, 21*(2), 153–174.

Rassenti, S. J., Smith, V. L., & Wilson, B. J. (2001). Using experiments to inform the privatization/deregulation movement in electricity. *Cato Journal, 21*, 515.

Rassenti, S. J., Smith, V. L., & Wilson, B. J. (2003). Controlling market power and price spikes in electricity networks: Demand-side bidding. *Proceedings of the National Academy of Sciences, 100*(5), 2998–3003.

Rosenbaum, P. R. (2002). Covariance adjustment in randomized experiments and observational studies. *Statistical Science, 17*(3), 286–327.

Torriti, J., Leach, M., & Devine-Wright, P. (2011). Demand side participation: Price constraints, technical limits and behavioural risks. The future of electricity demand: Customers, citizens and loads, 88–105. The link to this source: https://books.google.com/books?id=s1UiAwAAQBAJ&lpg=PA88&ots=gxKkcxBVe-&dq=Demand%20side%20participation%3A%20price%20constraints%2C%20technical%20limits%20and%20behavioural%20risks&lr&pg=PA88#v=onepage&q=Demand%20side%20participation:%20price%20constraints,%20technical%20limits%20and%20behavioural%20risks&f=false.

Vesterberg, M. (2018). The effect of price on electricity contract choice. *Energy Economics, 69*, 59–70.